

基于生物信息分析技术的精准医疗

犹龙江

(贵州大学动物科学学院 贵阳 550025)

〔摘要〕 对目前生物信息学分析流程、具体应用进行分析,包括精神疾病药物使用指导、敏感性药物基因筛查等方面,对相关人才培养、生物医学大数据、精准医疗争论等相关问题进行讨论,希望对相关学科的应用与发展有所启发。

〔关键词〕 精准医疗;生物信息学;信息分析技术

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2019.04.003

Precision Medicine Based on Analysis Technologies of Bioinformatics YOU Longjiang, College of Animal Science, Guizhou University, Guizhou 550025, China

〔Abstract〕 The paper analyzes the analysis procedure and specific application of bioinformatics, including guidance on the use of psychiatric drugs and drug sensitivity genetic testing, etc. It discusses the cultivation of concerned talents, biomedicine big data, arguments about precision medicine and other related issues, expecting to provide inspiration for the application and development of related disciplines.

〔Keywords〕 precision medicine; bioinformatics; information analysis technology

1 引言

2015 年 1 月 20 日美国总统奥巴马宣布启动精准医学计划并决定一期投入约 2 亿美元。2015 年 12 月 11 日在上海成立“中国个体化用药-精准医学科学产业联盟”,标志我国正式开始精准医疗的探索。同年召开首次精准医疗战略专家会议,最终决定在 2030 年前对精准医疗领域投入 600 亿元。精准医疗根据患者基因、生活环境和方式的个体特异性推测疾病的发生,从而采取有效、精准的疾病预防与治疗手段。精准医疗的发展是建立在目前科学技术快速发展基础上的,如人类全基因组测序、生物信息分析技术、大数据

分析工具等。当前精准医学用于人类癌症治疗靶向药物的创新临床实验、综合性用药选择以及克服个体耐药性。

人类基因组计划(Human Genome Project, HGC)的目标是解码生命,了解生命起源、成长与衰退;认识个体之间的差异性原因,对人类疾病的发生机制、致病机制提供科学依据,同时极大促进相关行业的发展。如基因测序技术、功能基因组学、比较基因组学、生物信息学、计算生物学等。其中生物信息学和计算生物学是对物种个体基因的基础且关键的分析技术,分析人员掌握技术的熟练程度与分析结果有密切联系。

2 生物信息学概述

2.1 基本技能

生物信息学与计算生物学之间界限比较模

〔修回日期〕 2019-04-21

〔作者简介〕 狄龙江,硕士研究生。

糊, 生物信息学侧重对数据的提取、挖掘, 而计算生物学侧重对数据的处理、运用。目前生物信息学和计算生物学在处理对象和方法上已无显著差别, 分析方法大同小异、流程相似。生物信息学需掌握的技能是运用开源数据分析工具挖掘海量的基因测序数据, 而对数据分析的能力决定对复杂生命认识与理解。在进行生物信息学分析之前应掌握以下几点: (1) 可利用 Shell、R、Python 等计算语言编写脚本将机械性的文件处理操作转化为简单的指令运行。(2) 使用强大的 Linux (Unix) 系统运行开源数据处理程序。(3) 掌握 R 语言的探索性数据分析方法。(4) 了解基因组数据文件格式。如 FASTA、FASTQ、SAM、BAM 等。(5) 具有一定的分子生物学知识背景。2001 年后至今被称为后基因组时代, 完成人类基因组测序、各种模式生物基因组测序, 基因组学的研究重心也转向功能基因组学。

2.2 研究方向与应用

经过十几年的发展现已形成多个研究方向, 根据分析数据来源不同, 主要分为全外显子测序分析、转录组测序分析、染色质免疫共沉淀测序。以往生物信息学研究领域重点在于基因表达谱、个体物质代谢网络、基因芯片设计和蛋白质数据分析。进入后基因组时代其研究领域有了重大发展——扩展至结构基因组学、功能基因组学、比较基因组学、蛋白质组学、药物基因组学、肿瘤基因组学、分子流行病学和环境基因组学。

