

多基因风险评分与复杂性疾病风险预测和精准预防：机遇和挑战

杭栋 沈洪兵

南京医科大学公共卫生学院流行病学系 211166

通信作者：沈洪兵，Email:hbshen@njmu.edu.cn

【摘要】 随着人类基因组学研究的飞跃发展，全基因组关联研究已成功鉴定出大量复杂性疾病的易感位点。多基因风险评分通过综合多个易感位点信息预测疾病风险，在人群风险预测、优化筛查方案、实施精准预防等方面初显应用价值。本研究就多基因风险评分的最新进展作一概述，总结其应用机遇和挑战。

【关键词】 多基因风险评分；全基因组关联研究；复杂性疾病；精准预防

DOI:10.3760/cma.j.issn.0254-6450.2019.09.001

Application of polygenic risk scores in risk prediction and precision prevention of complex diseases: opportunities and challenges

Hang Dong, Shen Hongbing

Department of Epidemiology, School of Public Health, Nanjing Medical University, Nanjing 211166, China

Corresponding author: Shen Hongbing, Email: hbshen@njmu.edu.cn

【Abstract】 Along with the rapid progress in the field of human genomics, genome-wide association studies have successfully identified numerous risk loci for complex diseases. Polygenic risk scores can predict disease risk by integrating the effects of multiple susceptibility loci, and begin to show good performance for improving risk prediction, screening strategy and precision prevention. This paper briefly reviews the recent progress of polygenic risk scores in disease prevention, and summarizes the opportunities and challenges of its application.

【Key words】 Polygenic risk score; Genome-wide association study; Complex diseases; Precision prevention

DOI:10.3760/cma.j.issn.0254-6450.2019.09.001

目前关于心血管疾病、恶性肿瘤等复杂性疾病的预防建议和防控实践，虽然强调高危人群策略，但关于高危人群的定义以及不同危险度的人群分层，仍存在诸多的不一致性，可能导致无法达到预防措施的真正效果，并增加防控实践的成本。因此，公共卫生的一项关键任务是识别疾病的真正高危人群，从而促进精准预防和个体化干预，达到降低疾病发病率和死亡率的目的。由于大多数复杂性疾病与遗传因素密切相关，一个重要的策略是在环境危险因素的基础上进一步根据遗传标志物对人群进行发病风险预测，以实现个体化的精准预防。

一、多基因风险评分应用于复杂性疾病精准预防的理论基础

近 20 年来，伴随着基因组学研究技术的飞跃发展，全基因组关联研究 (genome-wide association

study, GWAS) 作为分子流行病学研究的重要方法，在复杂性疾病的遗传因素研究中取得了巨大成功。据美国国家人类基因组研究所和欧洲生物信息研究所 (The NHGRI-EBI GWAS Catalog) 统计，目前 GWAS 已经应用于 200 余种疾病和 400 余类性状研究，鉴定了 1.5 万余个与疾病或性状相关的遗传易感位点^[1]。此外，新一代测序技术的飞跃进步，促进了全基因组、全外显子组和候选基因区域深度测序的快速发展，与 GWAS 策略相互补充，为全面揭示复杂性疾病的遗传机制提供了前所未有的机遇。我国科学家在心血管疾病、恶性肿瘤等常见疾病的遗传易感性研究中也取得了丰硕成果，在中国人群中发现了大量的疾病易感基因和位点，为揭示复杂性疾病的遗传机制以及从基因组水平进行疾病风险预测奠定了科学基础。笔者课题组与国内外多个课题组合

作,利用中国人群9种常见恶性肿瘤的GWAS数据进行遗传度分析,发现基因组常见变异在胃癌、食管鳞癌、结直肠癌、肺癌和上皮性卵巢癌中所能解释的遗传度分别为20.26%、19.86%、16.30%、15.17%和13.31%,在肝癌、前列腺癌、乳腺癌和鼻咽癌中所解释的遗传度在10%左右^[2]。

由于复杂性疾病的发生受控于多基因多位点,单个或少数基因位点的效应较弱,无法准确预测疾病,因此需要综合多基因多位点信息,而多基因风险评分(polygenic risk score, PRS)是目前的常用策略,也是复杂性疾病遗传易感性研究的新阶段。PRS旨在量化多个基因或位点的累积效应,将数十、数百、数千甚至更多的基因组变异信息浓缩成衡量个体疾病易感性的分值。其最常用的构建方法包含两个步骤:首先是“变量选择”的过程,以确定哪些易感性位点需要包含在模型中;其次是“权重估算”,以获得需要附加到所选变量的系数或权重的过程。笔者课题组前期介绍了5种遗传风险评分计算的原理和方法,可为研究者构建评分提供参考^[3]。近期研究表明,PRS通过整合多个遗传易感性位点的信息,能够提高人群风险预测、筛查及干预的效果,是实现复杂性疾病精准预防的关键。

