

• 专家述评 •

检验组学在精准医疗中的应用价值^{*}刘 伟^{1,2}, 张 阳¹, 黄姣祺¹, 田晖艳¹, 府伟灵^{1△}

(1. 陆军军医大学第一附属医院检验科, 重庆 400038; 2. 西南大学材料与能源学部, 重庆 400715)

摘 要:随着多组学时代的到来, 检验组学这一综合性科学应运而生。检验组学是综合基因组学、蛋白组学、代谢组学、免疫组学及药物基因组学等多组学的检验数据结果, 并利用大数据的计算方法对各个组学的数据指标进行分析、筛选、分类、整合的多领域交叉学科。通过大数据、高通量的数据分析方法, 可得到更加精准的病理指标置信区间, 完善疾病的早期预警诊断体系, 形成个体化的生物学医学信息库, 以满足精准诊断、精准治疗的要求。尽管检验组学大数据分析在实际应用中处于起步阶段, 但通过与相关科学技术的结合发展, 检验组学将会成为临床检验医学领域的重要发展方向。

关键词:检验组学; 临床检验数据; 大数据分析; 多组学整合

DOI:10.3969/j.issn.1673-4130.2018.09.001

中图法分类号:R446

文章编号:1673-4130(2018)09-1025-04

文献标识码:A



府伟灵

1 检验组学诞生的时代背景

精准医疗, 也称为个性化的 (personalized)、预测性的 (predictive)、预防性的 (preventive)、参与式的 (participatory) “4P 医疗方式”, 是一种新型的个性化医疗实践方法^[1]。2012 年, 美国国立卫生研究院 (NIH)

确识别的准确患者提供准确药物的准确剂量方案。

精准医疗的内涵不仅包括精准治疗, 同时也对疾病的精准检验以及精准诊断提出了更高的要求。检验医学是现代实验室技术与临床医学的结合, 是运用现代物理化学方法及手段进行医学诊断的一门学科, 是包括临床检验基础、临床生物化学、微生物学、免疫学、血液学、寄生虫学检验等多个专业方向的交叉领域学科。随着组学研究的发展, 许多组学新指标的出现极大地丰富了临床医学检验项目。基因组学、蛋白组学、代谢组学、免疫组学等多组学的整合研究方法为临床疾病诊断、疗效评价及预后判断提供了全面、快速、准确的实验室数据。多组学整合分析是通过来自不同组学的数据源进行归一化处理、比较分析, 建立不同组间数据的关系, 综合多组学数据对生物过程从基因、蛋白和代谢水平进行全面深入的阐释, 从而更好地对疾病的各项指标进行汇总研究, 以更精准地诊断疾病。多组学整合, 多项指标联合分析是医学检验领域未来发展的方向。

然而, 医学检验的多组学整合分析应用于患者临床大数据时也面临一些挑战, 由于数据量巨大而且结构复杂, 医务人员很难直接从中获取有用的信息。医学检验的多组学整合分析是基于大规模人群的基因数据、生物样本数据 (包括蛋白质分析、细胞数量、代

将精准医疗定义为“建立在了解个体基因、环境及其生活方式基础上的新型疾病预防和治疗方法”^[2]。其本质是利用医学前沿技术以及基因组学、蛋白组学、代谢组学等组学技术^[3-6], 对于特定疾病与大样本群体进行生物标记物的分析、鉴定、验证与应用, 进而更精确地找到致病因素以及治疗靶点^[7], 进而对同种疾病的不同病程进行精细分类, 以最终实现对于疾病和患者的个性化治疗, 从而提高疾病预防与诊治的效益^[8-10]。精准医疗包含 5 个层面, 即精准剂量、精准药物、精准甄别、精准患者与精准时间。其核心思想为通过精准诊断获得患者全维度的疾病状态信息, 从而选择疾病精准治疗方案, 即在准确时间段内, 为经准

^{*} 基金项目: 国家重点基础研究发展计划 (“973” 计划) 资助项目 (2015CB755402); 国家重点研发计划 “精准医学研究” 项目 (2017YFC0909902); 西南医院智慧医疗重点项目 (SWH2017ZDCX4210)。

通信作者简介:府伟灵, 男, 陆军军医大学附属第一医院检验科主任, 教授, 博士生导师, 全军医学科学委员会检验医学专委会主任委员, 中国研究型医院学会检验医学专委会主任委员, 中国医促会临床检验医学专委会主任委员, 中国医师协会检验医师分会副主任委员, 国家重点基础研究计划 (“973” 计划) 项目首席科学家。△ **通信作者,** E-mail: fwl@tmmu.edu.cn。

本文引用格式:刘伟, 张阳, 黄姣祺, 等. 检验组学在精准医疗中的应用价值[J]. 国际检验医学杂志, 2018, 39(9): 1025-1028.