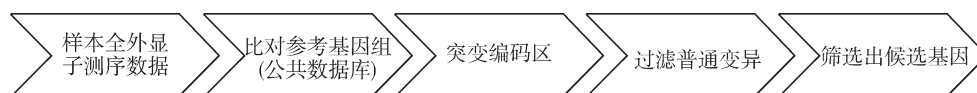


图 1 全外显子测序数据的生物信息学分析流程

3.2 转录组测序分析技术

指特定组织或细胞发生在某一阶段或功能状态下转录出来的核糖核酸 (Ribonucleic Acid, RNA) 总和, 主要包括 mRNA 和非编码 RNA (ncRNA)。转录组学是对基因结构、功能新的研究出发点, 是对研究对象在特定发育阶段下的分析。相比全基因

3 生物信息学分析流程

3.1 全外显子测序分析技术

通过捕获全基因组的外显子区域、富集后进行高通量测序, 能够直接发现与蛋白质功能变异相关的遗传变异。人类基因组中外显子仅占约 1%^[1], 但作为编码蛋白质的序列区域, 且超过 85% 的孟德尔遗传病都与基因组外显子变异有密切联系, 更适用于临床发现、确诊部分孕期疾病^[2]。目前产前诊断常用方法是 B 超影像学检查, 发现胎儿异常后采用核型分析或染色体微阵列芯片等技术, 综合诊断率约 14 ~ 16%^[1-2], 更高效、精确的诊断技术迫切需要。WES 的临床推广运用可能将孟德尔遗传病诊断率提高到 20 ~ 80%。全外显子测序分析具体流程, 见图 1。外显子测序后获得原始数据, 过滤除去高通量测序时插入的接头引物, 然后与从公共数据库 (如 NCBI、Ensemble 等) 获取的参考基因组进行比对, 找出突变编码区, 筛选出与遗传疾病相关的基因变异, 成为候选基因并进行讨论和分析, 总结实验结果。通过全基因组外显子测序分析技术不仅能有效、准确地定位致病基因位点, 还能发现新的突变位点的致病基因, 对阐明相关疾病发病机理、机制、遗传诊断具有重要的研究和推广价值。此外不依靠昂贵、精确的大型设备仪器, 而是多依据计算机对测序数据的处理及分析人员对公共数据库的理解和运用。该技术在临床疾病研究中具有推广与应用价值。

组分析, 转录组分析具有更高的灵敏度, 可以精确识别可变剪切位点、基因的表达水平; 对任意物种的全基因组分析甚至可能发现新的转录本、检测融合基因^[3]。组学分析将可能为中医理论中的阴阳学说进行科学分析与数据结果实践证明。通过分析阳类和阴类急性缺血性中风, 可以使用部分差异 lncRNA 和 miRNA 作为潜在疾病诊断的标记物, 是十

分具有潜力的研究方法。

4 基于生物信息分析技术的精准医疗

4.1 精神疾病药物使用指导

4.1.1 第 2 类精神药物依赖 近年修改《麻醉药品和精神药品管理条例》有关规定,国家食品药品监管总局、公安部、原国家卫生计生委决定将含可待因的复方口服液体剂(包括口服溶液剂、糖浆剂)列入第 2 类精神药品管理^[4]。其原因在于对敏感药物基因的临床实验发现 CYP2D6 超快代谢哺乳期妇女,分泌到乳汁中可待因可能出现药物过量,导致乳儿可能出现致死性不良反应^[5]。随着研究人员对药物代谢相关基因的研究,对部分药品的使用说明进行修改,体现出现代医学对精准用药的要求。随着测序技术以及相关行业的发展,目前进行指定药物相关基因的检测费用约 3 000 元/次,周期为 3~4 天,检测结果终身有效。

4.1.2 指导精神药物使用 武汉市优抚医院通过精神疾病患者的临床症状、基因检测分析结果为其制定精准治疗方案^[6]。取患者口腔上皮细胞进行药物基因检测,结合临床用药结果分析,该患者应选用 CYP2D6 基因型广泛代谢型药物,典型药物有奋乃静、氟哌啶醇等。

