

多基因风险评分在冠心病风险评估中的研究进展

邹发龙¹ 王学懂¹

[摘要] 冠心病的发病受多种环境与遗传因素的影响。多基因风险评分作为研究遗传易感性与复杂疾病之间关系的新兴方法,通过整合多种变异位点的累积效应,实现高危患者的早期风险评估,从而有助于冠心病等复杂疾病的精准预测和防控。本文就冠心病多基因风险评分的研究进展作一综述。

[关键词] 冠心病;风险评估;多基因风险评分

DOI:10.13201/j.issn.1001-1439.2023.03.006

[中图分类号] R541.4 **[文献标志码]** A

Research progress of polygenic risk score in the risk assessment of coronary heart disease

ZOU Falong WANG Xuedong

(Department of Cardiology, Second Affiliated Hospital of Harbin Medical University, Harbin, 150086, China)

Corresponding author: WANG Xuedong, E-mail: doctor_wang@vip.126.com

Abstract The incidence of coronary heart disease is affected by many environmental and genetic factors. The newly emerged polygenic risk score can make the early assessment of high-risk patients by investigating the relationship between genetic susceptibility and complex diseases, and integrating the cumulative effect of multiple variant loci. This scoring method can realize the accurate prediction and prevention of complex disease such as coronary heart disease. This article reviews the research progress of polygenic risk score for coronary heart disease.

Key words coronary heart disease; risk assessment; polygenic risk score

冠心病(coronary heart disease, CHD)是一种由于冠状动脉(冠脉)粥样硬化导致心肌缺血缺氧甚至坏死而严重威胁人类生命健康的常见心血管疾病。据 *J Am Coll Cardiol* 统计报告,自1990年以来CHD所造成的人口死亡数量始终保持增高趋势,是人类公共卫生的主要威胁^[1]。CHD的发病机制复杂,主要与遗传以及生活方式等因素有关,且研究表明两种因素与CHD的易感性独立相关^[2]。受限于生物技术与遗传学的发展水平,过去对于CHD的风险评估主要基于生活方式等传统危险因素。然而基于传统危险因素构建的风险模型对心血管事件的预测效应已难以满足现代医学的需求^[3]。近年来,随着基因技术的进步,多基因风险评分(polygenic risk score, PRS)模型被成功构建,且已被证实其对CHD的风险预测与分层具有良好的效应^[2,4-5]。但由于构建PRS模型所采用的基因数据常受数量与种族等因素限制,不同研究所得结果也不尽相同,故仍需进一步的探索与研究^[6-7]。本文旨在对近年来PRS用于CHD风险预

测的研究进展作一综述。

1 CHD风险评估的历史回顾

CHD作为威胁人类生命健康的重大疾病之一,如何实现对高危人群的早期和精准预测始终是心血管领域亟需解决的关键问题。目前已有相关研究取得重大突破。其中,美国的Framingham心脏研究(Framingham Heart Study, FHS)在CHD的病因及预测方面做出了最为显著的贡献,该研究不仅开创性地提出高血压、血脂水平、超重等是CHD发展的危险因素,还首次强调了父母患CHD作为后代疾病风险的独立危险因素,提示遗传因素可能在CHD的发病中发挥作用,开拓并奠定了早期CHD风险评估的研究方向^[8]。近代以来,随着人类遗传学的进步以及相关技术的发展,研究者们开始深入研究遗传因素对人类疾病的影响,以帮助人类更好地对疾病进行预测与管理。其中,通过全基因组相关研究(genome-wide association study, GWAS)以及单核苷酸多态性(single nucleotide polymorphism, SNP)构建的PRS风险预测模型是最有效的方法之一。该方法将众多遗传位点对某种性状的累积效应计算为定量指标,从而对基因风险进行量化。2005年,Horner等^[9]进行了一项研

¹哈尔滨医科大学附属第二医院心血管内科(哈尔滨, 150086)

通信作者:王学懂, E-mail: doctor_wang@vip.126.com

究,首次应用 PRS 对 CHD 进行风险预测,结果显示不同 PRS 组间 CHD 风险具有显著差异,体现出 PRS 对复杂疾病风险预测的潜力。随后 PRS 风险模型被广泛应用于 CHD 的风险评估并取得良好的效果。

