

DOI: 10.13288/j.11-2166/r.2016.14.006

思路与方法

全基因组及全表型组关联分析在中医药研究中应用的可能性与前景

周欣, 薛文达, 吴颢昕, 陈刚*

(南京中医药大学转化系统生物学与神经科学研究中心, 江苏省南京市栖霞区仙林大道 138 号, 210023)

[摘要] 中医辨证论治强调对患者进行个体化治疗, 与现代医学所倡导的精准治疗有相通之处。中医体质与证候部分由先天禀赋决定, 即遗传发挥了重要影响。全基因组关联分析可直接将基因变异性与疾病或生理性状相关联, 全表型组关联分析可进一步获得特定的遗传变异位点所影响的不同表型。未来中医药研究中, 可以通过全基因组关联分析和全表型组关联分析从全基因组的层面上来判别影响特定体质、证候与病证结合对证治疗的分子遗传基础, 揭示中医辨证论治的现代生物学本质, 为提高中医诊断与治疗的精准性与稳定性提供参考。

[关键词] 全基因组关联分析; 全表型组关联分析; 精准医学; 中医证候; 中医体质; 基因表型

精准医学是现代医学的一个重要发展方向, 是用以改进目前对于复杂疾病个体化差异下的医疗缺乏个性化而采用的针对性的医疗方式^[1]。精准医学是个性化医学理念的新发展, 是基于基因组学技术革命性的突破并将其应用于医疗实践中。现代医学力图实现个性化, 其路径是根据每个人基因型的差异性进行疾病病因的判断并选择性的使用药物进行治疗, 如肿瘤基因诊断与分子靶向治疗。这种根据基因与遗传信息进行精准医疗的转变过程可能是长期的, 也具有一定的不确定性。中医学是典型的个性化医学, 根据患者所处的证候阶段和禀赋体质给予不同的治疗。中医学在临床实践中实现了个性化, 但人们对这种个性化医疗与基于基因组变异而提出的个性化医学之间的深层次联系并不清楚。患者基因的变异性是否导致同一疾病下证候表现的差异性, 辨证论治是否具有内在自然科学依据, 基因组的变异性是否为辨证论治的一个重要依据, 这些问题可以从全基因组关联分析 (genome-wide association study, GWAS) 与全表型组关联分析 (phenome-wide association study, PheWAS) 研究中获得启示。

基金项目: 江苏省自然科学基金 (BK20141565, BK20151568); 江苏省高层次创新创业人才引进计划

* 通讯作者: chengang@njutcm.edu.cn

1 GWAS 研究的形成与发展

人类基因组计划的完成^[2], 大量疾病相关基因的发现, 促进了传统生物医学模式向可预测性、可预防性、个体化和参与性的基因组医学模式转变, 为未来预防、诊断、治疗长期困扰人类的由遗传和环境因素共同作用引起的重大复杂性疾病开辟了新途径, 也为基因科学的产业化提供了良好机遇。随着人类基因组计划和基因组单倍体图谱计划的实施, 研究者开始对影响人类性状形成和复杂性疾病产生的遗传特征进行探索, 并在此基础上成功建立了人类基因组 DNA 遗传变异数据库^[3], 为相关疾病的潜在影响基因关联性研究奠定了基础, 使得人类复杂疾病的 GWAS 得以开展。

遗传对疾病的影响是广泛的, 复杂疾病并不是由某个单一的致病基因所引起, 而是受到包括基因多态性、等位基因异质性、基因-基因相互作用等多重遗传因素以及与环境因素的共同作用和 (或) 相互作用的结果, 而这些基因对常见疾病的遗传力可能只有很小的贡献, 这使得鉴定那些与常见疾病相关的基因成为难点和热点^[4]。GWAS 不同于研究疾病单个基因点突变位点的家系基因连锁分析, 其应用基因组中数以百万计的单核苷酸多态性 (single nucleotide polymorphism, SNP) 作为分子遗传标记,

