

分 类 号: R-741  
密 级:

学校代码: 10062  
学 号: 2013011331



**天津医科大学**  
TIANJIN MEDICAL UNIVERSITY

**硕 士 学 位 论 文**  
**MASTER'S DISSERTATION**

**论文题目: 血清 25OHD 和 FGF23 水平与  
缺血性卒中的遗传因果研究**

**T I T L E      The genetic causality study of serum  
25OHD and FGF23 levels on ischemic stroke**

**专业学位名称: 临床医学**

**专业学位领域: 神经病学**

**论 文 作 者 : 林灵敏**

**导            师 : 郝峻巍**

**天 津 医 科 大 学 研 究 生 院**

**二〇二〇年五月**

分类号: R-741

密 级:

学位类别: 科学学位 ☐ 专业学位 ☒

学校代码: 10062

学 号: 2013011331

学科门类: 医学



天津医科大学

TIANJIN MEDICAL UNIVERSITY

# 硕士学位论文

MASTER'S DISSERTATION

论文题目: 血清 25OHD 和 FGF23 水平与  
缺血性卒中的遗传因果研究

T I T L E     The genetic causality study of serum  
25OHD and FGF23 levels on ischemic stroke

专业学位名称: 临床医学

专业学位领域: 神经病学

论 文 作 者 : 林灵敏

导         师 : 郝峻巍

导师组成员 : 刘桂友

天津医科大学研究生院

二〇二〇年五月

## 学位论文原创性声明

本人郑重声明：所呈交的论文是我个人在导师指导下独立进行研究工作取得的研究成果。除了文中特别加以标注引用的内容和致谢的地方外，论文中不包含任何其他个人或集体已经发表或撰写过的研究成果，与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：林灵敏 日期：2020年4月13日

## 学位论文版权使用授权书

本学位论文作者完全了解天津医科大学有关保留、使用学位论文的规定，即：学校有权将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构送交论文，并编入有关数据库。

保密 ☐，在 \_\_\_\_\_ 年解密后适用本授权书。

本论文属于

不保密 ☒。

(请在相对应的方框内打“√”)

学位论文作者签名：林灵敏 日期：2020年4月13日

导师签名：郝增磊 日期：2020年4月13日

## 中文摘要

**背景：**日益增加的全球负荷对卒中的综合预防提出了更高的要求，可调节的维生素 D 是否对缺血性卒中存在保护性的效应目前尚无定论。成纤维细胞生长因子 23 (FGF23, Fibroblast growth factor 23) 是心血管疾病的独立危险因素，但 FGF23 与缺血性卒中的流行病学关联结果并不一致，且 FGF23 与其是否有因果关联也不明确。FGF23 可抑制维生素 D 的活化并促进活性维生素 D 的降解，有观点认为 FGF23 的部分生物学效应是通过降低活性维生素 D 水平形成的。因此，在特定人群中补充维生素 D 可能是有益的。

**目的：**本研究通过孟德尔随机化 (Mendelian randomization, MR) 框架评估遗传倾向于增加的循环 25-羟基维生素 D (25OHD, 25-hydroxyvitamin D) 水平和 FGF23 水平与缺血性卒中或其亚型风险的因果关联，以探索在不同人群中补充维生素 D 的可能效应。

**方法：**本研究利用来自英国生物库 UK BioBank 中 417,580 名欧洲裔的 25OHD 相关全基因组关联研究 (GWAS, genome-wide association study) 数据、来自 7 个队列荟萃研究 (N = 16,624) 的 FGF23 水平显著相关 GWAS 数据 (欧洲裔) 及 MEGASTROKE 卒中与其亚型的大型 GWAS 数据构建工具变量。通过遗传变异的相关性 ( $P < 5.0E-08$ )、连锁平衡性 ( $r^2 < 0.1$ )、可能的多效性关联以及工具强度筛选遗传变异。本研究通过两样本 MR 框架分析 25OHD、FGF23 水平与卒中或其亚型之间的遗传因果关联。我们以逆方差加权法 (MR-IVW, inverse-variance weighted method) 分析因果关联的性质，利用 Cochran Q、留一法以及加权的中位数法 (MR-WME, weighted-median method) 对结果的稳健性进行评估；此外，我们通过 MR-Egger 法、MR-PRESSO (Pleiotropy RESidual Sum and Outlier test) 方法分析遗传工具潜在的水平多效性，以 MR-Steiger 评估因果关联方向。最后，我们评估因果估计的统计效能。

**结果：**本研究未发现血清 25OHD 水平与缺血性卒中或其亚型存在因果关联。通过 60 个筛选后的工具变量分析关联性质，每增加 1 个标准差 (standard deviation, SD) 的 25OHD 水平，(MR-IVW 法) 急性缺血性卒中的比值比 OR (Odds Ratio) 为 1.02 (95% CI 0.94-1.11)，大动脉粥样硬化性卒中 (large artery atherosclerotic stroke, LAS) 的估计风险为 0.94 OR (95% CI 0.76-1.16)。我们

共获得 3 个通过筛选的 FGF23 相关工具变量, MR-IVW 评估结果为每增加 1 个 SD 的循环 FGF23 水平, 急性缺血性卒中风险为 0.90 OR (95% CI 0.74-1.09)。研究结果提示循环 FGF23 水平与 LAS 风险可能存在因果关联, MR-IVW 估计为 1.74 OR (95% CI 1.08-2.81), MR-WME 估计为 1.76 OR (95% CI 1.04-2.99), Cochran Q = 0.02 ( $P = 0.993$ ); 留一法提示移除单个遗传工具对结果的估计效应影响在 -0.13% 至 5.8% 间, CYP24A1 相关遗传变异在 FGF23 与 LAS 的因果关联中起到驱动作用。工具变量 Egger 截距为 -0.001, 未提示存在定向多效性 ( $P = 0.145$ )。因果方向未提示存在矛盾 (MR-Steiger,  $P = 3.20 \times 10^{-14}$ )。所获得的 3 个 FGF23 相关遗传工具共解释了 0.53% 的循环 FGF23 水平变化, FGF23 与 LAS 的因果估计统计效能为 94%。

**结论:** 本研究未发现遗传增加的循环 25OHD 水平与缺血性卒中或其亚型降低的风险存在遗传学因果关联, 因此不支持也不反对在普通人群中补充维生素 D。我们的结果提示遗传倾向于增高的循环 FGF23 水平与增加的大动脉粥样硬化性卒中风险存在潜在的遗传学因果关联, 其中的机制可能与维生素 D 和  $\alpha$ Klotho 相关。在 LAS 风险人群 (如颅内动脉粥样硬化) 中适当补充维生素 D 可能有益。FGF23 与 LAS 的相关机制以及  $\alpha$ Klotho、维生素 D 在其中的作用需要进一步的实验评估。

**关键词:** 缺血性卒中 成纤维细胞生长因子 23 孟德尔随机化 维生素 D 动脉粥样硬化

## Abstract

**Backgrounds:** The increasing global burden of stroke requires a more comprehensive approach for disease prevention; whether the modifiable vitamin D has protective effects on ischemic stroke or subtypes remains a question. Fibroblast growth factor 23 (FGF23) is an emerging independent risk factor on cardiovascular diseases, but its association with ischemic stroke are not consistent. Besides, the causalities between FGF23 and ischemic stroke or subtypes are unclear. Considering the crucial role of FGF23 on vitamin D levels through inhibiting activation and promoting degradation, the function of FGF23 was thought partly dependent on its effects on vitamin D. Thus, vitamin D supplements may benefit for specific groups.

**Objective:** We applied the Mendelian randomization (MR) framework to infer the causalities of circulating vitamin D (25OHD, 25-hydroxyvitamin D) levels and FGF23 levels on ischemic stroke and subtypes, to investigate potential effects of vitamin D supplement among different groups.

**Methods:** The circulating 25OHD genome-wide association study (GWAS) data (N=417,580, the white British group) from the UK BioBank, 5 SNPs from a seven cohort GWAS meta-study associated with serum FGF23 levels, and the summary statistics of acute ischemic stroke (34,217 cases, 406,111 controls) and subtypes in MEGASTROKE study (European ancestry) were chosen to construct instrumental variants. We filtered genetic variants by association ( $P < 5.0E-08$ ), linkage disequilibrium ( $LD\ r^2 < 0.1$ ), possible pleiotropic associations and instrument strength. The inverse-variance weighted method (MR-IVW) was picked to investigate the causal relationship of 25OHD and FGF23 levels on ischemic stroke or subtypes. The weighted-median method (MR-WME), Cochran'Q, and the leave-one-out analysis were employed to evaluate the robustness of causal estimates. Moreover, the Egger intercept test and the PRESSO (Pleiotropy RESidual Sum and Outlier) test were utilized on sensitive analysis for potential horizontal pleiotropies; and the MR-Steiger method was utilized to evaluate the vertical direction of the causal estimate. Furthermore, we calculate the power of the causal estimate.

**Results:** We did not find causal relationships of 25OHD with acute ischemic stroke

(AIS) or subtypes. This study used 60 filtered instrument variables to estimate (MR-IVW) an OR (Odds Ratio) of 1.02 for acute ischemic stroke with a 95% confidence interval (95%CI) of 0.94-1.12, and 0.94 OR (95% CI 0.76-1.16) for large artery atherosclerotic stroke (LAS) per SD (standard deviation) increased 25OHD. Three instruments passed the filtering. There were no causal relationships of FGF23 on acute ischemic stroke (MR-IVW; 0.90 OR, 95%CI 0.74-1.09;  $P = 0.271$ ). We found a suggestive genetic causal association of FGF23 on LAS (MR-IVW; 1.74 OR, 95% CI 1.08-2.81), and the MR-WME result was 1.76 OR (95% CI 1.04-2.99). There was no Cochran heterogeneity ( $Q = 0.02$ ,  $P = 0.993$ ); leave-one-out analysis showed that the effects of each genetic variables on causal estimate were perturbations (-0.13% to 5.8%), and the validated variant near the CYP24A1 gene drove the causal association of FGF23 and LAS. No horizontal pleiotropy was found (MR-Egger intercept -0.001,  $P = 0.145$ ), and the causal direction was FGF23 to LAS with no conflict (MR-Steiger,  $P = 3.20 \times 10^{-14}$ ). Three genetic instruments accounted for 0.53% circulating FGF23 levels, and the causal estimate power of FGF23 on LAS was 94%.

**Conclusion:** Our study did not find a genetic predisposition of a higher circulating 25OHD associated with a lower risk of acute ischemic stroke or subtypes; thus, we neither support nor reject vitamin D supplementation in the general population. We found a genetic predisposition of increased FGF23 associated with a higher risk of large arterial atherosclerotic stroke. Our results suggested a beneficial effect of vitamin D supplements in people with a high risk of LAS (like intracranial atherosclerosis). Further studies should investigate the associated mechanisms of FGF23 in LAS, and the roles of vitamin D and  $\alpha$ Klotho in the association of FGF23 and LAS.

