

doi:10.3969/j.issn.1006-9852.2021.07.002

• 国外医学动态 •

基于全基因组测序筛选腰椎管狭窄症
易感基因突变位点

摘 要 识别腰椎疾病的遗传易感因素有助于了解该病的潜在发病机制和探索新的治疗方法。通过电子病历系统、基因组网络以及格伊辛格卫生系统纳入 100,811 名参与者,对基因型和电子病历资料进行全基因组关联分析 (GWAS),使用有效算法和临床诊断代码将其分为病例组和对照组。病例组包括需医疗干预的腰痛 (LBP-HC)、腰骶神经根综合征 (LSRS) 和腰椎管狭窄症 (LSS)。GWAS 采用多因素 Logistic 回归分析,根据年龄、性别、特定区域因素和种族进行累积效应校正,并进行了固定效应反方差加权荟萃分析。全基因组意义上的遗传变异 ($P < 5 \times 10^{-8}$) 可以在 UKB 的独立样本中进行重复验证。研究数据显示 LBP-HC 的患病率为 48.8%,LSRS 的患病率为 19.8%,LSS 的患病率为 7.9%,其中 LBP-HC 无显著相关易感基因突变;LSRS 存在相关易感基因突变位点 (rs146153280: C > G, G 的比值比 $OR = 1.17, P = 2.1 \times 10^{-9}$),但未得到重复验证;与 LSS 相关联的易感突变基因是位于 2 号染色体上编码 *GFPT1*, *NFUI* 和 *AAK1* 基因的位点 (rs13427243: G > A, A 的比值比 $OR = 1.10, P = 4.3 \times 10^{-8}$),并可重复验证 ($OR = 1.11, P = 5.4 \times 10^{-5}$)。这是使用电子病历数据对腰椎疾病进行 GWAS 的第一个荟萃分析。通过本研究确定了与 LSRS 和 LSS 关联的两个新的易感基因突变位点,后者可重复验证。

关键词 遗传学;腰痛;坐骨神经痛;椎间盘突出;椎管狭窄

腰痛 (low back pain, LBP) 是美国公民就医的第五大病因,也是致残的主要原因。LBP 发病率高,社会影响大,但病因尚未明确。有效识别 LBP 的遗传易感因素将是探索发病机制和探究新疗法的一种有效途径。LBP 的广义遗传力高达 40%,显示其与遗传因素密切相关。基于两项全基因组关联分析 (GWAS) 数据确定了 LBP 相关易感基因突变位点,表明遗传因素是 LBP 的生物-心理-社会模式的基础,开启了 LBP 遗传机制研究的大门。在医学上 LBP 作为一种临床症状,而非一项诊断。临床上被普遍接受的两种常用诊断分别为腰骶神经根性疼痛或腰骶神经根综合征 (LSRS) 和腰椎管狭窄症 (LSS),其他诊断缺乏诊断标准。因此, LBP 的研究多集中在“非特异性 (原发性) 腰痛”方向,并未考虑潜在的外周致病因素,忽略了与 LBP 相关疾病的表型异质性可能会降低通过 GWAS 探索致病基因统计学效力。临床上也可通过疾病特定的“内在表型”进行诊断,例如公认的 LSRS 和 LSS 临床综合征。通常在病人自诉不清的情况下可在规范化临床诊疗中通过电子病历 (EHR) 关联的遗传生物库数据进行诊断。该遗传生物数据库可提供病人及受试者较为全面的病史,进而实现更准确的疼痛表型分型。

本项研究中,利用 EHR 生物库数据资源探讨了 LBP 和腰椎疾病的潜在遗传学发病机制。首次发现 LBP 相关易感基因突变位点,同时确定与 LSRS 和 LSS 相关的易感基因突变位点,最后利用 EHR 生物库数据进行 GWAS 分析确定了 LBP 和腰椎疾病相关的易感基因突变位点的表现型。