进行全基因组水平上常见遗传变异与所研究疾病的对照分析或相关性分析,以此来发现可能影响复杂疾病的相关基因变异^[5]。基于人类基因组测序的完成和高通量基因分型技术的发展,在全基因组水平上对复杂疾病进行全面基因关联分析,尤其是多中心、大样本、反复验证的关联研究,可以全面揭示疾病的发生、发展及与治疗相关的基因及网络,2005 年,Science 杂志首次报道了年龄相关性视网膜黄斑变性 GWAS 结果^[6],此后一系列的 GWAS 研究相继被报道,如 2006 年波士顿大学医学院等多个研究机构报道了肥胖的 GWAS 结果^[7],2007 年 Saxena 等多个研究组联合报道了与 2 型糖尿病关联的多个位点^[8]、Samani 等发表了冠心病 GWAS 结果^[9],2008 年 Barrett 等通过 GWAS 发现了 30 个与克罗恩病 (Crohn's disease) 相关的易感位点^[10],2009 年 Weiss 等通过 GWAS 发现了与具有高度遗传性的神经发育疾病——自闭症关联的染色体区域^[11]。我国学者则通过对 12 000 多例汉族系统性红斑狼疮患者以及健康对照者的 GWAS 研究发现了 5 个红斑狼疮易感基因,确定了 4 个新的易感位点^[12]。截至 2015 年 2 月,已发现了 15 000 多个关联的 SNP (<http://www.genome.gov/gwastudies>)。但到目前为止,通过检索 Web of Science 和中国知网,没有获得任何关于中医药的 GWAS 研究报道。

2 GWAS 在中医相关表型研究中应用的可能性与优势

中医学的特色和优势在于辨证论治,实质上就是对于复杂疾病的全面把握和对患者的个体化治疗。中医证候及复方的本质研究作为中医药现代化的关键和最具有标志性的方面,仍面临着非常大的困难^[13]。证候在不同体质人群中的表现具有差异性,而这种体质上的差异又是由于先天禀赋即遗传背景与后天环境共同作用的结果^[14],是遗传物质在生物整体层面的表现。这种表现可以如任何可测量的物理或生物化学特征一样进行分析,即可以作为表型。证候不同,遣方用药就会相应地产生各种差异,包括方药组成、剂量等。利用 GWAS 研究方法,将中医证候作为特定的表型,从全基因组层面上来鉴别与说明影响特定证的分子遗传基础,可以帮助我们有效地揭示中医证候的生物学本质。例如在抑郁症的中医证候研究中,可以选取有代表性的证候类型如肝郁脾虚证、肝郁化火证、心脾两虚

证等^[15-16]进行多中心、大样本的研究,将证候类型与全基因组关联,有助于发现其影响特定证候的易感位点或基因,从而在本质上揭示抑郁症中医证候的生物学机制。

转录组学、蛋白组学、代谢组学、表观遗传组学等与基因组学一样,可以在一定程度上通过发现分子网络系统变化来说明中医药的系统作用机理,但其应用都需要落实到具体的器官与组织之中。由于许多导致证候的中医表型的组织、器官、细胞等并不清楚,很多情况下是多系统、多组织的,只研究某一个或几个组织、器官的变化来说明中医表型会导致错失发现真正作用物质基础的可能性。GWAS 研究的鲜明特点在于其不需要在研究之前构建任何假设,即不需要预先假设某些特定的基因或位点,也无需假设作用的生理解剖机制^[17],而仅仅通过关联研究就可以发现大量与疾病相关的基因,所以 GWAS 对中医相关表型如证候、复方等作用机制不明表型的研究具有独特优势。

3 PheWAS 对 GWAS 研究的补充

GWAS 研究仅关注于预先定义和有限的表型。从已有的研究模式中,基于多重比较测试的要求,GWAS 的显著性被定义为 $P < 10^{-8}$ ^[18-19],能达到此显著水平的 SNP 数量比较少,且对表型的贡献都比较微弱,能够解释的遗传率也非常少,远远达不到双胞胎家系研究所获得的水平,这一现象被称为遗传性缺失现象。而疾病表现性状与基因组之间的关系受多种因素的影响,这也导致了有些 GWAS 研究所获得的结果难以解释某些病因。此外,即使 GWAS 分析出显著的 SNP, GWAS 本身也不能鉴别该 SNP 所能影响的临床相关性状。严格地说, GWAS 分析出的 SNP 是与性状关联的“标记”,大量 GWAS 获得的显著的 SNP 都是基因间的,基因内的相对较少,但是鉴别这些基因间序列的功能目前还非常困难。针对这一情况,一方面需要改善 GWAS 的分析方法,包括更多地考虑各种基因-基因相互作用和基因-环境相互作用的研究模型,以及加强基因组中非编码基因功能的认识;另一方面,还可以将 GWAS 的研究结果进行进一步利用,扩大其应用成果, PheWAS 就是其中的代表。

GWAS 是固定研究某个表型,将其变异性与全基因组水平的 SNP 变异性相关联;而 PheWAS 则采用与 GWAS 相反的过程,针对若干个 SNP,从成百上千的表型的变化中寻找与这些 SNP 关联的