**Key words:** Ischemic stroke, Fibroblast growth factor 23, Mendelian randomization, Vitamin D, Atherosclerosis

## 目 录

中文摘要 .....	I
Abstract.....	III
缩略语/符号说明 .....	VII
前言 .....	1
1. 研究现状、成果 .....	1
2. 研究目的、方法 .....	2
研究对象和方法 .....	3
1. 研究对象 .....	3
1.1 血清 25OHD GWAS 数据集 .....	3
1.2 血清 FGF23 GWAS 数据集 .....	3
1.3 缺血性卒中 GWAS 数据集 .....	4
2. 研究方法 .....	6
2.1 MR 研究原理 .....	6
2.2 MR 遗传工具评估 .....	8
2.3 MR-IVW 因果估计 .....	9
2.4 MR 敏感性分析 .....	10
研究结果 .....	13
1. 循环 25OHD 与缺血性卒中的因果关联 .....	13
1.1 循环 25OHD 相关工具变量 .....	13
1.2 循环 25OHD 与 AIS 的因果估计 .....	17
1.3 循环 25OHD 与卒中亚型的因果估计 .....	18
2. 循环 FGF23 与缺血性卒中的因果关联 .....	20
2.1 循环 FGF23 相关工具变量 .....	20
2.2 循环 FGF23 与 AIS 的因果估计 .....	21
2.3 循环 FGF23 与卒中亚型的因果估计 .....	23
讨论 .....	25
1. 与其他研究比较 .....	25
1.1 缺血性卒中亚型差异 .....	25
1.2 25OHD 相关比较 .....	26



1.3 FGF23 相关比较 .....	27
2. FGF23 相关机制 .....	28
2.1 内皮损伤与免疫炎症 .....	28
2.2 血管钙化与血管硬化 .....	29
2.3 维生素 D 相关 .....	30
3. 临床转化 .....	33
4. 优势与局限 .....	34
5. 进一步研究方向 .....	35
参考文献 .....	37
附录 .....	53
附表 1 血清 25OHD 显著相关遗传工具（移除） .....	53
附表 2 MEGASTROKE 缺血性卒中及其亚型匹配的 FGF23 关联效应 .....	56
附表 3 FGF23 相关遗传变异可能的水平多效性 .....	57
附图 1 循环 25OHD 与急性缺血性卒中单遗传工具因果估计 .....	58
附图 2 循环 25OHD 与 LAS 留一法分析 .....	59
综述 .....	60
摘要 .....	60
1. 概述 .....	60
2. 关联研究 .....	61
2.1 全基因组关联研究 .....	61
2.2 全转录组关联研究 .....	62
3. 因果关联 .....	62
4. 网络分析 .....	64
5. 深度学习 .....	66
6. 展望与小结 .....	66
综述参考文献 .....	68
致谢 .....	78
个人简历 .....	80

## 缩略语/符号说明

英文缩写	英文全称	中文全称
25OHD	25-hydroxyvitamin D	25-羟基维生素 D
RINT	Rank-based inverse-normal transformation	基于秩的逆正态转换
COJO	Conditional and joint	条件联合（分析）
VDR	Vitamin D receptor	维生素 D 受体
FGF23	Fibroblast growth factor 23	成纤维细胞生长因子 23
cFGF23	COOH-terminal FGF23	FGF23 羧酸片段
iFGF23	Intact FGF23	FGF23 完整片段
AIS	Acute Ischemic stroke	急性缺血性卒中
LAS	Large artery atherosclerotic stroke	大动脉粥样硬化性卒中
CES	Cardioembolic stroke	心源性脑栓塞性卒中
SVS	Small-vessel stroke	小血管性卒中
TOAST	Trial of Org 10172 in Acute Stroke Treatment	Org 10172 低分子类肝素急性卒中治疗临床试验
BMI	Body mass index	体质指数
HIF-1 $\alpha$	Hypoxia-inducible factor-1 alpha	低氧诱导因子 1 $\alpha$
TNF- $\alpha$	Tumor necrosis factor alpha	肿瘤坏死因子 $\alpha$
CKD	Chronic kidney disease	慢性肾脏病
GWAS	Genome-wide association Studies	全基因组关联研究
SNP	Single-nucleotide polymorphism	单核苷酸多态性
LD	Linkage Disequilibrium	连锁不平衡
MR	Mendelian Randomization	孟德尔随机化（框架）
RCT	Randomized controlled trial	随机对照试验
CI	Confidence interval	置信区间
SE	Standard Error	标准误

## 缩略语/符号说明

英文缩写	英文全称	中文全称
OR	Odds Ratio	比值比
SD	Standard deviation	标准差
EAF	Effect allele frequency	等位基因频率
IVW	Inverse-variance weighted	逆方差加权法
WME	Weighted median estimate	中位数加权法
InSIDE	INstrument Strength Independent of Direct Effect	工具优势独立于直接效应 (假设)
NOME	NO Measument estimate	无测量估计
PRESSO	Pleiotropy RESidual Sum and Outlier test	多效性残差和与异常值检测
GLGC	Global Lipids Genetics Consortium	全球脂质遗传协会
MAGIC	Meta-Analyses of Glucose and Insulin-related traits Consortium	葡萄糖荟萃分析与胰岛素相关特征协会
DIAGRAM	Diabetes Genetics Replication and Meta-analysis	糖尿病遗传重复与荟萃研究
GIANT	Genetic Investigation of Anthropometric Traits	人体测量特征的遗传探索
ICBP	International Consortium of Blood Pressure	国际高血压协会
ARIC	Atherosclerosis Risk in Communities Study	社区动脉粥样硬化风险研究
CHS	Cardiovascular Health Study	心血管健康研究
SUNLIGHT	Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits	维生素 D 的潜在遗传决定因素和高度相关性状的研究
VITAL	VITamin D and OmegA-3 Trial	维生素 D 和 $\omega$ -3 脂肪酸(补充) 试验

## 前言

### 1. 研究现状、成果

卒中是世界上死亡率、致残率最高的疾病之一<sup>[1-2]</sup>。在美国每年约有超过 795,000 例发生卒中事件，每年卒中相关估计经济损失超过三十三亿美元<sup>[3]</sup>。在中国，卒中是死亡率的首要病因，在世界卒中死亡人数中占据了高达三分之一的比重<sup>[4]</sup>。缺血性卒中是最常见的类型，在中国约占 69.6% 的卒中比例<sup>[3-4]</sup>。逐渐增加的全球负荷要求不同风险层次的人群进行卒中的综合预防<sup>[5-6]</sup>。缺血性卒中与多种因素相关：经典的危险因素包括高血压、糖尿病、血脂异常、高体质指数（body mass index, BMI）、高腰臀比、吸烟、酗酒等<sup>[7-9]</sup>；生活方式与心理社会因素中还包括高盐饮食<sup>[10]</sup>、压力<sup>[11]</sup>、抑郁<sup>[12]</sup>等；此外，炎症相关因子白介素 6、单核细胞趋化因子-1 等也与高卒中风险相关<sup>[13-16]</sup>。复杂的表型关联可能是混杂因素关联，也可能提示着潜在的疾病干预靶点<sup>[17]</sup>。

近年来，研究者们对维生素 D 缺乏与卒中发病风险之间的关联产生了极大的兴趣。循环维生素 D 水平在人群中广泛缺乏<sup>[18-19]</sup>。作为一种可调节的固醇类激素前体，维生素 D 除了钙磷调节的功能之外，还参与多种生物调节过程，如细胞生长，神经肌肉调节、免疫调节等<sup>[20-21]</sup>。流行病学研究指出较低的维生素 D 水平与较高的卒中风险相关。Zhou 等人通过系统性回顾纳入 19 个研究进行荟萃分析，结果显示维生素 D 缺乏与增加的缺血性卒中风险存在流行病学关联，相对危险度（RR, relative risk）为 1.62，95% 置信区间（confidence interval, CI）为 1.34-1.96，其中与缺血性卒中风险的关联为 RR 2.45（95% CI 1.56-3.86）<sup>[22]</sup>。此外，较低的血清 25-羟基维生素 D（serious 25-hydroxyvitamin D, 25OHD）也与缺血性卒中较高的复发率以及较差的预后相关<sup>[23-25]</sup>。

然而人群中维生素 D 补充的随机对照试验（randomized controlled trial, RCT）未能证实口服补充维生素 D 对卒中具有保护性作用<sup>[26-27]</sup>。社区人群每日高浓度补充维生素 D3 在卒中发病上未显示获益<sup>[26]</sup>。Barbarawi 等研究者综合 21 个 RCT 研究发现补充维生素 D 并未降低卒中终点发生率（RR 1.06；95% CI 0.98-1.15； $P = 0.16$ ）<sup>[27]</sup>。观测性研究与维生素 D 补充的随机对照研究之间的矛盾性结果对维生素 D 与卒中之间是否具有潜在的因果关联提出了疑问。

成纤维细胞生长因子 23 (FGF23, Fibroblast growth factor 23) 是心血管疾病的独立危险因素, 与卒中存在显著的流行病学关联<sup>[28-29]</sup>。此外, 有研究指出循环 FGF23 水平也与亚临床脑血管损伤<sup>[30]</sup>、颈动脉斑块高负荷<sup>[31]</sup>等表型相关。FGF23 与缺血性卒中的关联并不一致, FGF23 与缺血性卒中之间是否存在因果关联也不明确<sup>[32-35]</sup>。一项 149 人(平均估算的肾小球滤过率  $36 \text{ ml} \cdot \text{min}^{-1} \cdot 1.73 \text{ m}^2$ ) 的研究发现, FGF23 与卒中风险显著相关<sup>[35]</sup>。另一项荟萃分析结果显示高水平 FGF23 与卒中的发生存在显著性正相关, RR 1.29 (95% CI 1.10-1.52), 与缺血性卒中的风险不具有统计学显著性, RR 为 1.12 (95% CI 0.92-1.37) <sup>[32]</sup>。评估 FGF23 水平与缺血性卒中及其亚型风险之间的关联性质有助于寻找可能的疾病干预靶点, 优化卒中的诊疗方案。

FGF23 分子是一种骨源性激素, 与多种生理病理过程相关<sup>[36]</sup>。FGF23 主要作用于肾脏、甲状旁腺, 是钙、磷调节的核心因子, 且参与维生素 D 水平调节。FGF23 可通过肾脏远端肾小管内化  $\text{Na}^+$  依赖的转运受体减少磷的重吸收, 并通过改变甲状旁腺激素以及活性维生素 D 水平调节钙磷的平衡。此外, FGF23 还可能在心肌重构、血管钙化、免疫炎症等过程中扮演着重要的角色。考虑到 FGF23 可以通过减少维生素 D 活化并且促进活性维生素 D 降解, 有观点认为 FGF23 对组织的损伤可能部分是通过降低活性维生素 D 水平形成<sup>[36-37]</sup>。

## 2. 研究目的、方法

因此, 本研究期望通过分析循环维生素 D 和 FGF23 水平与卒中或其亚型风险的因果关联性质, 以初步评估在不同人群中补充维生素 D 的可能效应, 以期进一步优化卒中的综合管理方案。

本研究首先利用与循环 25OHD、FGF23 水平独立并且相关的全基因组关联研究 (GWAS, Genome-wide association study) 的统计摘要数据 (欧洲裔) 构建遗传工具, 通过两样本的孟德尔随机化 (Mendelian Randomization, MR) 框架来探索遗传倾向于增高的 25OHD 以及 FGF23 分子与卒中及其亚型风险的关联性质, 并利用不同方法评估因果推断的稳健性, 水平多效性, 因果关联方向、统计效能以综合分析因果估计的效应。然后, 本研究回顾并讨论不同人群中维生素 D 补充的可能效应。