谢物、DNA 和 RNA 以及全基因组测序等)、日常生活信息等数据的整合而发展起来的,大量的数据集合在一起,需要工具发掘其中的价值。为解决这一难题,检验组学应运而生。检验组学利用大数据高通量分析方法为海量临床检验数据提供了有力的运算技术支持,实现了计算机与医疗的跨界协同发展。大数据分析能对生物医疗大数据进行有效的分析和挖掘,有利于对疾病的发病机制进行深入的研究并对疾病的诊断与治疗有强大的推进作用。

2 检验组学的概念与内涵

检验组学是综合基因组学、蛋白组学、代谢组学、免疫组学及药物基因组学等多组学的检验数据结果,并利用大数据的计算方法对各个组学的的数据指标进行分析、筛选、分类、联合的多领域交叉学科。通过大数据、高通量的数据分析方法,可得到更加精准的病理指标置信区间,完善疾病的早期预警诊断体系,形成个体化的生物医学信息库,以满足精准诊断、精准治疗的要求。

基因组学是以生物体内基因组的全部基因为研究对象,从整体水平上探索全基因组在生命活动的作用及其内在规律和内外环境影响机制的科学。通过对全基因组整体水平的研究分析,认识生命活动的规律。检验组学可通过对某些基因组数据测试分析,进而评估某些个体疾病复发的危险率以及化疗效果,这有助于医生获得更多的治疗信息并进行个性化医疗。

蛋白组学是以蛋白质组为研究对象,研究细胞、组织或生物体蛋白质组成及其变化规律的科学。检验组学通过对健康个体及病理个体间的蛋白质组比较分析,可以找到某些“疾病特异性的蛋白质分子”,它们可能成为新药物设计的分子靶点,也为寻找疾病的早期诊断分子标志提供了新方法。

代谢组学是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的研究方式。疾病导致机体病理生理过程变化,最终引起代谢产物发生相应的改变,运用检验组学方法对某些代谢产物进行大数据、高通量分析,并与健康人群的代谢产物对比,可寻找疾病代谢过程中的生物标志物。

检验组学包含了基因组学、蛋白组学、代谢组学、免疫组学及药物基因组学等研究的实验室检查数据结果,并通过对多组学数据的大数据整合分析,发展联合多项指标,可对疾病进行更精确的诊断、治疗及预后。与过去的检验医学研究相比,检验组学的诞生使其他相关组学研究由碎片连接为整体,具有数据容量大、动态性强、准确度高、预后性好等特点。

检验组学对于多组学检验数据处理的帮助主要

体现在两个方面:一是针对大样本群体的某项或多项生理指标进行统计处理分析得到更加精准的病理指标置信区间,以及更加精准的疾病预警指标体系;二是对于个体不同阶段各项检验生理指标的动态处理分析,以建立个体化的生物医学信息库。检验组学通过对大量临床医学检验历史数据进行数据提炼,生成测试数据集与训练数据集。测试数据集用于模型验证并持续训练形成机器学习模型;训练数据集则可直接持续改进为机器学习模型。机器学习模型与应用系统进行信息交互,机器模型复制进入应用系统,应用系统不断改进并反馈给机器模型,从而实现相互促进共同发展,直至能达到临床实践要求。

3 检验组学的临床应用价值

3.1 检验组学有望发现新的疾病预警指标 目前临床检验往往只用了有限的检验指标数据对疾病诊断结果做出判定。然而,很多临床检验数据(例如红细胞均值、血小板均值等),尽管并没有明显超过现有指标的 normal 阈值,但其可能在某个区间内存在上升或下降的趋势。通过大数据分析、云计算等统计学方法,可实现对检验数据与病例症状的逻辑关系分析,从而建立分析方法,可能从容易忽视的数据中挖掘出与疾病相关的新的疾病标志物。检验组学还可将某些常规的指标与多种疾病进行数据对比计算分析,进而发现其数据层面的对应关系,再通过恰当的病理学验证,有望发现新的疾病预警指标。