一个或几个表型^[20]。当然 SNP 通常已经在 GWAS 研究中揭示出与某些表型有显著关联性, 而 PheWAS 研究一方面可以检验该 SNP 与已经报道的表型的关联, 另一方面还可以研究该 SNP 对应的多功能性, 即对其他表型的作用。GWAS 研究中表型的选择是成功的关键, 相应地对 PheWAS 来说, 合适的 SNP 标记物的选择亦尤为重要。PheWAS 研究策略首次建立于 2010 年^[21], 其在医学的研究上主要利用电子病历来定义表型组。现有的电子病历中包含了基于《国际疾病分类标准》的一千多种疾病类型记录, 包括疾病、体征和症状。当然 PheWAS 也存在很多类似于 GWAS 的局限性, 如基于电子病历的 PheWAS 研究只能够研究来自于临床研究收集的表型数据^[21-22], 而最大的挑战是大量研究所积累的表型数据的有效性以及如何保证高质量的系统表型测定。尽管如此, 越来越多的研究证明 PheWAS 有能力鉴别出与 GWAS 对应的关联性。

4 GWAS 和 PheWAS 在中医药研究中的挑战与前景

考虑到目前中医临床的标准化与规范化还有待进一步提高, 开展多中心、大样本 GWAS 研究可能会因中医证候表型的一致性、复方药材的差异性等问题而产生困难。同时不同中医师对同一患者的脉象、舌苔等可有不同诊断, 以及在同样或类似的证候要素基础上, 因为中医师的主观性产生的证候分型不同, 同时部分证候随时间而发生明显变化。此外, GWAS 本身需要足够大的样本, 但由于证候类型越多而能够获得的人群数量就相应地减少, 增加了研究的难度与成本。PheWAS 的中医研究也面临相似问题。但是比较而言中医学更注重各种证候表现, 因而会产生更多表型数据, 在获得了与特定的中医表型关联的 SNP 后, 可以使用 PheWAS 对中医的证素、症状、生理指标等进行研究, 以揭示该 SNP 与其他中医表型可能的关联性, 进而揭示出各种中医表型的相关性。这需要有完善的患者电子病历, 其中要包含详尽的中医诊断数据, 这方面许多中医院还有待建立与完善。应该指出, 这些问题已经越来越多地得到了解决, 可以从研究基础较好、一致性强、相对易于操作的中医体质、证候或复方着手进行研究, 鉴别其作用的分子遗传学和分子网络基础。尽管到目前为止还没有 GWAS 和 PheWAS 在中医药研究方面的应用报道, 但是其有良好的应用前景。例如, 目前已经有各种中医体质的数据, 对受试者进行基因组分型就可以进行

GWAS 分析, 从而获得决定特定体质的特异性分子遗传证据, 说明体质差异的客观遗传基础, 以揭示中医理论的深刻内涵, 同时也可以反过来作为生物标记物帮助确定个体的体质类型, 因而在临床上极具应用潜力。相应地, PheWAS 方法也有揭示病证之间相互关联的分子证据等潜力, 值得深入发掘。

5 小结

证候是中医特色的辨证论治个体化治疗的基础, 证候与体质部分地由遗传决定, GWAS 和 PheWAS 为鉴定中医证候、体质与复方的遗传物质基础提供了新的方法与手段。此外, 还可以利用模式动物及合适的遗传种群对中医的重要表型进行 GWAS 和 PheWAS 分析, 结合各种组学数据, 获得确定的关于中医表型分子组构的系统生物学的知识, 并予以验证。这些人群与动物研究都将深入推进对中医精准治疗疾病的科学内涵的认识。随着精准医学时代的到来, 个体全基因组信息读取成本的降低, 这些信息也可以为中医临床所用, 将其与辨证论治进行结合, 以推动中医临床向更精准、更有预测性、更可靠的方向迈进。