## 研究对象和方法

### 1. 研究对象

#### 1.1 血清 25OHD GWAS 数据集

循环 25OHD 遗传工具来源于 Joana A Revez 提供的欧洲裔 GWAS 数据<sup>[38]</sup>。本文选取经过 COJO (a conditional and joint) 分析相互独立且 COJO 分析前后均显著相关 ( $P < 5.00E-08$ ) 的遗传变异构建工具变量<sup>[38]</sup>。选取的遗传工具限定在常染色体双等位基因的遗传变异, 共 113 个。

该数据样本 ( $N = 417,580$ ) 筛选自 UK Biobank 英国生物库, 样本来源人群在 40-69 岁之间。血清 25OHD 通过 Diasorin 方法检测。协变量校正年龄、性别、月份、检测中心, 前 40 个主要成分以及批次效应、饮食补充维生素 D。25OHD 基因型数据经质控后含 8,806,780 个遗传变异, 其次要等位基因频率大于 1% 并遵循 Hardy-Weinberg 平衡定律 ( $P > 1.00E-05$ )。效应值基于秩的逆正态转换 (RINT, rank-based inverse-normal transformation) 标准化, 1 个标准差 (standard deviation, SD) 的 RINT (25OHD) 约为 21 nmol/L 25OHD<sup>[38]</sup>。

#### 1.2 血清 FGF23 GWAS 数据集

成纤维细胞生长因子 23 相关遗传工具从一项 GWAS 荟萃研究中获取, 该研究中共包含 7 个队列, 16,624 例欧洲裔样本 (表 1)。队列中循环 FGF23 通过酶免疫耦联检测获得。由于肾脏疾病对 FGF23 的作用可能掩盖了单个遗传变异对 FGF23 水平的影响, 该研究排除估计肾小球滤过率小于  $30\text{ml/min} \cdot 1.73\text{m}^2$  的人群, 提高遗传因素对表型效应的评估准确性<sup>[39]</sup>。

我们选取第一阶段的 GWAS 数据, 以模型一的效应值 (校正年龄、性别及前 10 个种族相关主要成分) 进行后续分析。该研究中共发现 5 个与 FGF 显著相关 ( $P < 5.00E-08$ ) 且相互独立的遗传变异, 分别位于 CYP24A1、ABO、RGS14、LINC01506、LINC01229 基因附近 (详见表 6)<sup>[39]</sup>。效应值为自然对数转换后的 FGF23 ( $\log\text{ pg/ml}$ )。我们用 Quan 所提出的方法计算 1 个 SD 的 FGF23 水平, 即利用  $\log$  转换前的分布与  $\log$  转换后的正态分布进行换算, 计算 1 个 SD 的  $\log$  FGF23 在转换前所对应的值<sup>[39-40]</sup>。

表 1 血清 FGF23 相关 GWAS 荟萃研究队列

Characteristic	N	eGFR (CKD-EPI) ml/min per 1.73 m <sup>2</sup>	FGF23 (Intact) pg/ml	FGF23 (C-terminal) pg/ml
ARIC	8594	87.8 (13.2)	45.5 (17.3)	-
CHS	1988	72.1 (17.7)	-	12.9 (21.3)
Indiana	1128	103.5 (7.4)	39.4 (17.4)	-
MESA	2163	76.0 (17.1)	41.9 (17.6)	-
MrOS GBG	937	71.6 (19.3)	46.2 (22.5)	-
MrOS Malmo	894	68.8 (19.3)	55.9 (27.7)	-
OPRA	920	68.0 (12.6)	-	8.3 (9.1)

注：FGF23 检测方式包括完全的 FGF23 分子以及 C-端片段两种检测方式。ARIC, the Atherosclerosis Risk in Communities Study; CHS, the Cardiovascular Health Study; Indiana, the Indiana Sisters Cohort; MESA, the Multi-Ethnic Study of Atherosclerosis; MrOS GBG, the Osteoporotic Fractures in Men Study – Goteborg; OPRA, the Osteoporosis Prospective Risk Assessment Study; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration equation。数据以均值（标准差）的形式表示，“-”表示无数据。

### 1.3 缺血性卒中 GWAS 数据集

本研究结局性状为急性缺血性卒中（acute ischemic stroke, AIS）或其亚型，效应值（log OR; OR, odds ratio）源自 MEGASTROKE 协会的 GWAS 数据<sup>[41]</sup>。

MEGASTROKE 协会是由国际卒中遗传学协会发起的大型国际组织，综合来自不同国家的 25 个卒中相关的 GWAS 队列，首次发表于 2018 年。该数据集是迄今为止数据量最大、最全面的卒中相关遗传变异综合数据集<sup>[41]</sup>。为满足暴露与结局的数据来源人群的遗传背景一致，我们摘取其中欧洲裔的相关 GWAS 摘要数据进行因果推断。该数据由 17 个队列构建（见表 2），各队列内校正年龄、性别等协变量信息，所选取的欧洲裔共包括卒中人群 40,585 例，对照人群 406,111 例。

急性缺血性卒中的诊断依据世界卫生组织的定义，即出现迅速发展的局灶性或全局性脑功能障碍的临床表现，持续时间超过 24 小时或导致死亡，除血管因素外没有其它明显的病因，结合临床表现及影像诊断。MEGASTROKE 协会根据 TOAST（Trial of Org 10 172 in Acute Stroke Treatment）分型，进一步区分

缺血性卒中亚型。一共包含（欧洲裔）大动脉粥样硬化性卒中（large artery atherosclerotic stroke, LAS）4,373 例、心源性脑栓塞性卒中（cardioembolic stroke, CES）7,193 例、小血管性卒中（small-vessel stroke, SVS）5,386 例<sup>[41]</sup>。

表 2 缺血性卒中 MEGASTROKE 协会 GWAS 荟萃研究队列

Study	Stroke	AIS	LAS	CES	SVS	Control
Barcelona	520	520	119	215	-	315
CADISP	555	555	67	211	31	9,259
CHARGE	4,348	3,028	-	602	-	80,613
deCODE	5,520	4,483	512	1,346	615	255,213
EPIC-CVD	4,347	2,226	-	-	-	7,897
FINLAND	501	501	99	187	62	1,813
Glasgow	599	599	72	105	137	1,775
HVH1	681	577	62	92	175	1,331
HVH2	124	103	-	28	39	570
ICH	1,545	-	-	-	-	1,481
INTERSTROKE EUR	911	826	188	208	243	863
MDC	202	202	-	-	-	4,295
METASTROKE	10,307	10,307	1,817	1,859	1,349	19,326
SAHLIS	298	298	-	35	85	596
SIFAP	981	981	184	170	104	1,824
SiGN	7,743	7,743	1,253	2,135	1,534	17,970
Young Lacunar DNA	1,403	1,268	-	-	1,012	970
Total European	40,585	34,217	4,373	7,193	5,386	406,111

注：MEGASTROKE 欧洲裔数据集包含 17 个队列。AIS、LAS、CES、SVS 分别表示急性缺血性卒中（acute ischemic stroke）、大动脉粥样硬化性卒中（large artery atherosclerotic stroke）、心源性脑栓塞性卒中（cardioembolic stroke）以及小血管性卒中（small-vessel stroke, SVS）。



当遗传变异在结局变量的 GWAS 摘要数据中无法获取时，我们以千人基因组计划 3 期（“EUR”）的数据为参考<sup>[42]</sup>，选取与该遗传位点高度连锁不平衡（Linkage Disequilibrium, LD）（ $r^2 > 0.80$ ）并且距离在 $\pm 20\text{kb}$  范围的遗传变异作为替代工具<sup>[43-44]</sup>。

## 2. 研究方法

### 2.1 MR 研究原理

孟德尔随机化设计最早由 Katan MB 提出，被广泛用于探索表型之间的因果关联<sup>[45-47]</sup>。该框架遵循孟德尔遗传自由分离定律，认为在群体中亲代的等位基因随机分配给子代，因此是一种天然的随机对照设计。同时，在合子形成时，个体的基因就已确定，基于基因型决定表型的假设，MR 框架可以有效避免混杂因素和逆向因果的影响，以评估表型之间的关联性质。因此，通过利用暴露因素（循环 25OHD、FGF23 水平）相关单核苷酸多态性（Single-nucleotide polymorphism, SNP）位点构建遗传工具，可推断遗传倾向于增加的暴露因素与结局变量（卒中或其亚型）之间是否存在因果关联<sup>[48]</sup>。

此外，孟德尔随机化框架有助于弥补随机对照试验的不足。随机对照试验常常由于伦理等因素限制难以实行，MR 框架不失为一个经济有效的因果评估方法<sup>[49]</sup>。此外，孟德尔随机化框架也有助于药物疗效<sup>[50-51]</sup>以及副作用<sup>[52]</sup>的评估，这有助于减少 RCT 研究不必要的经济损失并且提高药物靶点筛选的准确性。随着二代测序的快速发展，以及测序成本的下降，生物序列数据快速增多，这为遗传关联结构的探索提供了坚实的基础，推动了基于算法评估表型间因果关联性质的进程<sup>[53]</sup>。

孟德尔随机化因果估计是一种有向无环图的推断（图 1）。准确的因果推断需满足三大工具变量假设（图 1）：① 工具变量与暴露因素显著相关（相关性）；② 工具变量与暴露、结局的混杂因素不相关（独立性）；③ 工具变量仅仅通过暴露影响结局（排他性）<sup>[48]</sup>。第一个假设可以直接通过相关性检验。由于混杂因素的未知性，第二个假设在 MR 研究中难以检验，通常要求暴露与结局二者的工具变量遗传背景一致<sup>[54]</sup>。第三个假设与工具变量的多效性相关，遗传工具对结局变量（卒中或其亚型）的作用为通过暴露因素对结局的间接效应以及其通过其他途径对结局的水平多效性的效应之和。