例如,红细胞体积分布宽度(RDW)是一个描述红细胞体积变异性的血常规指标,以往认为 RDW 仅是一个血液病指标,其临床价值主要在于贫血病因的鉴别。然而,近年来的研究显示,RDW 在很多非血液系统疾病中具有重要的临床价值,例如在普通个体中,RDW 越高,全因死亡风险和心血管疾病发病风险也越高;在恶性肿瘤和心血管疾病患者中,RDW 越高则预后越差;在肝病患者或自身免疫性疾病患者中,RDW 增高提示疾病病情较重或活动度较高。通过检验组学对多种常见指标数据与常见疾病的高通量无差别对比分析,将对疾病预警体系的丰富与完善有极大的促进作用。

3.2 检验组学将精准化疾病早期预警诊断体系 众所周知,生物标志物的检测一直是筛查、诊断某些疾病的重要手段。组学时代,各个组学的飞速发展作为众多疾病生物标志物的发现与发展提供了广阔的前景。

蛋白组学技术是癌症生物标志物和治疗靶点研究的一个重要工具。例如,缺乏早期诊断及有效治疗手段使得肺癌的病死率很高,而早期诊断出肺癌的患者则可大大提升其 5 年生存率。研究表明在肺癌临床研究中,蛋白组学技术总结了大量与肺癌的发病机

制^[11]、转移^[12]、早期诊断^[13-17]及治疗^[18-19]相关的蛋白,如 S100A14、STC2、PRX1、PTX3、ENO1、transgelin-2 等。这些肿瘤标志物的研究可为肺癌的临床研究提供更多的依据。

代谢组学分析在肿瘤标志物筛选中亦具有重要意义。通过对肿瘤患者血清、尿液、胆汁、脑脊液等体液中代谢产物的分析能够发现一些有别于健康人群的异常表达的潜在标志物,其诊断效率显著高于常规的检测项目,并且部分标志物与肿瘤的分期、分化等预后因素密切相关。目前已有学者将代谢组学成功应用于胃癌、前列腺癌、乳腺癌、肠癌^[20-22]等肿瘤标志物识别中,并显示出它们在肿瘤早期诊断和研究中的优势和价值。

然而,单纯研究某一层次生物分子组学(基因组学、蛋白组学或代谢组学等)的变化,即单个生物标志物的变化,其敏感性或特异性往往偏低,现已经很难满足临床对于相关疾病诊断及预后越来越高的研究期望。理论上和实践上都提倡一次同时测定多种标志物,以提高敏感性和特异性。检验组学对于多组学数据进行分析处理正是实现这一设想的有力武器。基于多组学大数据分析的检验组学可以更加完善地检测多种临床疾病生物标志物。通过整合各个组学筛选的生物标志物,并进行合理的检验组学数据处理分析,将大大提升对疾病诊断的灵敏性及特异性,可以更加精准地实现疾病的早期诊断以及预后。

3.3 检验组学与个体化生物医学信息库的建立 正如德国哲学家莱布尼茨所说:“世界上没有完全相同的树叶”,世界上自然不可能存在完全相同的,没有特异性的两个人。不同人的各项医学检验指标亦是如此,甚至于同一个人不同时间的某项检验数据都会存在波动。因此,每个个体不同阶段的各项医学检验指标数据也是在不断波动,不断变化的,数据量不可小觑。

检验组学基于大数据、高通量的数据分析方法,整合了多组学以及临床检验多项指标数据,通过生物信息学数据统计处理软件的逐步研发,指导个性化的治疗,使个体在不同阶段的各项检验指标、数据的统计分析得以实现,进而可以针对不同个体建立个体化生物医学信息库,从而更加精确地对各项指标进行评估,达到对不同个体疾病更加精准化地预防、治疗、诊断。