参考文献

- [1] COLLINS FS, VARMUS H. A new initiative on precision medicine [J]. N Engl J Med, 2015, 372(9): 793-795.
- [2] BALTIMORE D. Our genome unveiled [J]. Nature, 2001, 409(6822): 814-816.
- [3] 武雪梅, 肖华胜. 人类基因组结构变异检测研究进展 [J]. 中国科学 (C 辑: 生命科学), 2009, 39(3): 237-244.
- [4] 王莹, 王艺, 黄薇. 全基因组关联分析在人类复杂疾病研究中的应用 [J]. 诊断理论与实践, 2010, 9(5): 420-432.
- [5] LU YF, GOLDSTEIN DB, ANGRIST M, et al. Personalized medicine and human genetic diversity [J]. Cold Spring Harb Perspect Med, 2014, 4(9): a008581. doi: 10.1101/cshperspect.a008581.
- [6] KLEIN RJ, ZEISS C, CHEW EY, et al. Complement factor H polymorphism in age-related macular degeneration [J]. Science, 2005, 308(5720): 385-389.
- [7] HERBERT A, GERRY NP, MCQUEEN MB, et al. A common genetic variant is associated with adult and childhood obesity [J]. Science, 2006, 312(5771): 279-283.
- [8] SAXENA R, VOIGHT BF, LYSENKO V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels [J]. Science, 2007, 316(5829): 1331-1336.
- [9] SAMANI NJ, ERDMANN J, HALL AS, et al. Genome-

- wide association analysis of coronary artery disease [J]. *N Engl J Med*, 2007, 357(5): 443-453.
- [10] BARRETT JC, HANSOUL S, NICOLAE DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease [J]. *Nat Genet*, 2008, 40(8): 955-962.
- [11] WEISS LA, ARKING DE, DALY MJ, et al. A genome-wide linkage and association scan reveals novel loci for autism [J]. *Nature*, 2009, 461(7265): 802-808.
- [12] HAN JW, ZHENG HF, CUI Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus [J]. *Nat Genet*, 2009, 41(11): 1234-1237.
- [13] 江峰, 王宗殿. 反思中医证候本质研究 [N]. *中国医药报*, 2006-01-07(07).
- [14] 刘向哲, 王新志, 王永炎. 试论禀赋与体质的关系 [J]. *北京中医药大学学报*, 2011, 34(7): 441-443.
- [15] 唐启盛, 曲淼, 包祖晓, 等. 抑郁症中医证候的贝叶斯网络研究 [J]. *中医杂志*, 2008, 49(11): 1013-1015.
- [16] 李涛, 梅妍, 张鑫, 等. 关于抑郁症辨证分型及常见症状的第 1 轮专家问卷分析 [J]. *中医药导报*, 2011, 17(11): 1-6.
- [17] TODD JA. Statistical false positive or true disease pathway? [J]. *Nat Genet*, 2006, 38(7): 731-733.
- [18] MCCARTHY MI, ABECASIS GR, CARDON LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges [J]. *Nat Rev Genet*, 2008, 9(5): 356-369.
- [19] RISCH N, MERIKANGAS K. The future of genetic studies of complex human diseases [J]. *Science*, 1996, 273(5281): 1516-1517.
- [20] PENDERGRASS SA, BROWN-GENTRY K, DUDEK SM, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery [J]. *Genet Epidemiol*, 2011, 35(5): 410-422.
- [21] DENNY JC, RITCHIE MD, BASFORD MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations [J]. *Bioinformatics*, 2010, 26(9): 1205-1210.
- [22] MCCARTY CA, CHISHOLM RL, CHUTE CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies [J]. *BMC Med Genomics*, 2011, 4(13): 1-11.

Possibility and Prospect of Application of Association Study on Genomewide and Phenome-wide in Chinese Medicine Researches

ZHOU Xin, XUE Wenda, WU Haoxin, CHEN Gang

(Center for Transformation System Biology and Neurosciences Research, Nanjing University of Chinese Medicine, Nanjing, 210023)

ABSTRACT Chinese medicine syndrome differentiation treatment emphasizes individualized treatment on patients, which is similar to precision treatment proposed by modern medicine. Chinese medicine constitution and syndrome is determined partly by natural endowment, which means that inheritance plays an important role. Genomewide association study could correlate gene mutation with disease or physiological characteristics directly. Phenomewide association study could further find out different phenotypes influenced by specific genetic variation site. In future Chinese medicine researches, molecular genetic basis, which is influenced by certain constitution, syndrome and disease-syndrome combination syndrome treatment, might be differentiated from genomewide aspect by genomewide association study and phenomewide association study. The essence of modern biology of Chinese medicine syndrome differentiation treatment might be revealed. This paper might provide reference for improving the precision and stability of Chinese medicine diagnosis and treatment.

Keywords genomewide association study; phenomewide association study; precision medicine; Chinese medicine syndrome; Chinese medicine constitution; gene phenotype

(收稿日期: 2015-11-25; 修回日期: 2016-03-10)

[编辑: 焦 爽]