对基于统计摘要的两样本孟德尔随机化研究而言，如果暴露相关的 GWAS 来源人群与结局相关的 GWAS 来源人群无重合，则遗传变异  $G_j$  对暴露、结局的效应 ( $\hat{\gamma}_j$ 、 $\hat{\Gamma}_j$ ) 相互独立。若每个遗传工具的 Wald 比率估计量  $\hat{\theta}_j$  线性且均质，则  $\hat{\gamma}_j$ 、 $\hat{\Gamma}_j$  符合正态分布  $\hat{\gamma}_j \sim N(\gamma_j, \sigma_{X_j}^2)$ 、 $\hat{\Gamma}_j \sim N(\hat{\theta}\gamma_j + a_j, \sigma_{Y_j}^2)$  ( $\sigma_{X_j}^2$ 、 $\sigma_{Y_j}^2$  分别为 SNP-暴露、SNP-结局的方差)， $\hat{\theta}_j$  回归于暴露因素与结局变量的因果关联效应量  $\theta$ <sup>[55]</sup>。公式中的  $U_x$ 、 $U_y$ 、 $\varepsilon_x$ 、 $\varepsilon_y$  分别表示混杂因素和其它因素对暴露、结局变量的影响 (公式 1，公式 2)。

$$X = \sum_{j=1}^J \gamma_j G_j + U_x + \varepsilon_x, \quad j=1, \dots, J \quad (1)$$

$$Y = \theta X + \sum_{j=1}^J \alpha_j G_j + U_y + \varepsilon_y \quad (2)$$

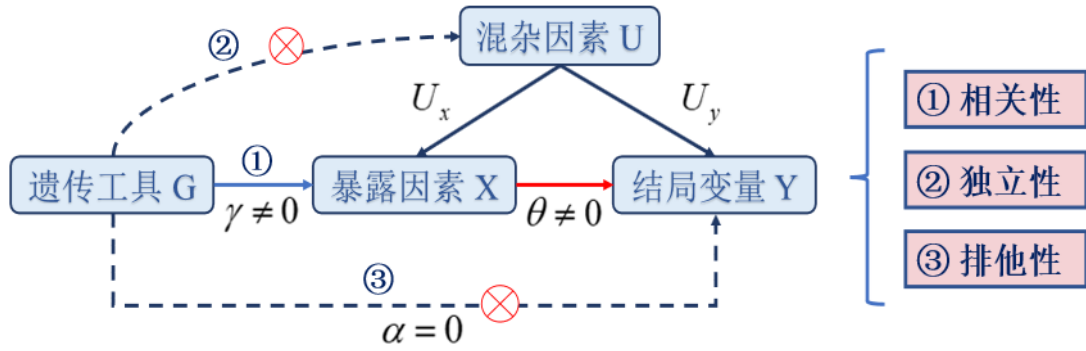


图 1 孟德尔随机化设计框架原理图

图注：孟德尔随机化因果推断要求遗传变异  $G$  为有效的工具变量，即满足工具变量三大假设：① 相关性 ( $\gamma \neq 0$ )，② 独立性，③ 排他性 ( $\alpha = 0$ )。在此基础上，若  $\theta \neq 0$ ，则可推断暴露因素  $X$  对结局变量  $Y$  存在因果效应。其中  $\alpha$ 、 $\theta$ 、 $\gamma$ 、 $U_x$ 、 $U_y$  分别表示遗传工具对结局变量的直接效应、暴露因素与结局变量的因果估计量、遗传变异对暴露因素的效应量以及混杂因素对暴露因素以及结局变量的效应量。

本文首先评估遗传工具的有效性，然后以随机效应的逆方差加权法 (MR-IVW, Inverse-variance weighted) 分析循环 25OHD、FGF23 水平与缺血性卒中及其亚型的关联性质<sup>[56]</sup>，并通过评估结果的稳健性、遗传工具的水平多效性、垂直方向以及统计效能判断结果的准确性 (图 2)。本研究通过 R 3.6.2 平台评估因果关联性质，所使用的工具以及相应的用途可见工具汇总 (表 3)。

表 3 使用工具汇总

Name	Usage	Version	Citation
TwoSampleMR	数据合并	0.5.1	[57]
MendelianRandomization	因果关联估计	0.4.1	[58]
MRPRESSO	ME-PRESSO 检测	1.0	[59]
LDlinkR	检索连锁不平衡关系	1.0.1	[60]
dplyr	数据提取	0.8.5	[61]
UpSetR	多重集合可视化	1.4.0	[62]

## 2.2 MR 遗传工具评估

本研究选取与暴露因素（血清 25OHD 水平、血清 FGF23 水平）显著相关（ $P < 5.00E-08$ ）且相互独立的遗传变异位点作为工具变量。我们通过千人基因组计划 3 期欧洲裔的基因型参考面板评估遗传工具的独立性，要求连锁不平衡关系  $LD\ r^2 < 0.1$ 。

在此基础上，我们借鉴 Larsson 等人的方案对遗传工具的水平多效性进行三阶段评估（见图 2）<sup>[63]</sup>。首先，我们通过 PhenoScannerV2 数据库检测与卒中相关的遗传工具（Bonferroni 多重校正， $P < 0.05/m$ ， $m$  表示工具变量数）<sup>[64-65]</sup>（检索时间 2020 年 3 月）。同时，我们选用大型公开的 GWAS 数据集以筛选与卒中相关混杂因素存在多效性关联的遗传工具<sup>[63]</sup>。我们通过 GLGC（Global Lipids Genetics Consortium）对主要血脂进行评估<sup>[66]</sup>；通过 MAGIC（Meta-Analyses of Glucose and Insulin-related traits Consortium）评估空腹血糖、空腹胰岛素的关联性<sup>[67]</sup>；通过 DIAGRAM（Diabetes Genetics Replication and Meta-analysis）对 2 型糖尿病进行评估<sup>[68-69]</sup>；通过 GIANT（Genetic Investigation of Anthropometric Traits）对 BMI、腰臀比进行评估<sup>[70-71]</sup>；通过 ICBP（International Consortium of Blood Pressure）协会对收缩压、舒张压进行评估<sup>[72]</sup>；此外我们以 Neale UK Biobank 的 GWAS 数据对上述表型以及吸烟、酗酒表型进行评估<sup>[73]</sup>。

若所选择的遗传变异与上述因素存在显著性关联，那么这个遗传变异可能与缺血性卒中或者其亚型存在多效性关联<sup>[74]</sup>。在此基础上，我们评估暴露与结局的因果关联性质。然后，我们使用 MR-Egger 方法<sup>[75]</sup>以及 MR-PRESSO（Pleiotropy RESidual Sum and Outlier test）方法<sup>[59]</sup>评估潜在的水平多效性。

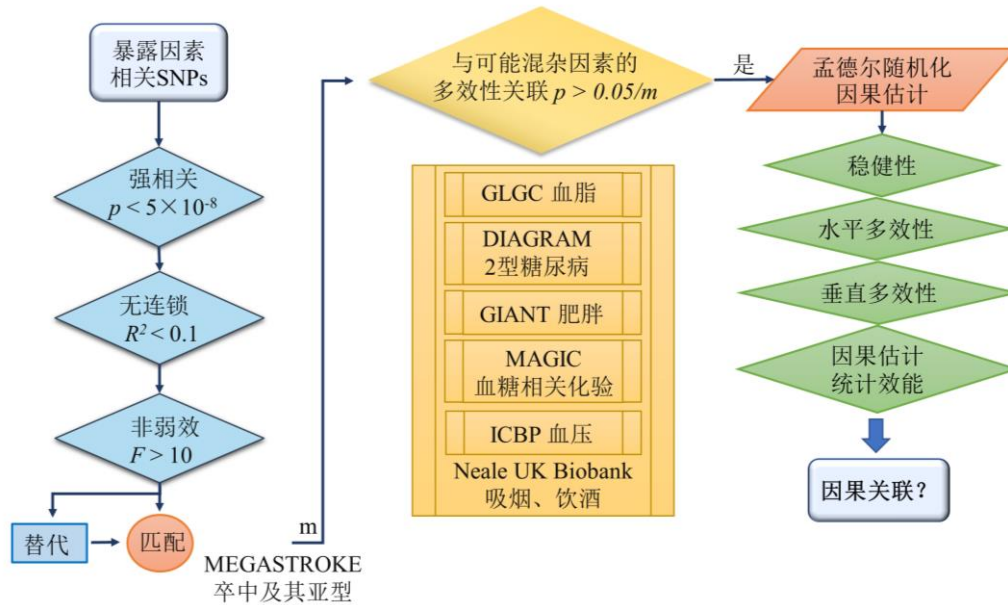


图2 孟德尔随机化研究分析流程图

图注：LD 表示连锁不平衡。GLGC（Global Lipids Genetics Consortium）<sup>[66]</sup>、MAGIC（Meta-Analyses of Glucose and Insulin-related traits Consortium）<sup>[67]</sup>、DIAGRAM（Diabetes Genetics Replication and Meta-analysis）<sup>[68-69]</sup>、GIANT（Genetic Investigation of Anthropometric Traits）<sup>[70-71]</sup>和 ICBP（International Consortium of Blood Pressure）<sup>[72]</sup>分别表示源自全球脂质遗传协会、葡萄糖荟萃分析与胰岛素相关特征协会、糖尿病遗传重复与荟萃研究、人体测量特征的遗传探索以及国际高血压协会的大型全基因组荟萃研究数据。血糖相关化验包括空腹血糖与空腹胰岛素，肥胖包括体质指数与腰臀比；SNPs（Single-nucleotide polymorphisms）表示单核苷酸多态性位点。

当工具变量对暴露遗传解释强度低时，工具变量对暴露的解释方差小于 SNP-暴露关联的标准误可解释的方差，那么 SNP-暴露的相关性有不可忽略的不确定性，此时该工具变量称为弱工具变量。在工具变量筛选过程中，我们利用均值 F 检验评估每个遗传变量的工具强度，要求每一个遗传工具 F 大于 10<sup>[76]</sup>，即所有采用的遗传工具均非弱工具变量。

### 2.3 MR-IVW 因果估计

本文使用逆方差加权法（MR-IVW）评估循环 25OHD、FGF23 水平与缺血性卒中或其亚型风险的遗传因果关联。为方便结果的临床转换，我们将结果标准化为增加 1 个 SD 暴露因素（25OHD、FGF23 水平）相应的卒中风险（OR）。

MR-IVW 方法考虑两样本 MR 的因果效应量为大样本的回归估计，被广泛用于 MR 研究的因果关联评估<sup>[46]</sup>。因果估计量为多个工具变量的逆方差荟萃估计量  $\hat{\theta}_{IVW}$ （公式 3）；MR-IVW 假设  $\hat{\gamma}_j / \hat{\gamma}_j$  分母的不确定性可以忽略（ $\hat{\gamma}_j \approx \gamma_j$ ），方差  $v_{IVW}$  用泰勒展开式的一阶项估算（公式 4）<sup>[76]</sup>。随机效应模型在因果推断中考虑残差项估计量的影响， $\sigma_\varepsilon^2$  为因果效应线性回归的残差的方差。若  $\sigma_\varepsilon^2$  为 1（公式 4），则为 MR-IVW 的固定效应模型；若每个遗传变量均无水平多效性，即  $\alpha_j = 0$  对所有遗传工具均成立，则固定效应的估计结果与随机效应一致<sup>[77]</sup>。当遗传工具的数目较多时，本研究使用随机效应的模型评估因果关联；当遗传工具数目较少时，使用固定效应的模型评估因果关联。

$$\hat{\theta}_{IVW} = \frac{\sum_j w_j \hat{\theta}_j}{\sum_j w_j}, \quad w_j = \frac{\hat{\gamma}_j^2}{\sigma_{Y_j}^2} \quad (3)$$

$$v_{IVW} = \frac{\sigma_\varepsilon^2}{\sum_j \hat{\gamma}_j^2 \sigma_{Y_j}^{-2}} \quad (4)$$

## 2.4 MR 敏感性分析

MR 推断结果的有效性依赖于严格的工具变量假设及建模过程中的假设。我们通过加权的中位数（MR-WM, Weighted median）法评估多数工具变量下的因果效应<sup>[78]</sup>，并结合 Cochran Q、留一法评估结果的稳健性。在此基础上，通过 MR-Egger 法<sup>[75]</sup>及 MR-PRESSO 法<sup>[59]</sup>评估潜在的水平多效性，通过 MR-Steiger 法评估 MR 因果关系的垂直方向<sup>[79]</sup>。最后，基于 SNP 的遗传力计算 MR 估计的统计效能<sup>[56,80]</sup>。

