

◇ 综 述 ◇

全基因组关联研究在冠心病中的研究进展

高 敏¹ 夏 青² 刘 永³ 综述

摘要 目前认为冠心病(CAD)是一种由多因素交互作用的慢性复杂性疾病,是危害人类健康的重大疾病之一,对CAD遗传机制的研究一直备受关注。近些年,全基因组关联研究(GWAS)的引入使CAD易感基因的研究更加快速、可靠,世界各地的医学研究者应用此研究策略已成功发现了多个CAD遗传易感基因或易感位点。现就近些年GWAS的进展、面临的挑战及其发展趋势进行阐述。

关键词 全基因组关联研究;冠心病;遗传易感位点;亚洲;汉族人群

中图分类号 R 541.4

文献标志码 A **文章编号** 1000-1492(2015)04-0558-04

全基因组关联研究(genome-wide association study, GWAS)是一种高通量基因分型技术^[1]。目前, GWAS已成功地在许多复杂性疾病中寻找到相关遗传易感基因或变异位点,例如阿尔兹海默症、肿瘤及类风湿性关节炎等^[2-3],为这些疾病的诊断、药物设计及个性化防治提供了坚实的理论基础。冠心病(coronary artery disease, CAD)是一种由环境因素和遗传因素共同作用的慢性复杂性疾病,是世界范围内人类健康的头号杀手。据世界卫生组织预测,2030年世界缺血性心脏病的死亡人数将占总死亡人数的13.1%^[4], CAD的防治刻不容缓。GWAS的引入为CAD相关易感基因的搜寻提供了新的机遇,此研究策略突破传统的候选基因连锁分析的一些不足,成功地发现了多个CAD相关易感区域。现就近些年CAD GWAS的研究进展进行如下报道。

1 CAD的GWAS

GWAS是一种通过高通量基因分型技术测定成

2015-01-04 接收

基金项目:2012年安徽省青年自然科学基金项目(编号:1208085QH155)

作者单位:¹安徽医科大学附属省立医院心电科,合肥 230001

²安徽医科大学公共卫生学院流行病与卫生统计学系,合肥 230032

³安徽医科大学公共卫生学院09级妇幼卫生专业,合肥 230032

作者简介:高 敏,女,副主任医师,责任作者, E-mail: gmbaauty@163.com

千上万个单核苷酸多态,结合临床表型确定疾病相关易感基因或易感位点,从而探索复杂性疾病遗传机制的方法^[1]。与在已知基因内寻找疾病易感位点的候选基因连锁研究相比, GWAS具有样本量大、重复性好、可多中心研究等优点;其在研究方法上主要分为3种:单阶段研究、两阶段研究及多阶段研究。两阶段研究和多阶段研究应用的最多,即先通过测定小样本量的单核苷酸多态性(single nucleotide polymorphisms, SNPs),从中找出阳性的SNPs,后期阶段再针对这些阳性SNPs进行重复验证。

第一个有实际意义的CAD GWAS是日本科学家Ozaki et al^[5]在Nature genetics上发表的关于淋巴瘤毒素-α基因(lymphotoxin-α gene, LTA)功能位点与心肌梗死的关联研究。此研究对94例心肌梗死病例和658例对照的92 788个SNPs位点进行高通量基因分型,然后在1 133例心肌梗死病例和1 006例对照人群中进行重复验证。结果显示, LTA基因中有2个SNPs位点与心肌梗死相关联($P < 1 \times 10^{-6}$)。目前,运用GWAS方法找到的CAD相关基因位点已达30多个,阳性结果在世界部分种族中得到重复验证。

2 CAD的GWAS最新国内外研究现状

2.1 国外现状 CAD GWAS起先主要在欧美等发达国家取得进展,发现了许多有重要价值的CAD相关易感基因和变异,例如9p21、1q41、12q24、6q25等区域,这些结果已在世界部分种族人群中得到了验证。近些年在此基础上,相关的研究进一步深入且新的研究又不断涌现。

2.1.1 欧洲 2007年,一项大型的GWAS WTC-CC^[6]在英国人群中展开,研究首次显示9p21区域rs1333049与高加索人群CAD显著关联($P < 1 \times 10^{-12}$)。2013年, van Setten et al^[7]检测了NELSON实验中2 620男性个体的2 500 000个SNPs,结果显示9p21区域与冠状动脉钙化显著关联(rs1537370, $P = 2.3 \times 10^{-11}$)。