4.2 敏感性药物基因筛查

4.2.1 通过敏感性药物基因检测实现个体化用药 根据用药患者个体差异引起的药物吸收、转运、代谢等不同而实现有效的联合用药、精准治疗,降低患者对药物的耐受性和用药成本。药物基因检测常用方法有 3 种:实时荧光定量 PCR (qPCR)、荧光原位杂交 (FISH)、基因测序技术。qPCR 技术常用于对药物干扰基因 mRNA 的表达水平的检测,判断使用相关药物的疗效。FISH 是使用荧光探针技术对药物或疾病相关基因的检测。而测序技术通过对个体全基因组测序后进行生物学分析,筛选出潜在的遗传疾病和相关治疗药物蛋白基因多态性。如细胞色素 P450 为自身氧化的亚铁血红蛋白家族,又称多功能氧化酶,多参与内源性物质的代谢。

CYP2D6 基因型为 CYP450 酶系中重要的一种并且是唯一一种不能被诱导的酶。研究表明 CYP2D6 参与临床常用药物的体内代谢^[7],如奋乃静、氟哌啶醇等抗精神疾病药物,氯丙嗪、美沙酮等镇静类药物,吉非替尼片、他莫昔芬等抗肿瘤药物。

4.2.2 肿瘤领域个体化用药案例 香港中文大学莫树锦教授的临床实验说明药物基因检测的潜在价值。该试验目的是探究易瑞沙是否提高亚洲肺癌患者的生存期望,首次引入基因检测因素,将服用药物的患者进行表皮生长因子受体 (Epidermal Growth Factor Receptor, EGFR) 基因检测。实验结果证实易瑞沙对敏感 EGFR 突变患者具有良好的治疗效果,也提示非敏感患者应接受常规治疗方案。该实验是肿瘤治疗领域率先引入的个体化用药案例之一。

4.3 个体疾病潜在的精确诊断方式

4.3.1 诊断疾病的分子标记物 随着测序、序列分析技术的不断发展,许多常见和罕见的系统疾病分子标记物被确定,如阿尔茨海默症 (Alzheimer's Disease, AD)、帕金森症 (Parkinson's Disease, PD)、多发性硬化症 (Multiple Sclerosis, MS) 等。尽管疾病的发生存在多因素与异质性的影响,但是分子标记物的确定能作为疾病精确诊断的方式。就 AD 而言,目前已证实遗传因素在其病理机制中扮演重要角色^[8]。2011 年的文献证实基因 APP、RS-EN1 和 PSEN2 的遗传突变直接导致早发性阿尔茨海默症 (Early Onset Alzheimer's Disease, EOAD),但是之后在出现 EOAD 的临床患者中并没有检测到这 3 个基因的突变。随着研究深入目前已发现 3 个机制不明的小家系 EOAD 的遗传因素。第 1 个为 NOTCH3 基因的错义突变^[9];第 2 个在 SORL1 的错义和无意义的突变^[10];第 3 个为在脑小胶质细胞中表达的免疫吞噬受体 TREM2 外显子 2 变异与 AD 的早发性和迟发性具有密切联系^[11]。通过不断的研究发现疾病的致病因子很多,机制也很复杂。

4.3.2 肿瘤筛查领域 在肿瘤疾病研究领域基因变异检测是精准医疗中重要的组成部分^[12]。临床肿瘤患者多用组织活检技术,结果不完全准确,存在

65% 的假阳性^[13]。但基因检测技术可以将癌症的诊断提前到细胞癌变阶段,为肿瘤的预防提供基因水平上的指导。

5 讨论

5.1 生物医学大数据应用

5.1.1 价值与意义 目前正处于由基因组学、蛋白质组学、代谢组学、影像学、临床诊断数据等构成的生物医学大数据时代。对生物医学数据信息的理解、运用将扩展为现阶段对患者从诊断到预防再到个性化治疗的疾病管理新模式。如果使用得当大数据将成为具有重要价值的资源,用于改善当前的医疗服务和降低医疗成本。到 2022 年美国医疗保健占 GDP 的比例预计达到 19.9%^[14]。中国是人口大国的新兴经济体,尤其是新的高通量测序平台、实时成像系统、亚洲基因数据库建立将极大促进中国未来生命科学发展。