### 2.4.1 稳健性分析

我们结合 MR-WME 法、Cochran Q 和留一法评估因果估计的稳健性<sup>[81]</sup>。MR-WME 在大部分遗传工具有效时，可获得一致的因果估计。MR-WME 法以所有遗传变异比率估计量的中位数评估因果效应，有效避免了极端值的影响，因此适用于部分遗传变异（< 50%）不符合工具变量假设时的因果估计<sup>[78]</sup>。MR-WME 法将各个遗传变异的逆方差权重标准化，并按照顺序进行排列以获得新的权重  $w'_j$ （ $j=1, \dots, J$ ）；通过计算前  $j$  项的权重之和  $s_j$ ，该方法可获得因果估计的经验性分布。MR-WME 法考虑 SNP 对因果估计量经验性分布的影响与

其权重  $w_j$  成正比, 因果估计量  $\hat{\theta}_{WM}$  为第  $j$  个有序比率估计量  $\hat{\theta}_j$  与其百分位数所构成的分布的中位数 (公式 5), 其中  $\lfloor J/2 \rfloor$  表示  $J/2$  的向下取整<sup>[78]</sup>。估计量  $\hat{\theta}_{WM}$  的置信区间可以通过参数引导方法获得。

$$\hat{\theta}_{WM} = \hat{\theta}_t + (\hat{\theta}_{t+1} - \hat{\theta}_t) \times \frac{50 - s_t}{s_{t+1} - s_t}, \quad t = \lfloor J/2 \rfloor \quad (5)$$

Cochran Q 检验可用于检验全局异质性。Cochran Q 为每一个工具变量的 Wald 比率估计量  $\hat{\theta}_j$  与  $\hat{\theta}_{IVW}$  之间的误差的加权平方和 (公式 6), 权重见公式 3。如果全局无异质性, 则 Cochran Q 符合自由度为  $J-1$  的卡方分布<sup>[82]</sup>。

$$Q = \sum_j Q_j = \sum_j w_j (\hat{\theta}_j - \hat{\theta}_{IVW})^2 \quad (6)$$

留一法评估单个 SNP 对评估结果的影响。该方法将  $J$  个遗传工具的 Wald 比率估计量分为 2 份,  $J-1$  份作为训练, 留下的 1 份作为测试, 迭代  $J$  次, 然后计算平均误差, 以评估因果估计结果的稳定性。若存在异质性, 则提示 1 个或多个遗传工具违背遗传工具假设<sup>[83]</sup>。

#### 2.4.2 水平多效性

若遗传工具存在多效性, 则难以获得一致的估计量。我们以 MR-Egger 法和 MR-PRESSO 法评估工具变量的潜在多效性。

MR-Egger 法将多效性引起的偏差近似为小型研究偏倚, 以检测因果推测中定向多效性的存在; 该方法在 MR-IVW 方法的基础上估计截距量, 允许截距不为 0<sup>[75]</sup>。MR-Egger 法借鉴荟萃研究中的 Egger 回归分析, 截距量  $\hat{\theta}_{0E}$  表示平均多效性; 斜率  $\hat{\theta}_E$  为减去偏倚后的因果估计量 (公式 7), 其方差可以通过公式 8 计算,  $\hat{\sigma}_{\varepsilon E}^2$  为残差  $\varepsilon_{E_j}$  的方差<sup>[77]</sup>。

MR-Egger 以较弱的 InSIDE 假设 (INstrument Strength Independent of Direct Effect), 即水平多效性与因果效应相独立, 来替代严格的工具变量第三假设<sup>[75]</sup>。即使所有的遗传工具都无效, 只要满足 InSIDE 假设, MR-Egger 可随着样本量及 SNP 数目的增大获得一致的因果估计<sup>[75]</sup>。InSIDE 假设经验上可通过评估遗传工具  $G_j$  的 SNP-结局相关性  $\alpha_j$  与 SNP-暴露相关性  $\gamma_j$  的关联大小分析, 要求二者的加权协方差为 0。即使所有的遗传变异都是无效的工具变量, Egger 检验也为因果假设提供了有效的检验与一致的因果效应估计, 常用于敏感性分析<sup>[75]</sup>。

当工具变量的效应对暴露的可解释方差弱时, MR-Egger 的截距估计结果将趋向于 0, 即 MR-Egger 平均多效性的估计量存在稀释。本文使用 Bowden 等人提出的  $I_{GX}^2$  异质性检验方法量化结果的稀释程度。若截距项无稀释, 即  $I_{GX}^2=1$ , 则截距可准确表示平均多效性<sup>[76,84]</sup>。当 InSIDE 假设满足时, 若截距  $\hat{\theta}_{0E}$  不为 0, 则所选取工具存在定向多效性; 如果  $\hat{\theta}_{0E}$  为 0, 则工具变量存在平衡多效性或者无水平多效性<sup>[77]</sup>。

$$\hat{\theta}_E = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j} - \frac{\hat{\theta}_{0E} + \varepsilon_{Ej}}{\hat{\gamma}_j}, \quad \varepsilon_{Ej} \sim N(0, \sigma_E^2 \sigma_{Y_j}^2) \quad (7)$$

$$v_E = \frac{\hat{\sigma}_{\varepsilon E}^2}{\sum_j (\hat{\gamma}_j - \hat{\theta}_E)^2 \sigma_{Y_j}^{-2}} \quad (8)$$

MR-PRESSO 基于多数有效的假设检测多效性, 常用于 MR 因果推断的水平多效性的异常值检测与校正<sup>[59]</sup>。该方法将因果估计量视为 SNP-结局关联对 SNP-暴露关联的回归, 允许遗传变异对暴露产生的扰动以及相应的结局观测值存在误差, 回归线斜率为因果效应估计量。MR-PRESSO 相对于 MR-Egger 截距测试检测水平多效性具有较低的假阳性率及较高的检测效能<sup>[59]</sup>。对于存在少数水平多效性 (<10%) 且满足 InSIDE 假设情况下的因果效应估计, MR-PRESSO 相比于 MR-IVW、多变量孟德尔随机化法具有更高的精确性。

MR-PRESSO 方法由全局水平多效性测试、异常值检测以及检测异常值移除前后因果估计显著性差异的失真检测三部分组成。水平多效性的全局测试中, MR-PRESSO 法计算移除一个 SNP 后其它遗传工具估计的该 SNP 的结局  $\hat{\theta}_{-j}\hat{\gamma}_j$  与实际  $\hat{\Gamma}_j$  的残差平方和 (residual sum of squares, RSS), 通过所有遗传变异计算全局误差观测值  $RSS_{obs}$  (公式 9); 然后评估以遗传工具-暴露模拟分布  $\hat{\gamma}_j^{random} \sim N(\hat{\theta}_{-j}\hat{\gamma}_j, \mathbb{V}(\hat{\gamma}_j))$  下的结局  $\hat{\theta}_{-j}\hat{\gamma}_j^{random}$  与  $\hat{\Gamma}_j^{random} \sim N(\hat{\theta}_{-j}\hat{\Gamma}_j, \mathbb{V}(\hat{\Gamma}_j))$ , 即 SNP-结局模拟分布的差异 (公式 10)。经验  $P$  值为  $K$  次无异常值模拟分布下观测误差低于期望的比例, 此处定义  $K=10,000$ 。对于多个异常点, 我们使用 Bonferroni 方法进行多重性校正<sup>[59]</sup>。

$$RSS_{obs} = \sum_j (\hat{\Gamma}_j - \hat{\theta}_{-j}\hat{\gamma}_j)^2 \quad (9)$$

$$RSS_{exp} = \sum_j (\hat{\Gamma}_j^{random} - \hat{\theta}_{-j}\hat{\gamma}_j^{random})^2 \quad (10)$$

### 2.4.3 垂直方向评估

由于可能的测量误差及未测量的混杂误差会影响因果效应的方向<sup>[85]</sup>，我们使用 MR-Steiger 测试评估所得因果效应的垂直方向。MR-Steiger 方法考虑 SNP-结局关联由 SNP-暴露关联与因果关联的效应共同决定，因此，SNP-暴露的关联性应大于 SNP-结局。同时，Steiger 方法假设 SNP-暴露关联与 SNP-结局关联不存在差异，进行统计学检验。MR-Steiger 方法通过 SNP-暴露关联与 SNP-结局关联的大小以及差异显著性检验评估 MR 因果估计的垂直方向<sup>[79,86]</sup>。

### 2.4.4 效能分析

我们通过估算基于遗传工具的遗传力并计算因果估计的统计效能。遗传力，即遗传变异可解释暴露因素（25OHD 或者 FGF23）水平的方差比例。我们估计遗传力  $h_{SNP}^2$ （公式 11），其中 EAF 为等位基因 A1 的频率， $\hat{\gamma}_j$  为标准化后的 SNP-暴露因素的关联系数<sup>[87]</sup>。我们 mRnd 平台上二值结局的方法基于估算 1 型错误率为 5% 的情况下，工具变量具有足够（>80%）统计效能拒绝零假设时，暴露与结局的因果估计的期望值，并利用该平台计算该研究的统计效能<sup>[88]</sup>。

$$h_{SNP}^2 = 2 \times EAF \times (1 - EAF) \times \hat{\gamma}_j^2 \quad (11)$$

## 研究结果

### 1. 循环 25OHD 与缺血性卒中的因果关联

#### 1.1 循环 25OHD 相关工具变量

本研究一共获取 107 个相互独立且与 25OHD 水平显著相关（ $P < 5.00E-08$ ）的遗传工具（见表 4）。MEGASTROKE 与 UK BioBank 研究来源人群所获得的数据一共匹配得 109 个 SNP 位点，其中 1 个替代工具 rs9616994（原 SNP 为 rs115621755）。替代遗传工具与原始 SNP 的 LD  $r^2$  为 0.996，与原始 SNP 间距为 0.77 kb。本研究通过欧洲裔参考基因组筛选除去 2 个连锁不平衡（LD  $r^2 > 0.1$ ）的遗传变异（rs61816761，rs613808）。至此，本研究一共获得 107 个独立并且相关的遗传工具。



通过 PhenoScanV2 数据库相关表型检测，本研究发现 1 个工具变量（rs964184）与卒中存在直接关联（“Coronary artery disease or large artery stroke”，“Coronary artery disease or ischemic stroke”， $P < 4.67E-04$ ； $m = 107$ ）<sup>[89]</sup>。同时 GLGC 中还发现该遗传变异与高密度脂蛋白、低密度脂蛋白、高密度脂蛋白胆固醇、低密度脂蛋白胆固醇、总胆固醇、三酰甘油显著相关<sup>[66]</sup>。

大型公开 GWAS 数据中一共发现 47 个（44%）与潜在混杂因素存在多效性关联（ $P < 4.67E-4$ ）的遗传工具（见图 3）。一共有 9 个遗传工具与超过 3 个的筛选的表型相关，其中遗传变异 rs1260326 同时与空腹血糖、空腹胰岛素，2 型糖尿病，饮酒频率，高血压，体质指数，三酰甘油，总胆固醇，低密度脂蛋白存在显著性关联；遗传变异 rs590215 与 2 型糖尿病，饮酒频率，体质指数，高血压，高密度脂蛋白同时相关。最后共得到 60 个未发现与所检查的潜在的混杂因素存在多效性关联的遗传工具（表 4）。所有工具变量的遗传强度  $F$  均大于 30。