2.1.2 美洲 Larson et al^[8]在弗拉明翰心脏研究基础上进行了一项以社区人群为基础的心血管疾病

的 GWAS 结果显示 9p21 区域的位点与 CAD 存在关联性 [$P = (2.5 \sim 3.5) \times 10^{-4}$]。Polfus et al^[9] 认为 GWAS 把过多的目光集中于基因对疾病的独立作用上,而忽略了环境与基因之间交互作用对疾病的影响。2013 年在 PLoS One 上发表了一篇基因与环境交互作用的 CAD GWAS 研究,研究基因变异联合吸烟等因素对冠状动脉钙化程度的影响。该研究的第一阶段选取 GENOA 中的同胞共 915 例,重复验证阶段在 FHS 同胞中选取 1 025 名欧裔美国人,并采用广义估计方程来解释吸烟组与非吸烟组同胞之间的关联。结果显示 TBC1D4 和 ADAMTS9 两个基因在吸烟暴露的情况下与冠状动脉钙化显著相关($P = 6.9 \times 10^{-5}$, $P = 7.1 \times 10^{-5}$)。同时,这些基因也被发现参与炎症通路,这种现象可能是吸烟诱导慢性系统性炎症而引起的。2014 年, Lusk et al^[10] 在欧裔美国人群中证实了染色体 9p21 区域的基因型变异与 CAD 的关联。研究者在 ARIC 研究中选取 7 589 名参与者(其中男性 3 720 名,女性 3 869 名),并选取了 9p21 区域的两个标记 SNPs 位点:rs1333049 和 rs3217992。结果显示这两个标记 SNPs 位点与 CAD 密切相关($P = 0.001$, $OR = 1.17$; $P < 0.001$, $OR = 1.22$)。

2.1.3 亚洲 2002 年, Ozaki et al^[5] 运用 GWAS 技术首次在日本人群中发现了 CAD 的一个易感基因 LTA,后期又发现 LGALS2 和 BRAP 基因与 CAD 有密切关联。2012 年,一项大型多阶段病例对照研究设计的 GWAS 在日本人群中展开^[11]。该研究的第一阶段运用 TaqMan assay 技术对 806 例 CAD 病例和 1 337 例健康对照的 451 382 个 SNPs 进行检测,以 $P < 5 \times 10^{-8}$ 为统计学显著性标准,在 3 052 例病例和 6 335 例对照的人群中进行重复验证,结果显示位于 12q24 上的 BRAP 和 ALDH2、位于 6p21 的 HLA-PQB1、位于 9p21 上的 CDKN2A/2B 基因与 CAD 显著相关($P = 1.6 \times 10^{-34}$, $P = 4.7 \times 10^{-7}$, $P = 6.1 \times 10^{-16}$)。Hirokawa et al^[12] 在 2014 年发表了一篇以日本人群为基础的大型心肌梗死(myocardial infarction, MI)基因组分析。该研究第一阶段采用 Illumina Human610-Quad 和 HumanHap550v3 Genotyping 两种芯片对 1 666 例 MI 病例和 3 198 例健康对照进行基因型分型检测,验证阶段使用 11 412 例 MI 病例及 28 397 例健康对照,结果显示两个新的 MI 易感基因位点:3p24.3 上的 PLCL2(rs4618210, $P = 2.60 \times 10^{-9}$, $OR = 0.91$);19p13.3 上的 AP3D1-DOT1L-SF3A2(rs3803915, $P = 3.84 \times 10^{-9}$, $OR =$

0.89)。此外,该研究验证了 14 个先前已报道的基因,通路分析结果还显示这些位点可能通过促进动脉粥样硬化而参与心肌梗死的发病过程。2013 年, Lee et al^[13] 以韩国人群为基础的进行了一项 CAD GWAS 研究,结果显示 1p13.3(rs99839)、9p21.3(rs4977574)、11q22.3(rs974819)3 个易感基因位点与 CAD 有显著关联。研究的前期阶段 Lee et al 使用基因填补技术,成功保留了 5 125 961 个基因型。Hager et al^[14] 在黎巴嫩人群中展开了一项大型的队列研究设计的 CAD GWAS 研究,研究的第一阶段包括 1 949 名个体,验证在另外的 2 547 名个体中进行,结果发现新的易感基因 PHACTR1 与 CAD 密切相关($OR = 1.34$, $P = 8.02 \times 10^{-10}$)。