2 多基因风险评分

2.1 PRS 概述

随着分子生物学的发展、基因组项目的结合以及强大基因分型平台的出现,大规模队列的 GWAS 得以实现。GWAS 专注于识别特定人群中常见疾病或性状相关的遗传变异,典型为 SNP,目前已发现众多位点与 CHD 有关^[10]。然而,这些变异中大多数仅具有很小的影响,对应于相关变异中的小部分,因此单独或少数的变异预测能力有限^[11]。PRS 作为唯一在个体水平上提供对性状遗传估计的方法,通过汇总基因组中变异的影响,估计遗传倾向并根据遗传特征对表型进行预测,其分析过程简述如下。首先,输入 Base(GWAS) data(基础数据)和 Target data(靶标数据)这两个数据集,前者是全基因组范围内遗传变异的基因型-表型关联的摘要统计信息(如 β 、 P 值),后者是目标样本中个体的基因型和表型。其次,对基础数据进行相应处理,随后与靶标数据联合产生 PRS 结果。最后通过分析结果对研究人群进行风险预测与治疗指导。这对于某些遗传因素影响发展的复杂性疾病如 CHD 进行更早更准确的风险预测具有重大价值。

2.2 PRS 的构建方法

PRS 的构建大体可分为 4 步。第 1 步:通过 GWAS 收集基础数据。第 2 步:对所得的基础数据进行相应处理以保证数据质量,其中选择的 P 阈值决定纳入 PRS 模型的 SNP 数量,会在很大程度上影响 PRS 风险预测模型的敏感性与特异性;另一个需要处理的因素是 SNP 间的相关性问题,若不对其进行削弱则会导致某些遗传变异的过度体现,影响 PRS 的评估。目前有两种方法对 SNP 间的相关性问题进行处理:一种为 clumping,也就是识别并选择一组相关 SNP 中与结果相关度更高(即 P 值更低)的 SNP,移除另一个,这种方式在减少 SNP 之间关联性的同时,能够保留更有价值的 SNP 数据以进行进一步分析;另一种为 shrinkage,这种方法允许纳入相关的 SNP,但要根据其相关性结构缩小其效应估计值。前一种方法与 P 阈值联合使用对 SNP 数据进行调整,而后一种方法则与全基因组 PRS 相关。第 3 步,将经第 2 步处理后的数据与靶标数据结合,生成 PRS 数据,并执行关联测试,从而确定最佳 P 值以接近最佳拟合的 PRS。第 4 步,测试最佳阈值并可视化 PRS 结果,报告相应指标。

2.3 PRS 的类型与计算

根据是否对 SNP 进行加权处理以及构建过程中对数据的处理方式,PRS 可分为 4 种。第 1 种为 Unweighted PRS,此类型将个体基因组中已知与相关疾病显著关联的 SNP 及其风险等位基因直接相加,统计其在人群中的数量分布情况。由于大多数复杂疾病其遗传结构中变异效应不等,因此该类型应用价值十分有限^[11]。第 2 种为 Weighted, limited clumping & thresholding PRS,该类型对常见变异风险等位基因进行加权处理,其中对 SNP 数据进行处理时采用了严格的 P 值阈值,该类型严格控制纳入 PRS 风险模型的 SNP 数量及其相关性,对 PRS 模型预测的特异性有益,在部分研究中有应用^[2]。第 3 种为 Weighted, expanded clumping & thresholding PRS,该类型同第 2 种对变异风险等位基因进行加权处理,但其 P 阈值设定则相对第 2 种更宽松,该类型 PRS 风险预测模型可纳入更多 SNP 数据。第 4 种,Weighted, genome-wide PRS,该类型同样对风险等位基因进行加权,但其却再次放宽乃至不存在 P 阈值,而是通过 Shrinkage 方式,根据不同风险等位基因相关性的程度对数据进行调整,该类型可容纳最多数量的 SNP 数据,运用 GWAS 所得全部 SNP 数据,有利于在更大范围群体中应用,该种类型目前应用最广泛^[12]。

在最简单同时也是最常见的形式中,PRS 是 m 个 SNP 效应的总和,基于 GWAS 汇总统计得到的 SNP 效应大小 β ,对数百至数千个 SNPs 进行汇总,从而将个人基因组的贡献转化为与特定疾病风险成正比的数字,其中 X_{ij} 是指第 i 个个体的第 j 个 SNP 基因型。

$$PRS_i = \sum_{j=1}^m X_{ij} \hat{\beta}_j$$

3 PRS 在 CHD 风险评估中的应用现状

3.1 与传统风险预测方式的结合与对比

随着 PRS 用于 CHD 风险预测的价值被证实,部分学者对其与传统风险预测联合的价值进行研究。其中 Inouye 等^[12]通过结合传统风险因素与 PRS 进行风险预测,肯定了二者联合应用对 CHD 风险预测的潜力,并证实了 PRS 对于 CHD 风险预测的优越性。而为量化将 PRS 添加到常规危险因素中对 CHD 风险预测的效益,Sun 等^[13]进行研究并证实,在常规危险因素中添加 PRS 提高了对 CHD 事件的预测能力。基于传统危险因素的临床指南作为指导临床治疗的主要依据,PRS 能否较其拥有更大的预测能力从而使更大范围群体获益具有关键意义。Aragam 等^[14]对此进行研究,并表明目前临床指南较 PRS 存在局限性,联合 PRS 将有助于识别更多预防性治疗的候选个体。除上述试