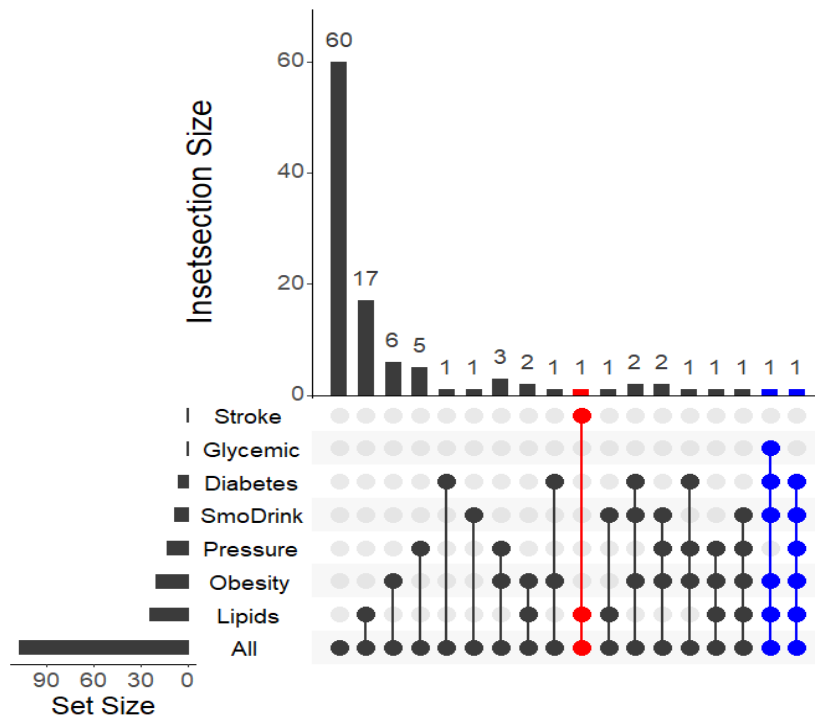


图 3 循环 25OHD 相关工具变量表型关联矩阵分布。

图注：矩阵中的黑圆表示交集的集合。由于遗传工具与其他表型相关（ $P < 4.63E-04$ ），因此柱形图中的与其它变量无关的“**All**”列遗传工具数目低于总的遗传工具数目。红色的集合为卒中直接相关的遗传变异所在的集合（rs964184），蓝色表示与 5 个表型均相关的遗传变异。Obesity 筛选自体质指数和腰臀比。“SmoDrink”表示吸烟、饮酒

表 4 血清 25OHD 显著相关遗传工具

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta	SE	P
rs10908419	ADAR	1: 154567699	G/A	0.51	0.012	0.002	5.81E-10
rs10908465	ASH1L	1: 155389688	C/T	0.73	-0.017	0.002	6.12E-14
rs12123821	FLG-AS1, RP11-107M16.2	1: 152179152	C/T	0.95	-0.079	0.005	2.55E-63
rs140371183	HDHD1P2	1: 152098428	A/G	0.97	-0.087	0.006	1.86E-53
rs1933064	FLG-AS1	1: 152301576	G/A	0.53	-0.016	0.002	9.80E-15
rs35408430	PADI1	1: 17560195	C/T	0.66	0.021	0.002	1.36E-24
rs6671730	PEX10	1: 2339139	G/A	0.57	0.015	0.002	1.92E-13
rs7522116	FOXO6	1: 41835685	C/T	0.43	0.013	0.002	2.97E-11
rs2012736	UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, UGT1A10	2: 234622379	C/A	0.92	0.048	0.004	1.16E-39
rs2710651	EHBP1	2: 63166379	G/A	0.47	0.011	0.002	9.51E-09
rs3849374	NPAS2, AC092168.2	2: 101443397	G/C	0.82	0.016	0.003	7.46E-10
rs7569755	HTR5BP	2: 118648261	G/A	0.71	-0.014	0.002	1.18E-10
rs9861009	RP11-271K21.11	3: 141654685	T/C	0.27	-0.014	0.002	4.86E-10
rs1352846	GC	4: 72617775	A/G	0.71	0.193	0.002	0.00E+00
rs28364331	RP11-696N14.1, ADH1A	4: 100201295	A/G	0.98	-0.069	0.007	4.18E-20
rs4364259	RP11-442P12.1	4: 15892159	G/A	0.80	-0.016	0.003	2.16E-10
rs4616820	REST	4: 57745481	C/T	0.54	0.012	0.002	1.13E-09
rs705117	GC	4: 72608115	C/T	0.15	0.033	0.003	1.06E-32
rs31612	AC012603.1	5: 108996643	T/C	0.83	0.015	0.003	4.15E-08
rs28374650	HLA-DQB1	6: 32623367	C/T	0.76	0.014	0.002	5.44E-09

表 4 血清 25OHD 显著相关遗传工具

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta	SE	P
rs72834856	RP1-209A6.1	6: 22801858	T/G	0.93	0.025	0.004	8.67E-11
rs9490317	RNU4-76P	6: 121859499	T/C	0.55	-0.011	0.002	3.95E-08
rs2346264	EXOC4	7: 133536351	A/C	0.22	0.014	0.002	1.20E-08
rs6966728	LINC01004	7: 104618318	C/T	0.54	0.012	0.002	8.01E-09
rs7784802	ZNF680	7: 64015379	A/T	0.64	-0.014	0.002	2.62E-11
rs12056768	LINC00536	8: 116988527	T/G	0.42	0.023	0.002	6.44E-31
rs10887718	MAT1A	10: 82042624	C/T	0.47	0.011	0.002	2.61E-08
rs3925446	KIF20B	10: 91495322	G/A	0.80	-0.015	0.002	1.09E-09
rs1149605	RP11-21L23.4	11: 76485216	T/C	0.83	-0.022	0.003	1.31E-16
rs116970203	PDE3B	11: 14876718	G/A	0.97	0.377	0.006	0.00E+00
rs117576073 <sup>b</sup>	CYP2R1	11: 14912573	G/T	0.99	0.147	0.009	5.91E-62
rs12794714	CYP2R1	11: 14913575	G/A	0.58	0.088	0.002	0.00E+00
rs12803256	AP002387.1	11: 71132868	A/G	0.22	-0.104	0.002	0.00E+00
rs1660839	AP002387.1	11: 71094232	G/A	0.75	-0.029	0.002	6.40E-37
rs2847500	POU2F3	11: 120114421	G/A	0.88	0.022	0.003	4.42E-13
rs72997623	DGAT2	11: 75488054	C/A	0.92	-0.028	0.004	1.25E-14
rs1038165	MDM1	12: 68665940	C/T	0.42	-0.012	0.002	2.31E-09
rs11182428	RNA5SP358	12: 38526387	T/C	0.48	0.013	0.002	3.23E-10
rs12317268	SLCO1B1	12: 21352541	A/G	0.85	0.021	0.003	6.19E-14
rs73413596	CUX2	12: 111582630	T/C	0.93	-0.022	0.004	1.41E-08
rs7149014	RP11-562L8.1	14: 29802911	T/C	0.37	0.013	0.002	5.42E-10
rs8018720	SEC23A	14: 39556185	G/C	0.18	0.038	0.003	1.26E-47
rs325384	MEF2A	15: 100229761	C/T	0.72	0.014	0.002	1.66E-10
rs11542462	SDR42E1	16: 82033810	G/A	0.87	0.023	0.003	1.27E-15
rs4327060	RP5-991G20.1	16: 72807438	C/T	0.95	0.024	0.004	2.92E-08
rs2037511	SERPINB11	18: 61366207	G/A	0.83	-0.018	0.003	1.35E-11

表 4 血清 25OHD 显著相关遗传工具

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta	SE	P
rs4575545	RP11-345M22.1, RP11-345M22.2	16: 79755446	G/A	0.70	0.016	0.002	7.47E-13
rs8091117	DSG1	18: 28919794	C/A	0.93	0.026	0.004	5.98E-11
rs10426	KLK10	19: 51517798	G/A	0.79	-0.026	0.002	4.64E-26
rs212100	SULT2A1	19: 48376995	T/C	0.16	0.066	0.003	1.60E-133
rs3814995	NPHS1	19: 36342212	C/T	0.69	0.013	0.002	5.18E-09
rs484195	APOC1	19: 45421877	A/G	0.38	0.016	0.002	1.37E-13
rs8113404	ZNF808, ZNF701	19: 53065579	C/T	0.70	-0.012	0.002	2.07E-08
rs17216707	CYP24A1	20: 52732362	T/C	0.82	0.038	0.003	3.47E-46
rs2585442	CYP24A1	20: 52737123	C/G	0.76	-0.036	0.002	6.70E-51
rs2762943	CYP24A1	20: 52790786	T/G	0.08	-0.046	0.004	2.10E-34
rs6123359	BCAS1	20: 52714706	A/G	0.90	-0.034	0.003	6.10E-25
rs115621755 <sup>a</sup>	PPP6R2	22: 50853134	C/T	0.67	0.012	0.002	4.76E-09
rs2074735	PLA2G3	22: 31535872	G/C	0.94	-0.028	0.004	8.23E-12
rs6003456	AP000362.1	22: 23356100	T/A	0.77	0.013	0.002	1.94E-08

注：该数据集来自 Joana A Revez 等人的 25OHD (25-羟基维生素 D) 大型 GWAS (Genome-wide association study) 统计摘要在表型筛选后保留的遗传工具。该数据集包含人群为欧洲生物库中的欧洲裔 (N = 417,580)。遗传变异为基于秩的逆正态转换标准化的 25OHD 浓度 (nmol/L)。SNP、Chr: Pos、EAF、SE、P 分别表示：单核苷酸多态性、遗传变异定位、等位基因 A1 频率、标准误、P 值。遗传变异定位基于人类参考基因组 hg19。右上角标 a 表示该 SNP 在 MEGASTROKE 中以 rs9616994 为替代工具匹配；b 表示该 SNP (rs117576073) 在小血管性卒中和心源性脑栓塞性卒中相关 GWAS 数据中不存在。

## 1.2 循环 25OHD 与 AIS 的因果估计

循环 25OHD 水平与急性缺血性卒中风险未显示存在因果关联。对 107 个独立且相关的遗传工具而言，25OHD 与 AIS 的估计风险 (随机效应的 MR-IVW

法) 为 1.00 OR (95% CI 1.56 OR-3.86 OR;  $P = 0.985$ ), 即每增加 1 个 SD 的 25OHD, 缺血性卒中风险无明显变化。稳健性分析显示 Cochran Q 为 172 ( $P = 4.89E-05$ ), 这提示遗传工具存在显著异质性。此时, 针对大多数工具变量估计因果效应的 MR-WME 法估计的 25OHD 与 AIS 的因果关联为 1.02 OR (95% CI 0.92-1.13;  $P = 0.704$ )。