2.2 国内现状 相对于国外 CAD GWAS 方面的成就,国内的发展较为滞后。国内相关研究显示,部分已发现的 CAD 易感基因位点在中国汉族人群中也得到验证。Li et al^[15] 开展了一项以医院为基础的中国汉族人群的病例对照研究,旨在重复验证国外 CAD GWAS 发现的 9p21 染色体区域,研究结果显示 MTAP 基因上的 rs101188757 位点可能与汉族人群的缺血性脑卒中有关联($OR = 1.31$, $P = 0.025$)。进一步研究还显示 rs101188757 位点和饮酒、高血压共同作用于汉族人群缺血性脑卒中的发病过程。随着国内 GWAS 的进一步发展,越来越多新的 CAD 易感基因位点被发现。2011 年, Wang et al^[16] 在 Nature 上首次发表了汉族人群的 CAD GWAS,该研究采用三阶段病例对照研究,以 $P < 0.01$ 为统计学显著标准来筛选阳性 SNPs 进入重复验证阶段。研究结果显示,位于 6p24.1 上的 C6orf105 与 CAD 易感性显著相关(rs6903956, $OR = 1.51$, $P = 4.87 \times 10^{-12}$)。Chen et al^[17] 在汉族人群中开展了一项两阶段病例对照设计 CAD GWAS,研究结果显示 KLF14(rs4731702)有助于降低心肌梗死的发生($OR = 0.72$, $P = 3.85 \times 10^{-3}$)。

3 Meta 分析

Meta 分析属于描述性的二次研究,相对于传统的描述性综述而言,Meta 分析可以对效应指标进行更准确、客观的评估,并能解释不同研究结果间的异质性。2012 年,一篇包括 2 个中国汉族人群 CAD GWAS 的 Meta 分析在 Nature 上发表^[18]。该研究第一阶段包括 CAD 病例 1 515 名和健康对照 5 019 名,在重复验证阶段选取 15 460 CAD 病例和 11 472 名对照。该研究以 $P < 5 \times 10^{-8}$ 为统计学显著性标

准,结果显示 TTC32-WDR35、GUCY1A3、C6orf10-BTNL2 及 ATP2B1 基因达到这一显著性标准,与汉族人群 CAD 易感性有关联。Davies et al^[19] 在 10 个大型 CAD GWAS 中采用 Meta 分析方法进行研究,研究的第一阶段检测了 5 个 GWAS 的 7 123 名病例和 6 826 名对照的近 5 000 000 个 SNPs,随后在另外的 5 个 GWAS 中的 5 211 名病例和 5 821 名对照中进行重复验证。研究结果显示,位于主要组织相容性复合体上的染色体 6p21.3 与 CAD 密切相关 ($rs3869109$, $P = 1.12 \times 10^{-9}$)。2013 年,Chan et al^[20] 在美国心脏病学会杂志中发表了一篇国际合作的 Meta 分析。该研究囊括了来自 21 个独立研究的 33 673 名参与者,研究结果显示 9p21 位点与 CAD 存在密切的关联性 ($OR = 1.31$, 95% $CI: 1.20 \sim 1.43$)。Meta 分析与循证医学的思想相符合,合理的运用可以促进 CAD GWAS 的发展,并增加研究结果的可信度,故 Meta 分析可能会成为未来 CAD GWAS 的主导方法之一。

4 CAD GWAS 研究的挑战性

4.1 提高 HapMap 数据进行基因型填补的力度

基因型填补是指依据已分型位点的基因型对数据缺失的位点或未分型位点进行基因型预测的方法。对分型缺失率较高的基因芯片利用 HapMap 数据进行填补可以提高全基因组遗传标记的覆盖率及研究效率,增加阳性关联位点的筛查成功率。根据 HapMap 数据进行基因型填补首先要利用 HapMap 数据,获得单体型信息,再根据样本观察到的基因型推测该样本最有可能携带的单体型,并据此将该单体型上相应位点的等位基因作为最有可能的填补值。目前在 GWAS 中,较为常用的填补软件包括: MACH、Impute 和 fastPHASE。另外,PLINK、TUNG 和 BEAGKE 等数据分析软件也包含了填补模块。HapMap 将为研究人员提供对人类健康和疾病以及对药物和环境的反应有影响的相关基因的关键信息。然而,检查填补后的数据往往可以显示,尽管填补后位点的个数增加了,部分位点仍然存在缺失值。故填补不是万能的金钥匙,想要达到理想的标准,相关问题仍需处理,如在填补之后,在分析时设定一些条件,剔除缺失比例仍然较高的位点,以及填充质量欠佳的位点。