3.4 检验组学与精准的治疗方案的选择 检验组学的应用还对临床精准治疗方案的选择有一定的指导作用。化疗是目前治疗癌症最有效的手段之一,和手术、放疗一起并称癌症的三大治疗手段。但其治疗过程往往伴随着骨髓抑制、脱发、肝肾功能损害等毒副

作用,使患者十分痛苦。然而,通过运用检验组学方法对大量临床病例、检测指标进行的合理化分析,可对术前化疗的必要性进行有效预测。在这个预测流程中用到了大量临床数据、医疗图像、分子数据等来构建预测模型。在临床实验中,患者可根据预测模型接受治疗。由此可见,检验组学通过大量数据处理并建立治疗预测模型,将在临床精准治疗方案的选择应用方面有重大意义。

4 检验组学面临的挑战

检验组学通过进行大规模、高通量的数据整合运算,实现诊断、治疗的精准化的美好蓝图固然令人向往,但海量数据的收集、整合、储存、共享以及其中涉及的伦理学问题,将是检验组学落实与发展的重大挑战。笔者认为检验组学大数据整合分析的门槛主要体现在以下几点。

4.1 临床大数据应用的专业性 数据本身不直接带来价值,其价值的体现要看最终的应用场景。对于多组学临床检验数据的应用来说,能否使数据专家和医学专家通过数据指标处理实现疾病精准诊断与精准治疗,最终得到的多组学生物标志物的整合诊断结果能否得到临床一线工作者和医药学家的认可,是区分检验组学是否能真正应用于临床的关键点。有志于大数据处理分析应用于医疗诊断方面的团队,除了要团结一批临床工作者外,还需要在团队里配备大数据计算的顶级专家。

4.2 医疗大数据存储的便捷性 组学技术的发展产生了海量数据积累,例如欧洲生物信息研究所(EBI)作为世界最大的生物数据库之一,2012 年储存约 2 PB 基因数据,且该数字以每年翻倍的速度继续增长^[23];另一方面,医学大数据的非结构性使得医学数据的存储和处理较其他领域的研究数据更为复杂。目前还没有综合、经济且安全的大数据存储解决方案,因此海量数据的快速处理和存储就成为重要的课题^[24]。

4.3 疾病数据与患者隐私 未来系统流行病学研究将涉及大量人群暴露疾病和组学信息,因此如何确保这些数据的隐私安全至关重要。已经有遗传学家证明即使是匿名数据库,个人的遗传隐私仍然可能被泄露,个人的遗传隐私无法得到保障可能引发伦理学纠纷。目前的解决方法主要是采取对数据库使用者进行实名登记,并公开使用者信息等监督机制^[23]。

5 检验组学前景展望

在各项组学飞速发展的信息时代,检验组学将传统临床检验信息与基因组学、蛋白组学、代谢组学、免疫组学及药物基因组学等多组学信息联系起来,基于大数据分析技术,对本身具有统计分布属性的检验结果进行数据挖掘,提炼成为有临床价值的重要信息,

以达到精准诊断、精准治疗的要求。

检验组学的建立将人工智能、大数据、高通量计算技术有机融合,通过建立检验多组学数据与病例症状的逻辑分析方法,发现新的疾病预警指标,明确精准的病理指标置信区间;通过收集、整合个体病例不同生理、病理状态下的各项组学与临床检验指标,建立针对每位患者的个体化生物学信息库;通过研究具有按需重构、随遇集成、智能服务、自主学习、持续演进等特征的智能化医学检验数据处理技术,建立医学检验大数据基础架构;通过综合利用各种疾病指标等信息资源,集成临床检验多组学信息系统,使各类疾病数据指标在医学检验中更加精准、高效。在大数据运算技术的支撑下,检验组学的应用研究发展将在检验医学领域中具有革命性意义。