移除与卒中直接相关的遗传工具 (rs964184) 后, 使用随机效应的 MR-IVW 方法显示, 每增加 1 个 SD 的循环 25OHD 水平的 AIS 的风险为 1.00 OR (95% CI 0.91-1.09;  $P = 0.942$ ), 结果存在明显异质性 (Cochran Q 171,  $P = 5.05E-05$ )。MR-WME 方法估计的结果为 1.02 OR (95% CI 0.92-1.13;  $P = 0.72$ )。MR-Egger 法提示无定向多效性 ( $P = 0.297$ )。移除与潜在的卒中相关表型存在多效性关联的变异后, 余下的遗传工具显示 25OHD 水平与 AIS 无因果关联 (表 5)。固定效应的 MR-IVW 因果估计结果为 1.02 OR (95% CI 0.95-1.10;  $P = 0.570$ ), 与随机效应估计结果一致。MR-WME 法的估计效应值为 1.01 OR (95% CI 0.91-1.12;  $P = 0.81$ )。留一法显示每个遗传工具对因果估计的影响为小幅度扰动, 95%CI 在 -0.137% 与 0.126% 之间。MR-Egger 截距无明显稀释, 无定向多效性 ( $P = 0.982$ ), MR-PRESSO 结果提示不会超过 1 个变异存在水平多效性 ( $P = 0.0284$ )。循环 25OHD 与 AIS 的 MR-Egger 法估计值与 MR-WME 结果一致。

### 1.3 循环 25OHD 与卒中亚型的因果估计

循环 25OHD 水平与大动脉粥样硬化性卒中风险未显示存在因果关联。对 107 个独立且相关的 SNPs 而言, 25OHD 与 LAS 的因果估计量 (随机效应的 MR-IVW 法) 为 0.96 OR (95% CI 0.79-1.16;  $P = 0.680$ ) 每增加 1 个 SD 的循环 25OHD, 结果存在异质性 (Cochran Q 143,  $P = 0.009$ )。MR-WME 估计结果为 0.92 OR (95% CI 0.82-1.14;  $P = 0.682$ )。

移除与卒中直接相关的 SNP 后结果与移除前无明显变化, MR-WME 因果估计为 0.92 OR (95% CI 0.71-1.18;  $P = 0.508$ )。通过血脂、血糖等表型筛选后保留的 60 个变异未提示 25OHD 与 LAS 存在因果关联 (表 5)。随机效应的 MR-IVW 的结果 (0.94 OR, 95% CI 0.76-1.16;  $P = 0.571$ ) 有异质性, Cochran Q 为 83 ( $P = 0.021$ ); MR-WME 估计结果为 0.92 OR (95% CI 0.71-1.17;  $P = 0.488$ )。MR-Egger 结果提示截距无稀释 ( $I^2_{GX} = 99.5\%$ ), 变异无定向多效性 ( $P = 0.902$ )。

表 5 循环 25OHD 与缺血性卒中及其亚型之间的因果关联

Trait	Method	m	OR/ Beta	95% lower	95% upper	Causal_P	Hetero- geneity	Hetero_P
AIS	IVW	60	1.02	0.94	1.11	6.32E-01	Q: 83	2.07E-02
	WME	60	1.01	0.91	1.12	8.10E-01	-	-
	Egger	60	1.02	0.92	1.14	6.97E-01	Q: 83	1.68E-02
	Egger Intercept <sup>a</sup>	60	0	-0.005	0.005	9.82E-01	I <sup>2</sup> <sub>Gx</sub> : 99.5%	-
	PRESSO	60	1.03	1.02	1.03	5.35E-01	-	2.84E-02 <sup>b</sup>
LAS	IVW	60	0.94	0.76	1.16	5.71E-01	Q: 83	2.01E-02
	WME	60	0.92	0.71	1.17	4.88E-01	-	-
	Egger	60	0.93	0.71	1.22	6.05E-01	Q: 83	1.63E-02
	Egger Intercept <sup>a</sup>	60	0.001	-0.011	0.012	9.07E-01	I <sup>2</sup> <sub>Gx</sub> : 99.5%	-
	PRESSO	60	0.94	0.93	0.95	5.73E-01	-	3.08E-02 <sup>b</sup>
SVS	IVW	59	0.99	0.83	1.19	9.33E-01	Q: 130	4.79E-02
	WME	59	0.95	0.75	1.19	6.44E-01	-	-
	Egger	59	0.9	0.72	1.13	3.77E-01	Q: 70	1.16E-01
	Egger Intercept <sup>a</sup>	59	0.007	-0.003	0.017	1.74E-01	I <sup>2</sup> <sub>Gx</sub> : 99.5%	-
	PRESSO	59	0.99	0.99	0.99	9.33E-01	-	1.34E-01 <sup>b</sup>
CES	IVW	59	1.03	0.88	1.2	7.38E-01	Q: 71	1.16E-01
	WME	59	1.11	0.91	1.35	2.89E-01	-	-
	Egger	59	1.03	0.88	1.2	7.38E-01	Q: 68	1.49E-01
	Egger Intercept <sup>a</sup>	59	0.121	-0.071	0.313	2.18E-01	I <sup>2</sup> <sub>Gx</sub> : 99.5%	-
	PRESSO	59	1.03	1.02	1.03	7.39E-01	-	1.39E-01 <sup>b</sup>

注： a 表示 Egger 截距及其 95% 置信区间效应单位为 log OR (Odds Ratio)，其余效应单位为每增加 1 个标准差的 25OHD 水平的 OR 值。 Egger 截距以  $I_{Gx}^2$  量化截距的稀释程度，Hetero\_P 值表示异质性/（截距稀释）统计学 P 值。b 表示全局异质性检验的统计学 P 值。 IVW、 WME、 Q、 PRESSO, m、 OR/Beta、 95% lower、 95%upper、 Causal\_P 分别表示随机效应的逆方差加权法、加权的中位数法、Cochran Q 检验、多效性残差和与异常值检测法；遗传工具数目、效应量、95% 置信区间上界、95% 置信区间下界、因果估计的 P 值；AIS、LAS、SVS、CES 分别表示急性缺血性卒中、大动脉粥样硬化性卒中、小血管性卒中以及心源性脑栓塞性卒中。

循环 25OHD 水平与小血管性卒中风险未显示存在因果关联。对 106 个独立且相关的遗传工具而言，25OHD 与 SVS 通过随机效应的 MR-IVW 法的因果

估计量为 0.97 OR (95% CI 0.82-1.14;  $P = 0.683$ ) 每增加 1 个 SD 的循环 25OHD 水平。MR-WME 估计的结果为 0.94 OR (95% CI 0.75-1.19;  $P = 0.625$ )。

移除与卒中相关的遗传变异后, MR-IVW 估计的结果无明显变化(0.97 OR, 95% CI 0.82-1.14;  $P = 0.683$ ), MR-WME 法估计为 OR 0.95 (95% CI 0.75-1.19;  $P = 0.625$ )。MR-Egger 估计的结果为 OR 0.95 (95% CI 0.77-1.19;  $P = 0.671$ )。通过筛选后的 59 个变异的估计结果见表 5, 随机效应的 MR-IVW 估计结果(0.99 OR) 存在异质性。MR-WME 的估计结果为 0.95 OR (95% CI 0.75-1.19)。Egger 截距测试结果提示遗传工具无定向多效性 ( $P = 0.174$ ), PRESSO 的结果提示无水平多效性 (全局异质性检测,  $P = 0.134$ )。

循环 25OHD 水平与 CES 风险未显示存在因果关联。对 106 个独立且相关的遗传工具而言, 25OHD 与 CES 通过随机效应的 MR-IVW 的因果估计效应量为 1.03 OR (95% CI 0.90-1.18;  $P = 0.640$ ), Cochran Q (116,  $P = 0.222$ ) 评估无异质性。MR-WME 估计结果为 1.11 OR (95% CI 0.92-1.35;  $P = 0.286$ )。MR-Egger 估计为 1.12 OR (95% CI 0.94-1.34;  $P = 0.201$ ), 截距 0.996 ( $P = 0.159$ )。

移除与卒中直接相关的遗传变异后, 随机效应的 MR-IVW 结果无明显变化 (1.03OR, 95% CI 0.90-1.18;  $P = 0.666$ ), MR-WME 结果为 1.11 OR (95% CI 0.91-1.35;  $P = 0.293$ )。血脂等表型筛选后保留的 59 个遗传变异的因果估计见表 5。随机效应的 MR-IVW 估计结果为 1.03 OR, 固定效应的结果与之无明显差异。MR-Egger 法与 MR-PRESSO 法的估计结果与 MR-IVW 一致, 且未提示存在定向多效性与水平多效性。MR-WME 结果为 1.11 OR (95% CI 0.91-1.35;  $P = 0.289$ ), 与移除前一致。

## 2. 循环 FGF23 与缺血性卒中的因果关联

### 2.1 循环 FGF23 相关工具变量

本研究共获取 5 个与循环 FGF23 水平显著相关的遗传工具 (见表 6), 卒中与之匹配的效应见附表 2。这 5 个遗传变异分别位于基因 CYP24A1 (rs17216707)、ABO (rs2769071)、RGS14 (rs11741640)、LINC01506 (rs17479566) 以及 LINC01229 (rs9925837) 附近。计算可得 1 个 SD (23 pg/ml) 的 FGF23 水平约为 0.538 log FGF。连锁结果显示处于同一条染色体的 rs2769071 以及 rs17479566 未处于连锁平衡状态 ( $LD r^2 < 0.1$ )。

表 6 血清 FGF23 显著相关遗传工具

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta <sup>c</sup>	SE <sup>c</sup>	P
rs17216707	CYP24A1	20: 52732362	T/C	0.80	0.1004	0.0093	3.00E-24
rs2769071 <sup>a</sup>	ABO	9: 136145974	G/A	0.37	0.0688	0.0093	6.10E-17
rs11741640 <sup>b</sup>	RGS14	5: 176792743	G/A	0.73	0.0725	0.0093	1.60E-16
rs17479566	LINC01506	9: 71198013	T/C	0.22	0.0576	0.0093	2.00E-09
rs9925837	LINC01229	16: 79927303	G/A	0.13	0.0651	0.0112	5.10E-09

注：FGF23 (pg/ml) 相关 GWAS (Genome-wide association study) 摘要数据获取自一项欧洲裔 GWAS 荟萃研究 (N = 16,624) [39]。SNP、Chr: Pos、EAF、SE、P 分别表示：单核苷酸多态性 (标记位点)、遗传变异在染色体的定位、等位基因 A1 频率、标准误以及与 FGF23 水平的关联 P 值。遗传变异定位选择 hg19 人类参考基因组。右上角标 a、b 分别表示与卒中、高血压表型存在的关联；c：遗传变异效应单位标准化为 1 个标准差的自然对数转换后 (0.538 log FGF23) 的值。

PhenoScanV2 数据库对这 5 个遗传变异进行相关表型检索 ( $P < 0.05/5=0.01$ )，结果在 UK Biobank 数据集中发现 1 个 SNP (rs2769071) 与卒中存在直接关联，分别为 “Self-reported stroke”， $P = 5.82E-03$  与 “Vascular or heart problems diagnosed by doctor: stroke”， $P = 8.40E-03$  [73]。同时，大型公开的 GWAS 中还发现该遗传变异与高密度脂蛋白、低密度脂蛋白、总胆固醇，以及 2 型糖尿病、高血压、舒张压相关。此外，在 UK Biobank 数据集中又发现遗传变异 rs11741640 与诊断的高血压 ( $P = 5.67E-04$ ) 和饮酒相关 ( $P = 3.74E-03$ ) 相关 [73]。表型结果见附表 3。所有变异 ( $m=5$ ) 的遗传强度均大于 30。