1. 研究方法

(1) 概述: GWAS 被称为“不可预知的”或“没有假设的”基因筛选方法。对两项独立的来源于电子病历、基因组网络和格伊辛格卫生系统的 GWAS 数据进行荟萃分析筛选易感突变基因,并在 UKB 的独立样本中对筛选的有显著差异的突变基因进行了体外验证。

(2) 腰痛和腰椎疾病的分型: 该研究严格按照 EHR 数据库和国际通用疾病分类编码第 9 版 (ICD-9) 中对腰痛和腰椎疾病的定义将纳入病例分为病例组和对照组。其中病例组包括: ①需医疗干预的腰痛 (LBP-HC 型); ②腰骶神经根综合征 (LSRS 型); ③腰椎管狭窄症 (LSS 型); ④腰椎源性神经病理性疼痛 (LSRS/LSS 型)。为确保所有观察者的有效随访时间,纳入年满 18 岁且规律接受医疗服务随访满 1 年的受试者。需医疗干预的 LBP-HC 型病

人临床上伴或不伴有下肢神经性疼痛,在疾病分类中属于非特异性(原发性)腰痛,很难划分到其他分型亚组中。可基于 EHR 的层次化分类算法来区分 LSRS 型和 LSS 型,LSRS 型病人通常合并有坐骨神经病变、梨状肌综合征、骶髂关节疼痛等症状;LSS 型病人主要表现为神经源性跛行和腰椎管狭窄为主的症状。

(3) 基因分型: eMERGE 网络数据主要使用 Illumina 平台和 Affymetrix 83 数组芯片技术对样本进行基因分型。单核苷酸突变位点 (SNVs) 的检测由密歇根基因插补网页工具完成。所有受试者均签署知情同意书,同时批准可获取 eMERGE3 数据库信息。

(4) 疾病表型和基因分型的体外验证: 初步探索显著差异的突变基因,随后通过 UKB 进行大样本体外验证,UKB 中研究人群与英国普通人群在疼痛发病率及相关社会人口和心理因素关系具有相似性。UKB 的腰骶神经根综合征 (LSR-UKB) 是指有一个或多个 ICD-10 诊断编码提示 LSR 或“坐骨神经痛”或“椎间盘突出/滑脱”的病例;将没有 ICD-10 诊断编码提示 LSR 和“坐骨神经痛”或“椎间盘突出/滑脱”的病例划分到对照组;LSS 和 LSS/LSRS 型的对照组和病例组使用相似的方法进行分组。

(5) 统计学分析: 参与者种族限于欧洲人群以便最大可能降低遗传异质性,其中三级亲属和近亲参与者排除在外。在 eMERGE3 数据库样本的 GWAS 分析中,设置过滤阈值如下: 次要等位基因频率 (MAF) < 0.005 , 插补系数 $r^2 < 0.3$, 哈迪温伯格平衡检验 (HWE) $P < 10^{-6}$, 基因分型率 < 0.98 , 个体率 < 0.98 。在 Geisinger 数据库样本的 GWAS 分析中,设置过滤阈值如下: MAF < 0.01 、插补系数 $r^2 < 0.3$ 或 < 0.4 (根据阵列来确定), HWE $P < 10^{-15}$, 基因分型率 < 0.90 , 个体率 < 0.90 。在 PLINK 1.9 中,对性别、年龄、特异性位点和种族自 1~10 进行了调整,并采用基因型加性模型对插补的 SNVs 进行 Logistic 回归分析。通过设置 SNVs 的滤过阈值进行了质量控制,滤过参数如下: MAF < 0.01 , 插补系数 $r^2 < 0.03$, HWE $P < 10^{-6}$, BETA ≥ 10 , SE ≤ 0 或 ≥ 10 , 预期次要等位基因计数 ≤ 100 。

荟萃分析使用 METAL 软件, $P < 5 \times 10^{-8}$ 作为在全基因组分析中具有统计学意义的标准。利用连锁不平衡评分回归方法计算遗传性。根据文献报道 LBP 的患病率 (84%) 和因 LBP 而寻求医疗干预比率 (58%), 推算普通人一生中因 LBP 寻求医疗干预的