5.1.2 构建相关数据库 生物信息学发展从个体基因组测序扩展到测量表观基因组数据,研究基因的表达过程不仅仅局限于 DNA 序列的修饰,如 DNA 甲基化修饰、His - tone 修饰^[15]。生物信息的获取除基因组学外还包括转录组组学、蛋白质组学、代谢组学等。而组学就是强调对生物体结构、功能和生物分子动力学进行集体的量化。大数据可以提高临床研究实用性,有助于有效精确地对患者进行分类,这是定制个性化医疗方案和充分利用医疗资源的关键点。从人口学角度出发,研究某人疾病的发生过程可以用来预测和预防同样疾病的发生,构建中国生物医学数据库具有重要的战略意义。

5.2 现阶段精准医疗争论

5.2.1 基因检测商业化 基因检测技术在医学检查、疾病诊断、精准用药等方面发挥巨大作用。这项检测使一些重大慢性疾病治愈率大幅提高,基因检测逐步被接受,用于商业化使用^[16]。基因测序是精准医疗的基础,药物基因的筛选虽然提高治疗药物的准确性,但是也限制治疗药物的使用。血液肿

瘤学家维奈·普拉萨德 (Vinay Prasad) 于 2016 年在《自然》发表相关文章,指出绝大多数的肿瘤患者并没有通过个体化的精准医疗方案提高生存几率、生活质量,相反带来的是额外的昂贵医疗费用^[17]。但笔者认为目前个体化治疗方案的失败更多的可能是目前方法学理论的失败^[18]。根据 2018 年 5 月在北京召开的中国消费级基因检测行业峰会,如果参照美国基因检测近 5 年的渗透率变化,2020 年中国的渗透率达到 3.5%。

5.2.2 药物精准使用 对患者进行药物敏感性基因筛查,除筛查结果终身可参考使用外还能为医生临床治疗用药提供指导,减少不敏感药物的使用,降低患者多次使用不同药物产生的药物耐受性,减少治疗负担。实现有效用药从而避免患者多次重复就诊和药物使用,充分有效利用医院和药品资源。

5.3 生物信息学人才培养

5.3.1 引导学生兴趣 基因组学研究基因突变或差异产生的个体差异性,转录组学和蛋白质组学研究某种外因对个体产生的影响,而代谢组学则是在研究个体目前的差异性变化。基因在细胞核中转录出信使 RNA,然后翻译为功能性蛋白质,参与复杂的生命活动行为。利用开源软件分析数据,掌握一门汇编语言,具有一定的分子生物学知识背景,是生物信息学分析的基础要求。学生开始学习时应多了解最新研究开发出的开源数据挖掘软件、分析算法、流程、数据库、大量的高质量文献和具有重复性的代码参考。然后根据兴趣选择研究方向,包括计算生物学、基因组学、转录组学等。

5.3.2 完善培养方案 相关文献每年发表总量增长说明生物信息技术广泛用于生命科学研究。在临床应用中医学信息学利用患者疾病现阶段的全外显子测序或转录组测序,筛选出致病突变基因,选择靶向用药,实现精准医疗^[6]。在精准医学用于人类肿瘤治疗靶向药物的创新临床实验、使用选择以及克服个体耐药方面,生物信息技术是最关键、最基础的分析技术^[12]。生物信息学在科学领域的研究和临床实践的实用价值以及专业人才分类说明研究行业对该专业人才的需求。高校应注重该专业人才培养

养, 提出相适宜的人才培养方案体系。

6 结语

目前的疾病诊断和治疗主要依赖于患者的主观感受、病史资料、医生的主观辨别和大量昂贵设备辅助诊断结果, 在临床治疗过程中不断进行验证和修改治疗方案。在未来的疾病诊断方面, 尤其是恶性疾病的早期诊断, 疾病相关基因分子标记物的筛查具有重要意义。随着基因组学研究、疾病机制和相关基因分子标记物不断发现和证实, 疾病基因分子标记物确定精确性不断提高, 未来有望实现血液中分子标记物的 PCR 定性检测和 qPCR 的定量检测, 大大提高疾病正确诊断率。个体化的精准医疗方案虽然因个体基因的多态性而缺乏特异性的靶向治疗药物, 但笔者认为这项技术为患者提供更适合的医疗方案, 又避免不敏感靶向治疗药物的使用, 同时也为新型靶向药物的研究提供方向。随着测序技术不断升级, 分析方案不断优化, 检测成本不断下降以及靶向治疗药物的开发, 未来有望将基因检测纳入医保政策。