二、多基因风险评分在复杂性疾病中的应用研究

首先,PRS在复杂性疾病的风险预测方面已显示出良好的应用前景。近期来自麻省总医院和Broad研究所的研究人员利用PRS识别5种常见疾病的高危人群,其中包括冠状动脉疾病、房颤、2型糖尿病、炎症性肠病和乳腺癌^[4]。研究人员首先基于大规模GWAS数据对上述5种疾病分别构建最优的PRS,然后对来自英国生物标本库(UK Biobank)的40多万名欧洲裔个体进行了风险评分算法的测试和验证。结果发现,PRS $>P_{92}$ 的高危人群发生冠状动脉疾病的风险是其余92%人群的3倍以上,类似于罕见单基因突变所导致的危险水平(如家族性高胆固醇血症基因突变,人群突变率0.4%),但PRS所识别的高危人群数量是罕见单基因突变的20倍。此外,对于房颤、2型糖尿病、炎症性肠病和乳腺癌,PRS可分别将总人群的6.1%、3.5%、3.2%和1.5%界定为高危人群,其发病风险可达其余人群的3倍。笔者课题组在国际上首次使用大规模前瞻性队列评价了PRS在肺癌发病风险预测中的应用效果^[5]。研究首先在中国人群中进行肺癌大样本多中心GWAS研究(包括13 327例非小细胞肺癌病例和

13 328例对照),系统筛选出19个独立的中国人肺癌易感性位点,以此构建了多基因风险评分PRS-19;随后,基于超大型前瞻性队列——中国慢性病前瞻性研究项目(China Kadoorie Biobank, CKB)中的近10万人(中位随访时间10.4年),前瞻性评价了PRS-19应用于预测肺癌风险的效能。结果显示,PRS $>P_{90}$ 的人群发生肺癌的风险是PRS $<P_{90}$ 人群的2.37倍;同时,PRS与吸烟存在显著的累积效应,遗传高风险(PRS $>P_{95}$)同时重度吸烟者其发生肺癌的风险与不吸烟者相比达到近4倍,支持PRS作为年龄和吸烟等因素之外的重要肺癌风险预测指标,优化高危人群的筛选标准。值得注意的是,遗传高风险的轻度吸烟者发生肺癌的风险与遗传低风险的重度吸烟者相近,这一部分人群也存在筛查价值,因而为扩展肺癌高危人群的筛查范围提供了科学证据。中国医学科学院阜外医院顾东风课题组基于亚洲人群GWAS鉴定了22个与血压相关的独立易感性位点并构建PRS,在包含26 262例中国人群的队列研究中前瞻性分析了PRS与心血管疾病发生风险的关联^[6]。结果显示PRS每增加一个标准差,心血管疾病发生风险升高16%;相比PRS $<P_{80}$ 人群,PRS $>P_{80}$ 人群罹患心血管疾病的风险增加43%,表明中国人群血压相关PRS可提高心血管疾病的风险预测效能,促进其高危人群的早期干预。

其次,在复杂性疾病筛查中应用PRS有望优化筛查方案,提高筛查的成本效益。英国National Health Service乳腺癌筛查项目推荐年龄47~73岁之间的女性应常规进行乳腺X线筛查。研究者估算47岁女性发生乳腺癌的10年绝对风险平均值为2.4%,采用PRS进行风险预测发现,对于风险 $>P_{90}$ 女性,她们在32~35岁时发生乳腺癌的10年绝对风险就已经达到2.4%,应纳入高危人群筛查,而对于PRS预测风险 $<P_{90}$ 女性,她们在一生中罹患乳腺癌的绝对风险始终 $<2.4%$ ^[7]。因此,相比于传统的仅依据一个年龄范围纳入高危人群的筛查项目,PRS有助于为不同遗传风险的女性制定针对性的筛查年龄。有研究进一步评估了乳腺癌筛查中应用PRS分层的成本效益,发现当筛查限制在PRS $>P_{70}$ 的高风险女性时,每1万人次的筛查成本将比单纯基于年龄的筛查节约54万英镑,过度诊断减少71.4%,而质量调整寿命年(quality-adjusted life year, QALY)增加443^[8]。因此,利用PRS风险分层可提供比单纯使用年龄范围更精准的筛查策略,避免低风险女性接