患病率为 48.7%, 以此代表 LBP-HC 的遗传易感性。同样, 基于 LSRS 和 LSS 的患病率 (26% 和 11%) 进而估计 LSRS 和 LSS 和 LSRS/LSS 的遗传易感性。暂无关于以人群为基础的 LSRS/LSS 患病率数据, 因此推测其患病率略高于 LSRS 的患病率 (29%)。为了量化疾病表型之间的总体共有遗传因素, 利用连锁不平衡得分回归计算了 LBP-HC、LSRS 和 LSS 之间的基于 SNV 的遗传相关性 (r_g), 鉴于 LSRS 和 LSS 分型是 LBP-HC 中的两个亚组。使用圣克鲁斯大学的基因组浏览器检验了全基因组范围内的重要 SNVs 位点。通过 LocusZoom 可视化方式来识别独立的 SNVs, 同时使用全基因组复杂性状分析软件包的汇总数据联合分析方法检测显著变异的突变位点。为了鉴定已筛 SNVs 位点的功能性, 在 GTEx 中对其进行检测, 并确定了与相邻基因差异表达相关的变异。

为了描述全基因组显著 SNVs 对其他性状的潜在共有遗传影响, 使用 Pheno Scanner 数据库第 2 版检测了主要的 SNVs 与其他性状的关系 ($P < 1 \times 10^{-5}$), 并在 eMERGE3 中的主要 SNVs 进行了全表型组关联研究。PheWASs 能系统地检测一种基因突变相关的所有表型。PheWAS 包含所有的 eMERGE3 EA 参与者的表型和 "phecodes" 定义的表型。研究了主要的 SNV 和 890 个具有至少 2 个 ICD 编码的 phecodes 之间的关联性, $P < 5.6 \times 10^{-5} (0.05/890)$ 作为每个主要 SNV 显著性的阈值。

体外研究验证显示有 3 个 SNVs 突变位点 (rs12310519 [SOX5], rs7814941 [CCDC26/GSDMC] 和 rs3180 [SPOCK2/CHST3]) 与腰痛有关。另外一项研究发现一个与冰岛人 LSRS 相关的 SNV (rs7831174), 其所采用的筛选阈值为 $P < 0.05 (0.05/1)$ 。

2. 结果

(1) 腰痛和腰椎疾病在 eMERGE3 及 Geisinger 数据库中的 GWAS: 荟萃分析研究显示 LBP-HC 的患病率为 48.8%, LSRS 为 19.8%, LSS 为 7.9%, LSRS/LSS 为 22.1%。全基因组水平研究显示 LBP-HC 无显著相关的 SNV 位点, LBP-HC 的全基因组 SNV 遗传率结果为观测尺度遗传力为 3.1% ($\pm 0.5\%$), 易感性尺度的遗传力为 4.9% ($\pm 0.8\%$)。

LSRS 的 SNV 遗传率在观察量表上为 4.2% ($\pm 0.5\%$), 在责任量表上为 9.4% ($\pm 1.0\%$)。研究发现 3 个突变位点与 LSRS 显著相关。差异最显著的突变位点位于 9 号染色体上 NXNL2 基因 (变异在 5'402 KB 处) 和 C9orf47/S1PR3/SHC3 基因 (变异在 3'15 KB 处) 区域间, 该基因位点为 rs146153280:

C > G (G 等位基因比值比 $OR = 1.17$, $P = 2.1 \times 10^{-9}$, G 等位基因频率为 0.04), 其中基因 *C9orf47*, *S1PR3* 以及 *SHC3* 彼此紧密排列和/或部分重叠。另 2 个突变位点位于该区域 rs146153280 两侧, 与之呈线性关系, 但在以 rs146153280 为主的条件性分析中并无显著差异。