参考文献

- 1 Ng S B, Turner E H, Robertson P D, et al. Targeted Capture and Massively Parallel Sequencing of 12 Human Exomes [J]. *Nature*, 2009, 461 (7261): 272–276.
- 2 Ng S B, Buckingham K J, Lee C, et al. Exome Sequencing Identifies the Cause of a Mendelian Disorder [J]. *Nature Genetics*, 2009, 42 (1): 30–35.
- 3 Maher C A, Kumar–Sinha C, Cao X, et al. Transcriptome Sequencing to Detect Gene Fusions in Cancer [J]. *Nature*, 2009, 458 (7234): 97–101.
- 4 中国麻醉药理学学会麻醉药理学专业委员会. 麻醉药品和精神药品管理条例 [C]. 全国第三次麻醉药理学学术会议论文集, 2012: 15.
- 5 杨传青, 王周然, 孔志, 等. 37 例可待因类止咳水成瘾患者临床特征分析 [J]. *中国药物滥用防治杂志*, 2018, 24 (2): 70–72.
- 6 董红霞. 精准医疗基因检测指导精神科用药方案的一例报告 [J]. *神经损伤与功能重建*, 2019, 14 (1): 51–52.
- 7 Maggo S D S, Savage R L, Kennedy M A. Impact of New Genomic Technologies on Understanding Adverse Drug Reactions [J]. *Clinical Pharmacokinetics*, 2016, 55 (4): 419–436.
- 8 Hollingworth P, Harold D, Jones L, et al. Alzheimer's Disease Genetics: current knowledge and future challenges [J]. *International Journal of Geriatric Psychiatry*, 2011, 26 (8): 793–802.
- 9 Kinsella E. Exome Sequencing Reveals an Unexpected Genetic Cause of Disease: NOTCH3 mutation in a Turkish family with Alzheimer's disease [J]. *Neurobiology of Aging*, 2012, 33 (5): 1008.
- 10 High Frequency of Potentially Pathogenic SORL1 Mutations in Autosomal Dominant Early-onset Alzheimer Disease [J]. *Molecular Psychiatry*, 2012, 17 (9): 875–879.
- 11 Pottier C, Wallon D, Rousseau S, et al. TREM2 R47H Variant as a Risk Factor for Early-onset Alzheimer's Disease [J]. *Journal of Alzheimers Disease* Jad, 2013, 35 (1): 45–49.
- 12 张乐吟, 孙磊涛, 沈敏鹤. 基因变异检测技术在恶性肿瘤精准医疗中的应用 [J]. *中国肿瘤生物治疗杂志*, 2019, 26 (1): 22–28.
- 13 Cao C, Mak L, Jin G, et al. PRESME: personalized reference editor for somatic mutation discovery in cancer genomics [J]. *Bioinformatics*, 2019, 35 (9): 1445–1452.
- 14 Cuckler G A. National Health Expenditure Projections, 2012–22: slow growth until coverage expands and economy improves [J]. *Health Affairs*, 2013, 32 (10): 1820–1831.
- 15 Perez J A, Poon C C Y, Merrifield R D, et al. Big Data for Health [J]. *Biomedical & Health Informatics IEEE Journal of*, 2015, 19 (4): 1.
- 16 廖庆宏. 健康管理中心基因体检的营销策略研究 [D]. 电子科技大学, 2014.
- 17 Prasad V. Perspective: the precision-oncology illusion [J]. *Nature*, 2016, 537 (7619): S63.
- 18 Abrahams, Edward, Eck, Stephen L. Precision Oncology is Not an Illusion [J]. *Nature*: 2016 (539): 357