参考文献

- [1] HOOD L, FRIEND S H. Predictive, personalized, preventive, participatory (P4) cancer medicine[J]. Nat Rev Clin Oncol, 2011, 8(3): 184-187.
- [2] 夏锋, 韦邦福. 精准医疗的理念及其技术体系[J]. 医学与哲学, 2010, 31(22): 1-3.
- [3] 苏晓娜. 新型医学概念及医疗模式——“精准医学”[J]. 解放军医药杂志, 2015, 27(5): 2.
- [4] 赵晓宇, 刁天喜, 高云华, 等. 美国“精准医学计划”解读与思考[J]. 军事医学, 2015, 39(4): 241-244.
- [5] 钱其军, 吴孟超. 肿瘤精准细胞免疫治疗: 梦想照进现实[J]. 中国肿瘤生物治疗杂志, 2015, 22(2): 151-158.
- [6] 《中国全科医学》编辑部. 全科医生小词典——精准医疗[J]. 中国全科医学, 2015, 18(24): 2951.
- [7] 李文磊, 杨露琦. 精准医疗时代的兴起[J]. 北京医学, 2015, 37(12): 1204-1206.
- [8] 弓孟春. 精准医学对医学信息学提出的挑战与任务[J]. 医学信息学杂志, 2016, 37(1): 2-7.
- [9] 江海燕, 张峰, 墨宏山. 谁来投入罕见病研究和药物研发: 埃博拉疫情暴发带给我们的警示[J]. 科学通报, 2015, 60(21): 1958-1962.
- [10] 徐鹏辉. 美国启动精准医疗计划[J]. 世界复合医学, 2015, 1(1): 44-46.
- [11] ZHANG X, LI W, HOU Y, et al. Comparative membrane proteomic analysis between lung adenocarcinoma and normal tissue by iTRAQ labeling mass spectrometry[J]. Am J Transl Res, 2014, 6(3): 267-280.
- [12] NA S S, ALDONZA M B, SUNG H J, et al. Stanniocalcin-2 (STC2): A potential lung cancer biomarker promotes lung cancer metastasis and progression[J]. Biochim

Biophys Acta, 2015, 1854(6): 668-676.

- [13] ROSTILA A, PUUSTINEN A, TOLJAMO T, et al. Peroxiredoxins and tropomyosins as plasma biomarkers for lung cancer and asbestos exposure[J]. Lung Cancer, 2012, 77(2): 450-459.
- [14] DIAMANDIS E P, GOODGLICK L, PLANQUE C, et al. Pentraxin-3 is a novel biomarker of lung carcinoma[J]. Clin Cancer Res, 2011, 17(8): 2395-2399.
- [15] YU L, SHEN J, MANNOOR K, et al. Identification of ENO1 as a potential sputum biomarker for early-stage lung cancer by shotgun proteomics[J]. Clin Lung Cancer, 2014, 15(5): 372-378.
- [16] JIN H, CHENG X, PEI Y, et al. Identification and verification of transgelin-2 as a potential biomarker of tumor-derived lung-cancer endothelial cells by comparative proteomics[J]. J Proteomics, 2016, 136(1): 77-88.
- [17] CLARK D J, FONDRIE W E, YANG A, et al. Triple SILAC quantitative proteomic analysis reveals differential abundance of cell signaling proteins between normal and lung cancer-derived exosomes[J]. J Proteomics, 2016, 133(2): 161-169.
- [18] WU Y H, HU C W, CHIEN C W, et al. Quantitative proteomic analysis of human lung tumor xenografts treated with the ectopic ATP synthase inhibitor citreoviridin[J]. PLoS One, 2013, 8(8): e70642.
- [19] WU X, LIANG W, HOU X, et al. Serum proteomic study on EGFR-TKIs target treatment for patients with NSCLC[J]. Onco Targets Ther, 2013, 6(14): 1481-1491.
- [20] AA J, YU L, SUN M, et al. Metabolic features of the tumor microenvironment of gastric cancer and the Link to the systemic macroenvironment[J]. Metabolomics, 2012, 8(1): 164-173.
- [21] TOSOIAN J J, ROSS A E, SOKOLL L J, et al. Urinary biomarkers for prostate cancer[J]. Urol Clin North Am, 2016, 43(1): 17-38.
- [22] KÜHN T, FLOEGEL A, SOOKTHAI D, et al. Higher plasma levels of lysophosphatidylcholine 18 : 0 are related to a lower risk of common cancers in a prospective metabolomics study[J]. BMC Med, 2016, 14(1): 13.
- [23] MARX V. Biology: the big challenges of big data[J]. Nature, 2013, 498(7453): 255-260.
- [24] COSTA F F. Big data in biomedicine[J]. Drug Discov Today, 2014, 19(4): 433-440.

(收稿日期: 2018-03-15 修回日期: 2018-04-10)