研究显示 LSS 的观测尺度遗传力为 2.2% ($\pm 0.5\%$), 易感性尺度的遗传力为 8.0% ($\pm 2.0\%$)。研究发现 4 个突变位点与 LSS 显著相关。其中差异最显著的突变位点位于 2 号染色体上 *AAK1* 基因的 5' 端非编码区, 该基因位点为 rs13427243: G > A (A 等位基因 $OR = 1.10$, $P = 4.3 \times 10^{-8}$, G 等位基因频率为 0.40)。这 4 个突变位点位于基因 *GFPT1*、*NFU1* 以及 *AAK1* 片段上共计 131, 465 对碱基, 余 3 个显著差异突变位点与 rs13427243 呈线性关联, 但在以 rs13427243 为主的条件性分析中并无显著差异。

LSRS/LSS 的观察尺度遗传力为 4.4% ($\pm 0.5\%$), 易感性尺度的遗传力为 9.3% ($\pm 1.0\%$)。GWAS 结果显示显著突变位点与 LSRS 型研究结果相同, 均为位于 9 号染色体上的 rs146153280 ($OR = 1.17$, $P = 1.4 \times 10^{-9}$)。另 2 个突变位点在以 rs146153280 为主的条件性分析中也无显著差异。

研究显示腰痛与 LSRS 的遗传相关系数 (r_g) 为 1.00, 与 LSS 的相关系数为 0.86。LSRS 和 LSS 的遗传相关系数为 0.67。

(2) 突变位点的基因型-组织表达 (genotype-tissue expression, GTEx) 特征: LSRS 和 LSRS/LSS 的基因型-组织表达研究结果未发现与显著差异基因突变位点 (rs146153280) 显著相关的数量性状位点 (eQTL)。研究显示 LSS 型中的显著差异突变位点 rs13427243 在包括肌肉和神经等 20 种组织中在 *GFPT1* 和 *NFU1* 基因上具有显著相关的表达数量性状位点 ($P = 1.8 \times 10^{-20}$ - 7.4×10^{-5})。

(3) 突变位点的表型特征: 研究显示与 LSRS 型相关的位于 9 号染色体上的显著差异位点 rs146153280 与蛋白 *Semaphorin-4D* 和 *C9orf47* 基因的表达显著相关。与 LSS 相关的位于 2 号染色体上的差异位点 rs13427243 与 *GFPT1*, *NFU1* 和 *ANXA4* 的基因表达以及 DNA 甲基化和组蛋白修饰表观遗传特征显著相关。全表型关联研究结果表明显著差异突变位点在校正统计检验后发现与其他表型并无显著关联。

(4) 显著突变基因的外部验证: 在 UKB 中研究结果显示 LSRS、LSS 和 LSRS/LSS 的患病率分别为 5.8%、1.2% 和 6.4%, 明显低于在 eMERGE/

Geisinger 数据库中相应的患病率 (19.8%、7.9% 和 22.1%)。上述在 LSRS 和 LSRS/LSS 疾病中发现的显著差异突变位点 rs146153280 在 UKB 中未得到重复验证。然而, LSS 相关差异突变位点 rs13427243 在该数据库中得到重复验证 ($OR = 1.11$, $P = 5.4 \times 10^{-5}$)。

(5) 腰痛和腰骶神经根综合征相关显著差异突变位点的外部验证: 在 UKB 中得到验证的 LBP-HC 相关显著差异位点 rs12310519: C > T (*SOX5*) ($OR = 1.05$) 在 eMERGE/Geisinger 数据库中同样得到验证 (T 等位基因的 $OR = 1.05$, $P = 0.011$)。此外, 同样在 UKB 中得到验证的 LBP 相关显著差异位点 rs7814941: G > A (*CCDC26/GSDMC*) ($OR = 1.04$) 在 eMERGE/Geisinger 数据库中也同样得到验证 (A 等位基因的 $OR = 1.03$, $P = 0.005$)。但在 UKB 中得到验证的 LBP 相关性突变位点 rs3180 (*SPOCK2/CHST3*) ($OR = 0.98$, $P = 0.09$) 在 eMERGE/Geisinger 数据库中无法重复验证。