受不必要的检查。

此外,PRS也有助于通过风险分层,提高预防性干预的人群获益。有研究者首先在一项前瞻性队列研究中构建并评估了含有27个冠心病易感性位点的PRS,继而利用4项随机对照试验数据,分析了处于不同PRS等级的人群在服用他汀类药物后冠心病相对风险和绝对风险降低的情况^[9]。结果显示,低遗传风险组($PRS < P_{20}$)、中等遗传风险组($PRS P_{20} \sim P_{80}$)和高遗传风险组($PRS > P_{80}$)在服用他汀类药物后,冠心病相对风险分别降低13%、29%和48%。低遗传风险组在使用他汀类药物后冠心病绝对风险从3.0%下降到1.9%,而高遗传风险组从6.6%下降到3.6%,在高遗传风险组中预防1例冠心病所需要治疗的人数(number needed to treat, NNT)仅为低遗传风险组的1/3。另一项随机对照试验研究发现,尽管高遗传风险组和其他风险组人群在服用他汀类药物后LDL-C下降幅度相似,但高遗传风险组的亚临床冠状动脉粥样硬化风险降低幅度最大,支持采用PRS优化心血管疾病高危人群的筛选标准,从而使这些人群最大程度地从预防性服用他汀类药物中获益^[10]。

三、多基因风险评分在复杂性疾病中应用的机遇和挑战

随着人类基因组学研究的飞速发展,GWAS成功鉴定出大量复杂性疾病的易感基因和位点。以此为基础,PRS在疾病风险分层、筛查及精准预防等方面已初显其应用价值。由于全基因组分型芯片费用不断降低,且个体易感基因和遗传变异等信息自出生以后即保持不变,因此预期PRS在疾病精准预防乃至临床应用中潜力巨大。然而,实现PRS的广泛应用仍面临一系列挑战。

首先,PRS识别的高危人群仅包含全部病例的一小部分。例如已有研究显示,在处于 $PRS > P_{99}$ 、 P_{95} 、 P_{90} 和 P_{80} 的人群中,乳腺癌病例分别占全人群乳腺癌病例的3.6%、12%、21%和35%^[11]。该问题在一定程度上归因于“遗传度缺失”。正在进行或即将开展的大规模测序研究和多中心大样本的GWAS将会发现一批新的疾病易感位点,包括常见、低频和罕见变异。因此未来PRS的构建需要纳入更多的疾病易感位点以提高模型的预测效能。

其次,基因-环境交互作用在复杂性疾病的发生过程中具有重要作用,而目前大多数疾病风险预测仅考虑遗传和环境因素的主效应,忽略了基因-环境交互作用。基因-环境相乘交互作用的传统检验方

法统计学效能较低。有研究者提出了综合多个GWAS进行基因-环境交互作用分析的新方法^[12],在随后的心血管疾病和血脂影响因素的研究中展现了良好的应用效果^[13-14]。此外,与PRS不同,环境因素可能随着时间的迁移而变化。前瞻性队列研究需要对环境危险因素暴露进行动态测量,并开发新型统计学方法,将具有时间依赖性的环境暴露信息纳入多基因预测模型,并对该模型的科学性和应用价值进行验证和评估。

最后,PRS旨在评估个体在未来某一时间段内罹患疾病风险的概率,而非判断个体是否患有疾病。医务工作者和公众对PRS应用于疾病风险预测尚缺乏足够的认识。如何在现有医疗体系下有效地实施疾病风险预测并进行科学的解读,也是医务工作者需要考虑的重要问题,包括解决风险评估所带来的法律和伦理问题、提高公众对新型风险预测方案的认可度、确定最佳的服务提供机制、开发新的风险咨询方式等^[15]。

总之,应用PRS预测复杂性疾病风险的机遇和挑战并存。通过推动大规模队列人群的基因组流行病学研究,综合环境和遗传因素构建更为完善的预测模型,并开展大规模的人群预测效果评估,将有助于优化现有疾病筛查指南、推动个体化精准预防,最终实现降低疾病负担、提高人群健康水平的公共卫生目标。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] 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):D1005-1012. DOI:10.1093/nar/gky1120.
- [2] Dai JC, Shen W, Wen WQ, et al. Estimation of heritability for nine common cancers using data from genome-wide association studies in Chinese population [J]. *Int J Cancer*, 2017, 140(2):329-336. DOI:10.1002/ijc.30447.
- [3] 王毓,戴俊程,孙义民,等. 遗传风险评分的原理与方法[J]. *中华流行病学杂志*, 2015, 36(10):1062-1064. DOI:10.3760/cma.j.issn.0254-6450.2015.10.005.
Wang C, Dai JC, Sun YM, et al. Genetic risk score: principle, methods and application [J]. *Chin J Epidemiol*, 2015, 36(10):1062-1064. DOI:10.3760/cma.j.issn.0254-6450.2015.10.005.
- [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. DOI:10.1038/s41588-018-0183-z.
- [5] Dai JC, Lyu J, Zhu M, et al. Identification of risk loci and a

polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese population[J]. Lancet Respir Med, 2019, In Press. DOI: [http://dx.doi.org/10.1016/S2213-2600\(19\)30144-4](http://dx.doi.org/10.1016/S2213-2600(19)30144-4).

[6] Lu XF, Huang JF, Wang LY, et al. Genetic predisposition to higher blood pressure increases risk of incident hypertension and cardiovascular diseases in Chinese [J]. Hypertension, 2015, 66 (4): 786-792. DOI: 10.1161/HYPERTENSIONAHA.115.05961.

[7] Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants [J]. J Natl Cancer Inst, 2015, 107 (5) : djv036. DOI: 10.1093/jnci/djv036.

[8] Pashayan N, Morris S, Gilbert FJ, et al. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model [J]. JAMA Oncol, 2018, 4 (11) : 1504-1510. DOI: 10.1001/jamaoncol.2018.1901.

[9] Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials [J]. Lancet, 2015, 385 (9984) : 2264-2271. DOI: 10.1016/S0140-6736(14)61730-X.

[10] Natarajan P, Young R, Stitziel 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. DOI: 10.1161/CIRCULATIONAHA.116.024436.

[11] Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes [J]. Am J Hum Genet, 2019, 104 (1) : 21-34. DOI: 10.1016/j.ajhg.2018.11.002.

[12] Manning AK, LaValley M, Liu CT, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients [J]. Genet Epidemiol, 2011, 35 (1) : 11-18. DOI: 10.1002/gepi.20546.

[13] Rao DC, Sung YJ, Winkler TW, et al. Multiancestry study of gene-lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: design and rationale [J]. Circ Cardiovasc Genet, 2017, 10 (3) : e001649. DOI: 10.1161/CIRCGENETICS.116.001649.

[14] Bentley AR, Sung YJ, Brown MR, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387 272 individuals identifies new loci associated with serum lipids [J]. Nat Genet, 2019, 51 (4) : 636-648. DOI: 10.1038/s41588-019-0378-y.

[15] Chatterjee N, Shi JX, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention [J]. Nat Rev Genet, 2016, 17 (7) : 392-406. DOI: 10.1038/nrg.2016.27.

(收稿日期: 2019-07-24)

(本文编辑: 李银鸽)

读者·作者·编者

本刊常用缩略语

本刊对以下较为熟悉的一些常用医学词汇将允许直接用缩写,即在文章中第一次出现时,可以不标注中文和英文全称。

OR	比值比	HBeAg	乙型肝炎核心抗原
RR	相对危险度	HBeAg	乙型肝炎 e 抗原
CI	可信区间	HBsAg	乙型肝炎表面抗原
P_n	第 n 百分位数	抗-HBs	乙型肝炎表面抗体
AIDS	艾滋病	抗-HBc	乙型肝炎核心抗体
HIV	艾滋病病毒	抗-HBe	乙型肝炎 e 抗体
MSM	男男性行为者	ALT	丙氨酸氨基转移酶
STD	性传播疾病	AST	天冬氨酸氨基转移酶
DNA	脱氧核糖核酸	HPV	人乳头瘤病毒
RNA	核糖核酸	DBP	舒张压
PCR	聚合酶链式反应	SBP	收缩压
RT-PCR	反转录聚合酶链式反应	BMI	体质指数
C_t 值	每个反应管内荧光信号达到设定的阈值时所经历的循环数	MS	代谢综合征
PAGE	聚丙烯酰胺凝胶电泳	FPG	空腹血糖
PFGE	脉冲场凝胶电泳	HDL-C	高密度脂蛋白胆固醇
ELISA	酶联免疫吸附试验	LDL-C	低密度脂蛋白胆固醇
A 值	吸光度值	TC	总胆固醇
GMT	几何平均滴度	TG	甘油三酯
HBV	乙型肝炎病毒	COPD	慢性阻塞性肺疾病
HCV	丙型肝炎病毒	CDC	疾病预防控制中心
HEV	戊型肝炎病毒	WHO	世界卫生组织