验,仍有诸多研究证实 PRS 较传统风险因素预测更优越,且两者结合可改善对 CHD 的风险预测,有助于 CHD 实现精准预防^[5,15]。

3.2 在不同人群的应用

基于基因组数据的 PRS 技术对种族背景高度敏感,而早期 GWAS 大多开展于欧洲人群,因此 PRS 对非欧洲人群的风险预测能力受到显著限制^[16-18]。倘若用基于某一人群的 GWAS 数据构建的 PRS 对另一人群进行预测时,可能导致严重的错误^[19]。因此,研究者们通过构建种群特异性的 PRS 模型验证其在不同种群的应用价值。目前已有多项研究,如针对中国人群、南亚人群、日本人群构建的种群特异性 PRS 模型均在对各自种群的 CHD 风险预测中取得良好效果^[20-22]。综上所述,对于变异频率、效应大小以及连锁不平衡模式等导致的 PRS 种群限制,可通过构建种群特异性的 PRS 模型提高其在不同种群中的应用价值,从而实现更广泛人群 CHD 的风险预测与预防管理。

3.3 遗传位点与机制研究

自 2007 年从染色体 9P21 区域发现第 1 批与 CHD 具有全基因组意义的常见遗传变异后,通过 GWAS 不断有新的变异位点被发现,但其中大多数变异位点增加 CHD 风险的机制处于未知状态^[23-24]。通过将 PRS 与疾病发病机制相关联从而指导治疗将具有重大意义。Natarajan 等^[25]研究发现,PRS 更高的人群可从他汀类药物治疗中获得更大的益处,并通过进一步研究证实了高 PRS 人群具有更大的亚临床动脉粥样硬化负担。Larsson 等^[26]将血清钙水平与 PRS 相关联,结果表明导致血清钙水平较高的遗传变异与 CHD 风险增加有关,提示钙负荷增高与 CHD 发病相关。上述研究有助于了解 CHD 的发病机制从而指导相应患者的治疗。

新位点的发现有助于指导新型靶向药物的研发以及构建更加准确的 PRS 风险预测模型,提高 PRS 对疾病的风险预测能力。Hartiala 等^[27]经研究发现,在冠状动脉疾病中,染色体 1P21.3 位点处的 SLC44A3 这一新位点与心肌梗死有关,而与冠状动脉粥样硬化无关。该研究明确了存在与冠状动脉粥样硬化无关的 CHD 风险位点,对研发新型靶向药物以及实现精准医疗具有重要意义。另有众多研究对易感位点影响 CHD 的机制进行研究,目前已知的机制主要包括促进内皮功能障碍与动脉粥样硬化、凝血与炎症过程、血管平滑肌细胞表型、冠状动脉钙化等,相应遗传位点有助于提高 PRS 的预测能力^[28-31]。对 CHD 相关遗传位点与机制的研究对提高 PRS 预测能力以及 CHD 新型药物的研发具有巨大潜力。

4 总结与展望

基于传统危险因素预测方式的预测方式已难以满足现代医学的需求。而通过 GWAS 构建的 PRS 能够更早更精准地对人群 CHD 风险进行预测与分层,且对传统危险因素预测构成补充,有效降低 CHD 对人类生命健康的威胁,促进精准医疗的进步。PRS 目前仍存在限制与不足。性别这一因素对人类复杂疾病的影响早已被证实,但目前仍未有研究构建性别相关性的 PRS 以降低性别因素对 CHD 风险预测造成的偏倚^[32]。对于 CHD 相关遗传位点以及相应机制的认知目前仍十分有限,对新位点以及机制的研究有着巨大潜力。相信随着研究的进展,PRS 的预测能力以及对治疗的指导效应将进一步提高,对人类健康的改善发挥巨大作用。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: Update From the GBD 2019 Study[J]. J Am Coll Cardiol, 2020, 76(25):2982-3021.
- [2] Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease[J]. N Engl J Med, 2016, 375(24):2349-2358.
- [3] Dalton JE, Rothberg MB, Dawson NV, et al. Failure of traditional risk factors to adequately predict cardiovascular events in older populations[J]. J Am Geriatr Soc, 2020, 68(4):754-761.
- [4] Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations[J]. Nat Genet, 2018, 50(9):1219-1224.
- [5] Elliott J, Bodinier B, Bond TA, et al. Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease[J]. JAMA, 2020, 323(7):636-645.
- [6] Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores may exacerbate health disparities[J]. Nat Genet, 2019, 51(4):584-591.
- [7] Mosley JD, Gupta DK, Tan J, et al. Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease[J]. JAMA, 2020, 323(7):627-635.
- [8] Andersson C, Naylor M, Tsao CW, et al. Framingham Heart Study: JACC Focus Seminar, 1/8[J]. J Am Coll Cardiol, 2021, 77(21):2680-2692.
- [9] Horne BD, Anderson JL, Carlquist JF, et al. Generating genetic risk scores from intermediate phenotypes for use in association studies of clinically significant endpoints[J]. Ann Hum Genet, 2005, 69(Pt 2):176-186.
- [10] Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary

- statistics 2019[J]. *Nucleic Acids Res*, 2019, 47(D1): D1005-D1012.
- [11] Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases[J]. *Nature*, 2009, 461(7265):747-753.
- [12] Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention[J]. *J Am Coll Cardiol*, 2018, 72(16):1883-1893.
- [13] Sun L, Pennells L, Kaptoge S, et al. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses[J]. *PLoS Med*, 2021, 18(1):e1003498.
- [14] Aragam KG, Dobbyn A, Judy R, et al. Limitations of contemporary guidelines for managing patients at high genetic risk of coronary artery disease[J]. *J Am Coll Cardiol*, 2020, 75(22):2769-2780.
- [15] Natarajan P. Polygenic risk scoring for coronary heart disease: the first risk factor[J]. *J Am Coll Cardiol*, 2018, 72(16):1894-1897.
- [16] Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations [J]. *Am J Hum Genet*, 2017, 100(4):635-649.
- [17] Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations [J]. *Nat Commun*, 2019, 10(1):3328.
- [18] Gurdasani D, Barroso I, Zeggini E, et al. Genomics of disease risk in globally diverse populations [J]. *Nat Rev Genet*, 2019, 20(9):520-535.
- [19] Kim MS, Patel KP, Teng AK, et al. Genetic disease risks can be misestimated across global populations [J]. *Genome Biol*, 2018, 19(1):179.
- [20] Lu X, Liu Z, Cui Q, et al. A polygenic risk score improves risk stratification of coronary artery disease: a large-scale prospective Chinese cohort study[J]. *Eur Heart J*, 2022, 43(18):1702-1711.
- [21] Wang M, Menon R, Mishra S, et al. Validation of a genome-wide polygenic score for coronary artery disease in south asians[J]. *J Am Coll Cardiol*, 2020, 76(6):703-714.
- [22] Koyama S, Ito K, Terao C, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease[J]. *Nat Genet*, 2020, 52(11):1169-1177.
- [23] Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease[J]. *N Engl J Med*, 2007, 357(5):443-453.
- [24] Erdmann J, Kessler T, Munoz Venegas L, et al. A decade of genome-wide association studies for coronary artery disease: the challenges ahead [J]. *Cardiovasc Res*, 2018, 114(9):1241-1257.
- [25] Natarajan P, Young R, Stitzel NO, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting[J]. *Circulation*, 2017, 135(22):2091-2101.
- [26] Larsson SC, Burgess S, Michaëlsson K. Association of genetic variants related to serum calcium levels with coronary artery disease and myocardial infarction[J]. *JAMA*, 2017, 318(4):371-380.
- [27] Hartiala JA, Han Y, Jia Q, et al. Genome-wide analysis identifies novel susceptibility loci for myocardial infarction[J]. *Eur Heart J*, 2021, 42(9):919-933.
- [28] Xu S, Xu Y, Liu P, et al. The novel coronary artery disease risk gene JCAD/KIAA1462 promotes endothelial dysfunction and atherosclerosis[J]. *Eur Heart J*, 2019, 40(29):2398-2408.
- [29] Howson J, Zhao W, Barnes DR, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms [J]. *Nat Genet*, 2017, 49(7):1113-1119.
- [30] Aherrahrou R, Guo L, Nagraj VP, et al. Genetic regulation of atherosclerosis-relevant phenotypes in human vascular smooth muscle cells[J]. *Circ Res*, 2020, 127(12):1552-1565.
- [31] 刘林鑫, 孟华, 杨冠旭, 等. HMGB1 及 HMGB2 与冠状动脉钙化相关性的分析[J]. *临床心血管病杂志*, 2021, 37(12):1085-1089.
- [32] Khramtsova EA, Davis LK, Stranger BE. The role of sex in the genomics of human complex traits[J]. *Nat Rev Genet*, 2019, 20(3):173-190.

(收稿日期:2022-04-02)