3. 讨论

本项 GWAS 研究结果主要通过利用多中心的 EHR 临床数据来识别需医疗干预的 LBP (LBP-HC) 和两种腰椎疾病 (LSRS 和 LSS) 相关的遗传标志物。研究纳入的样本量超过 10 万, 但全基因组分析未发现与 LBP-HC 具有显著关联的基因突变位点。虽 LSRS 和 LSS 型参与者病例数相对较少, 但全基因组研究新发现了与之相关的易感基因突变位点, 包括 LSS 相关的 1 个易感突变位点及 LSRS 相关的 1 个易感突变位点。随后在 UKB 中进行表型验证, 重复验证了突变位点 rs13427243 与 LSS 的显著关联性, 另外发现的突变位点未得到重复验证。

两项研究均报告了 LSS 相关遗传位点 (遗传力 67%~81%), 但在全基因组分析中未证实存在易感突变位点。本研究证实 LSS 相关显著易感突变基因位点 rs13427243: G > A [A 的 $OR = 1.10$, $P = 4.3 \times 10^{-8}$] 可在独立样本中重复验证 ($OR = 1.11$, $P = 5.4 \times 10^{-5}$), 该位点主要影响 *GFPT1*, *NFU1* 和 *AAK1* 基因编码, 进而影响多种组织中 *GFPT1* 和 *NFU1* 的表达。目前 *AAK1* 抑制剂治疗神经痛的研究已通过动物实验验证, 并完成了 1 期临床试验, 针对糖尿病周围神经痛的 2 期临床试验正在进行中。LSS 是腰椎相关性疼痛的常见病因, 通过在人类 GWAS 和以动物为模型的药物研发方面的结果均显示 *AAK1* 突变与 LSS 相关。临床上通常使用针对周围神经性疼痛的药物 (如加巴喷丁和普瑞巴林) 来控制 LSS 症状。如果将来 *AAK1* 抑制剂可安全有效地治疗糖尿病周围神经痛, 则可考虑 *AAK1* 抑制剂是否可用于

治疗 LSS 引起的神经病理性疼痛。GFPT1 基因可编码谷氨酰胺-果糖-6-磷酸转氨酶-1, 该限速酶在蛋白质和脂质糖基化过程中发挥重要作用。研究显示 GFPT1 基因的突变与先天性肌无力综合征有关, 临床表现为眼肌和延髓肌无力。NFUI 基因编码的蛋白质在铁硫簇生成过程中发挥重要作用, NFUI 基因的突变可引起线粒体功能障碍综合征。

本项全基因组荟萃分析研究显示在 9 号染色体 NXRS2 与 C9orf47 基因之间区域中的突变位点 rs146153280 与 LSRS 和 LSRS/LSS 表型显著相关。但在 UKB 中未得到重复验证, 可能因该数据库中对腰椎疾病定义的差异而导致。但并未影响 LSS 型中易感突变位点 rs13427243 的验证。UKB 中疾病的代码主要来自住院系统, 代码获得途径差异也造成独立样本重复验证中的不一致结果。鉴于 LSS 病人更多的接受住院手术治疗, LSRS 病人更多的接受保守治疗, 所以实际纳入的 LSS 病人数量较 LSRS 病人多。因此使用一致的疾病代码对 rs146153280 与 LSRS 的关联性进行重复验证。