4.2 深化基因-基因、基因-环境间交互作用的研究 现阶段的 CAD GWAS 将 CAD 的发病机理归结为多个独立的易感位点和区域,弱化了其位点-位

点之间、遗传-环境之间的交互作用。不同的环境外暴露混合后,暴露的混杂效应会削弱研究的把握度,致使大量的易感位点或区域被掩盖。Polfus et al^[9] 认为全基因组关联研究把过多的目光集中于基因对疾病的独立作用上,而忽略了环境与基因之间交互作用对疾病的影响。2013 年在 PLoS One 上发表了一篇基因与环境交互作用的 CAD GWAS 研究,研究基因变异联合吸烟因素对冠状动脉钙化程度的影响。结果显示 TBC1D4 和 ADAMTS9 两个基因在吸烟暴露的情况下与冠状动脉钙化显著相关 ($P = 6.9 \times 10^{-5}$, $P = 7.1 \times 10^{-5}$)。Zhang et al^[21] 创建了一个新的统计学方法 PLSPM 来检测基因-基因间的交互作用,研究结果显示该方法优于传统的单核苷酸多态性回归模型、PCA 模型及其他基因模型。

4.3 国际合作和 Meta 分析的运用 GWAS 多采用两阶段和多个阶段研究,其样本量一般在 1 000 以上,巨大的样本量使得研究对象的选取受限且耗费大量人力、物力及财力。然而,对于病例对照研究来说样本量越大研究的可信度就越大,特别是第一阶段的样本量对研究的影响更大。对于上述问题,Meta 分析可以很好地解决。Meta 分析通过合并多个研究的数据来增加样本量,从而提高检验效能。2011 年,CARDIOGRAM 协会运用 Meta 分析原理收集了 14 个关于冠心病及心肌梗死的 GWAS 数据^[22],其中包括 22 233 个病例及 64 762 个健康对照,样本人群均为欧洲血统。该研究显示了 13 个新的易感位点,并同时证实至少 10 个既往已有报道的易感位点。虽然 GWAS 的 Meta 分析在探索易感基因方面取得很多显著性的成就,但是随着样本量的不断增加,样本之间的异质性也会不断地凸显出来,如何平衡样本量和异质性之间的矛盾关系是需要深入探究的话题。

4.4 GWAS 的研究成果转换 首先,虽然 GWAS 在探寻 CAD 易感基因方面取得了很大成功,但是目前得到的 GWAS 结果只能解释小部分的 CAD 遗传发病机制,大部分的遗传易感基因仍未被发现。其次,已显示研究成果绝大部分未被运用到临床实践,仅少数在药物研制得到初步运用,例如药物基因组学通过 5-脂氧化酶通路的 CAD 易感基因研究开发了抑制该通路酶活性的物质和受体拮抗剂^[23]。这些新开发的药物,有望在使用降脂药物治疗 CAD 的基础上更好地控制和预防 CAD。在个体化治疗方面,由于每个个体间的遗传差异,治疗方法也不尽相同。如何有效地将 GWAS 研究成果用于诊断、预

防及治疗 CAD 将成为未来 CAD GWAS 的重要研究领域。

5 展望

随着人类基因组单倍型图计划以及“国际千人基因组计划”的完成,人类已经步入后基因组时代。后基因组时代,基因组研究的重心已由测定基因的 DNA 序列转移到从分子整体水平对生物学功能的研究上,在分子层面上探索人类健康和疾病的秘密。相信 GWAS 的不断完善会使得 CAD 的遗传机制越来越清晰,那时研究关注的重点则是如何将 GWAS 的研究成果有效地服务于临床实践,更好地服务人类的健康事业,从而推动人类冠心病的预防和治疗,降低冠心病发病率、致残率和死亡率,提高人类的健康水平和生活质量。