本项全基因组荟萃研究结果验证了已报道的 LBP 相关易感基因位点: SOX5 基因中的 rs12310519 及 CCDC26/GSDMC 基因间区域中的 rs7814941。这些突变位点在 eMERGE/Geisinger 数据库中和 UKB 数据库中均得到验证, 但另已报道的 LBP 相关易感基因 SPOCK2/CHST3 未超过显著性阈值, 但具有一定关联性。本研究同样验证了 LSRS 表型中易感突变基因位点 rs6651255 与需要减压手术的腰椎间盘突出症之间的关联性, 但在 eMERGE/Geisinger 数据库中关联性相对小 (C 等位基因 $OR = 0.95$)。研究已证实 LBP-HC 易感突变基因位点 rs12310519 和 rs7814941 以及位于 CCDC26/GSDMC 区域的 LBP-HC 易感突变基因位点 rs7814941, 通过 GWASs 分析发现 LBP 易感突变位点与该疾病严重程度同样具有关联性。本研究结果有效阐明了遗传因素在腰椎疾病的发病机制方面的重要作用。

本研究结果有效表明基于 EHR 的特定腰椎相关诊断 (如 LSRS 和 LSS) 表型的个体变异的遗传发现率更高, 这一发现可在临床上更好指导疾病型的诊断。在临床上可有效对 LBP 病人进行亚组诊断, 同时提供个体化诊治。例如, LSRS 或 LSS 引起的顽固性神经性下肢疼痛或肌无力的病人可优选腰椎减压手术, 但该术式不适用于下肢无神经性症状/体

征的轴性 LBP 病人。同时, 全基因组显著突变基因关联性分析显示 LSRS 和 LBP 的遗传力基本相同 ($r_g \sim 1.0$), LSS 和 LBP-HC 的遗传力高度相似 ($r_g = 0.86$), LSRS 和 LSS 的遗传力基本相同 ($r_g = 0.68$)。LSS 与 LBP 或 LSRS 之间的遗传力与已报道的不同部位的肌肉骨骼疼痛的遗传力相似。例如, 颈部疼痛和髋关节疼痛具有相似遗传力 ($r_g = 0.83$), 颈部疼痛和膝关节疼痛具有相似遗传力 ($r_g = 0.64$), 但全基因组研究发现 LSRS 和 LBP 的遗传力关联性更强。综上所述, 研究发现的易感突变位点与疼痛基因组学的研究结果一致, 即遗传因素在疼痛表型分类发挥重要决定作用, 特定的疼痛表型对应特定的遗传因素。

本研究存在一定局限性。首先 LSRS 和 LSS 临床诊断代码的差异。其中 LSRS 或 LSS 的诊断可通过影像学表现 (如腰椎间盘突出或腰椎管狭窄), 也可以通过该疾病的临床综合征诊断。因此脊柱疾病的分类越具体将直接影响研究结果的可靠性。本项全基因组研究发现的特定易感突变基因位点不一定是致病突变基因, 仅代表增加该疾病风险的可能。研究发现的位于基因 GFPT1, NFUI 和 AAK1 区域与 LSS 相关的突变位点在多种组织类型的基因 GFPT1 和 NFUI 功能性 eQTL 表达中发挥多种功能效应。尽管未发现 AAK1 基因表达与 LSS 有显著相关性, 可能是由于组织范围有限 (不包括 LSS 中央椎管的形态学特征) 或样本量有限影响。目前, 虽研究已证实 AAK1, GFPT1 和 NFUI 中的突变位点与 LSS 密切相关, 但具体作用机制仍未可知, 推测可能是相邻基因的差异表达产生一定的功能效应。本研究结果可作为 LSS 遗传机制探索的开端, 未来仍需进一步探索验证突变位点与疾病的关联性。尽管本研究发现的基因突变位点与 LSS 的关联性较小 (rs13427243: G > A, 初次发现时关联率为 10%, 验证阶段为 11%), 随着对基因突变位点的进一步深入研究, 预计遗传效力会有所增加。本研究纳入的对象为欧洲人群, 研究结果具有一定种族局限性。

(Suri P, Stanaway IB, Zhang Y, et al. Genome-wide association studies of low back pain and lumbar spinal disorders using electronic health record data identify a locus associated with lumbar spinal stenosis. Pain, 2021. doi: 10.1097/j.pain.0000000000002221. 北京大学第三医院疼痛医学中心, 梁辰译 李水清校)