参考文献

- [1] Pearson T A, Manolio T A. How to interpret a genome-wide association study [J]. *JAMA*, 2008, 299(11):1335-44.
- [2] Bailey J N, Pericak-Vance M A, Haines J L. The impact of the human genome project on complex disease [J]. *Genes (Basel)*, 2014, 5(3):518-35.
- [3] Wallace C, Newhouse S J, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia [J]. *Am J Hum Genet*, 2008, 82(1):139-49.
- [4] Mathers C D, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030 [J]. *PLoS Med*, 2006, 3(11):e442.
- [5] Ozaki K, Ohnishi Y, Iida A, et al. Functional SNPs in the lymphotxin- α gene that are associated with susceptibility to myocardial infarction [J]. *Nat Genet*, 2002, 32(4):650-4.
- [6] Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3 000 shared controls [J]. *Nature*, 2007, 447(7145):661-78.
- [7] van Setten J, Jsgum I, Smolonska J, et al. Genome-wide association study of coronary and aortic calcification implicates risk loci for coronary artery disease and myocardial infarction [J]. *Atherosclerosis* 2013, 228(2):400-5.
- [8] Larson M G, Atwood L D, Benjamin E J, et al. Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes [J]. *BMC Med Genet*, 2007, 8(Suppl 1):S5.
- [9] Polfus L M, Smith J A, Shimmin L C, et al. Genome-wide association study of gene by smoking interactions in coronary artery calcification [J]. *PLoS One* 2013, 8(10):e74642.
- [10] Lusk C M, Dyson G, Clark A G, et al. Validated context-dependent associations of coronary heart disease risk with genotype variation in the chromosome 9p21 region: the Atherosclerosis Risk in Communities study [J]. *Hum Genet*, 2014, 133(9):1105-16.
- [11] Takeuchi F, Yokota M, Yamamoto K, et al. Genome-wide association study of coronary artery disease in the Japanese [J]. *Eur J Hum Genet*, 2012, 20(3):333-40.
- [12] Hirokawa M, Morita H, Tajima T, et al. A genome-wide association study identifies *PLCL2* and *AP3D1-DOT1L-SF3A2* as new susceptibility loci for myocardial infarction in Japanese [J]. *Eur J Hum Genet*, 2015, 23(3):374-80.
- [13] Lee J Y, Lee B S, Shin D J, et al. A genome-wide association study of a coronary artery disease risk variant [J]. *J Hum Genet*, 2013, 58(3):120-6.
- [14] Hager J, Kamatani Y, Cazier J B, et al. Genome-wide association study in a Lebanese cohort confirms *PHACTR1* as a major determinant of coronary artery stenosis [J]. *PLoS One*, 2012, 7(6):e38663.
- [15] Li S J, Hu W L, Liu D T, et al. *MTAP* gene is associated with ischemic stroke in Chinese Hans [J]. *J Neurol Sci*, 2009, 284(1-2):103-7.
- [16] Wang F, Xu C Q, He Q, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population [J]. *Nat Genet*, 2011, 43(4):345-9.
- [17] Chen X, Li S, Yang Y, et al. Genome-wide association study validation identifies novel loci for atherosclerotic cardiovascular disease [J]. *J Thromb Haemost*, 2012, 10(8):1508-14.
- [18] Lu X, Wang L, Chen S, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease [J]. *Nat Genet*, 2012, 44(8):890-4.
- [19] Davies R W, Wells G A, Stewart A F, et al. A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex [J]. *Circ Cardiovasc Genet*, 2012, 5(2):217-25.
- [20] Chan K, Patel R S, Newcombe P, et al. Association between the chromosome 9p21 locus and angiographic coronary artery disease burden: a collaborative meta-analysis [J]. *J Am Coll Cardiol*, 2013, 61(9):957-70.
- [21] Zhang X, Yang X, Yuan Z, et al. A PLSPM-based test statistic for detecting gene-gene co-association in genome-wide association study with case-control design [J]. *PLoS One*, 2013, 8(4):e62129.
- [22] Schunkert H, König I R, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease [J]. *Nat Genet*, 2011, 43(4):333-8.
- [23] Funk C D. Leukotriene modifiers as potential therapeutics for cardiovascular disease [J]. *Nat Rev Drug Discov*, 2005, 4(8):664-72.