## 2.2 循环 FGF23 与 AIS 的因果估计

循环 FGF23 水平与急性缺血性卒中风险无因果关联。对 5 个独立相关的 SNPs 而言，每增加 1 个 SD FGF23 的 AIS 风险 (随机效应的 MR-IVW) 为 1.06 OR (95% CI 0.81-1.40;  $P = 0.985$ )，结果存在异质性 (Cochran Q 14.34;  $P = 6.30E-03$ )。MR-WME 估计为 1.02 OR (95% CI 0.83-1.26;  $P = 0.839$ )。

移除与卒中等表型相关的变异 rs2769071 之后，FGF23 与 AIS 风险的因果估计结果为 0.93 OR (95% CI 0.79-1.10;  $P = 0.400$ )，Cochran Q 无异质性 (Q 1.69,  $P = 0.640$ )。MR-WME 结果为 1.02 OR (95% CI 0.83-1.25;  $P = 0.840$ )，与移除



前一致。MR-PRESSO 未提示遗传工具存在水平多效性 ( $P = 0.525$ )，估计结果为 0.93 OR (95% CI 0.90-0.97;  $P = 0.343$ )。移除与高血压、饮酒相关的变异后结果显示 FGF23 与 AIS 风险无因果关联。MR-IVW 估计结果为 0.90 OR，无异质性 (Cochran Q 1.16)。MR-WME 结果 (0.92 OR) 与 MR-IVW 相似 (表 7)。

表 7 循环 FGF23 与缺血性卒中及其亚型之间的因果关联

Methods		m	OR/ Beta	95% lower	95% upper	Causal_P	Heterogeneity	Hetero_P
AIS	IVW	3	0.90	0.74	1.09	2.71E-01	Q: 1.16	5.60E-01
	WME	3	0.92	0.74	1.14	4.31E-01	-	-
	Egger	3	0.59	0.27	1.31	1.96E-01	Q: 0.05	8.29E-01
	Egger Intercept <sup>a</sup>	3	0.033	-0.028	0.095	2.91E-01	I <sup>2</sup> <sub>GX</sub> : 85.5%	-
LAS	IVW	3	1.74	1.08	2.81	2.30E-02	Q: 0.02	9.93E-01
	WME	3	1.76	1.04	2.99	3.56E-02	-	-
	Egger	3	1.80	0.26	12.46	5.49E-01	Q: 0.01	9.07E-01
	Egger Intercept <sup>a</sup>	3	-0.003	-0.150	0.145	9.70E-01	I <sup>2</sup> <sub>GX</sub> : 82.1%	-
SVS	IVW	3	0.91	0.58	1.42	6.72E-01	Q: 0.70	7.04E-01
	WME	3	0.94	0.57	1.55	8.21E-01	-	-
	Egger	3	0.92	0.15	5.52	9.29E-01	Q: 0.70	4.02E-01
	Egger Intercept <sup>a</sup>	3	-0.001	-0.138	0.136	9.87E-01	I <sup>2</sup> <sub>GX</sub> : 83.0%	-
CES	IVW	3	0.76	0.52	1.10	1.44E-01	Q: 0.48	7.85E-01
	WME	3	0.74	0.49	1.12	1.59E-01	-	-
	Egger	3	0.61	0.14	2.72	5.19E-01	Q: 0.40	5.28E-01
	Egger Intercept <sup>a</sup>	3	0.017	-0.097	0.131	7.70E-01	I <sup>2</sup> <sub>GX</sub> : 83.0%	-

注：a 表示 Egger 截距及其 95% 置信区间效应单位为 log OR (Odds Ratio)，其余效应单位为每增加 1 个标准差的 FGF23 水平的 OR 值。I<sup>2</sup><sub>GX</sub> 量化 Egger 截距的稀释程度，Hetero\_P 值为异质性或截距稀释的 P 值。IVW、WME、Q、PRESSO、m、OR/Beta、95% lower、95% upper、Causal\_P 分别表示随机效应的逆方差加权法、加权的中位数法、Cochran Q、多效性残差和与异常值检测法；遗传工具数目、效应量、95% 置信区间上界、95% 置信区间下界、因果估计的 P 值；AIS、LAS、CES、SVS 分别表示急性缺血性卒中、大动脉粥样硬化性卒中、心源性脑栓塞性卒中以及小血管性卒中。

### 2.3 循环 FGF23 与卒中亚型的因果估计

MR 结果提示循环 FGF23 水平与 LAS 风险存在潜在的遗传因果关联。所有独立且相关的遗传工具的 MR 因果关联结果（随机效应的 MR-IVW 法）显示，每增加 1 个 SD 的 FGF23，LAS 风险为 1.94 OR (95% CI 0.99-3.80;  $P = 0.055$ )；固定效应的评估结果具有统计学显著性 (1.94 OR, 95% CI 1.35-2.77;  $P = 3.04E-04$ )。Cochran Q (14.34,  $P = 6.30E-03$ ) 提示结果存在异质性。MR-WME 提示循环 FGF23 水平与 LAS 风险可能存在因果关联 (1.75 OR, 95% CI 1.06-2.90;  $P = 2.90E-02$ )。留一法提示 ABO 相关变异 rs2769071 驱动了二者间的因果关联，该 SNP 的因果估计为 6.95 OR (95% CI 3.08-15.65;  $P = 2.91E-06$ )。

移除与卒中直接相关的遗传工具 rs2769071 后，MR-IVW 估计结果为 1.42 OR (95% CI 0.95-2.12;  $P = 8.49E-02$ )，Cochran Q 未提示存在异质性 ( $Q = 2.32$ ,  $P = 0.509$ )。MR-WME 估计的结果为 1.70 OR (95% CI 1.03-2.80;  $P = 0.038$ )。留一法提示 rs11741640 对结果存在显著影响，移除该 SNP 可改变 57.67% 的效应，稳健性较差。遗传变异 rs11741640 同时也与饮酒、高血压等表型显著相关。移除该 SNP 后，MR-IVW 估计结果为 1.74 OR (95% CI 1.08-2.81;  $P = 0.027$ )，Cochran Q 提示无异质性 ( $Q = 0.05$ ;  $P = 0.993$ )。单个 SNP 对估计结果的改变在 -0.13% 到 5.8% 之间。MR-WME 估计结果与 MR-IVW 相似 (1.76 OR)。MR-Egger 估计结果为 1.80 OR，未提示存在定向多效性 (表 7)。这 3 个遗传工具的 Wald 比率估计结果可见图 4。

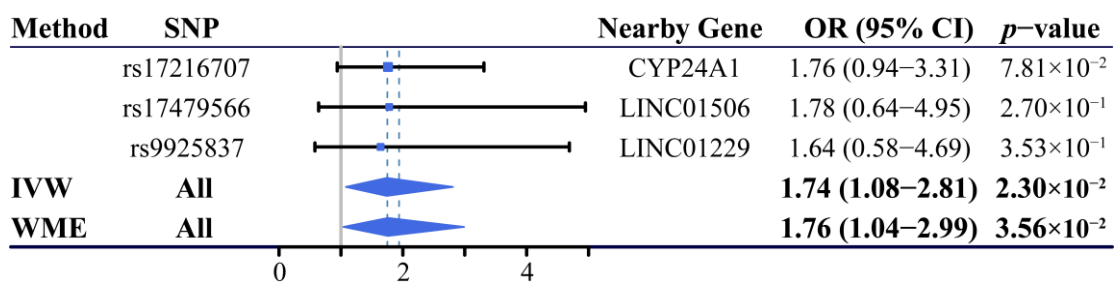


图 4 FGF23 与 LAS 的因果关联森林图

图注：FGF23 与 LAS 分别表示成纤维细胞生长因子 23 和大动脉粥样硬化性卒中，IVW、WME 分别表示固定效应的逆方差加权法与加权的中位数方法；SNP、OR 与 CI 分别表示单核苷酸多态性（位点），比值比，以及置信区间。单位标准化为 1 个标准差的自然对数转换效应值。

MR-Steiger 因果方向评估结果显示, SNP- FGF23 的相关性 ( $r^2 = 1.04E-02$ ) 大于遗传工具与 LAS 风险的相关性 ( $r^2 = 1.26E-05$ ), 二者存在统计学差异 ( $P_{Steiger} < 0.05$ ), 因果方向未出现矛盾性结果。这 3 个遗传工具共解释了 0.53% 的 FGF23 表型方差, 我们有足够的统计效能 ( $>80\%$ ) 检测 1.59 OR/SD log FGF23 的 LAS 风险 (LAS,  $N = 4,373$ ; 对照  $N = 406,111$ )。本研究中循环 FGF23 与 LAS 因果估计的统计效能为 94%。

循环 FGF23 水平与小血管性卒中无因果关联。所有独立且相关的遗传工具通过随机效应的 MR-IVW 法估计, 每增加 1 个 SD 的 FGF23 的 SVS 风险为 0.89 OR (95% CI 0.64-1.24;  $P = 0.488$ ), Cochran Q 无异质性 ( $0.75, P = 0.861$ )。MR-WME 评估结果为 0.91 OR (95% CI 0.62-1.35;  $P = 0.649$ )。MR-Egger 结果存在一定稀释 ( $I^2_{GX} 67.8\%$ ), 未提示存在定向多效性 ( $P = 0.377$ )。MR-PRESSO 全局检测未提示存在水平多效性 ( $P = 0.958$ ), 其估计结果为 0.89 OR (95% CI 0.84-0.94;  $P = 0.158$ ), 该结果与 MR-IVW、MR-WME 估计的结果相一致。

移除与卒中相关的遗传变异后, 随机效应的 MR-IVW 估计的结果没有明显变化 (OR 0.88, 95% CI 0.61-1.28;  $P = 0.861$ ), MR-WME 估计结果为 0.89 OR (95% CI 0.58-1.38;  $P = 0.613$ )。工具变量存在一定稀释, 未提示工具变量存在定向多效性 ( $P = 0.954$ ); MR-PRESSO 估计结果提示无水平多效性 ( $P = 0.894$ )。移除与高血压、饮酒相关的 SNP 后, MR-IVW 估计结果为 0.91 OR (95% CI 0.58-1.42;  $P = 0.672$ ), 未提示存在因果关联 (见表 7)。留一法显示移除单个 SNP 后结果在 -91.2% 至 66.0% 之间改变。

循环 FGF23 水平与心源性脑栓塞性卒中风险未显示存在因果关联。对 5 个工具变量而言, 随机效应的 MR-IVW 的结果显示每增加 1 个 SD 的循环 FGF23 水平, CES 风险为 1.10 OR (95% CI 0.62-1.95;  $P = 0.640$ ), 结果存在异质性 (Cochran Q 16.79,  $P = 2.1E-03$ )。MR-WME 估计结果为 0.94 OR (95% CI 0.64-1.38;  $P = 0.754$ )。MR-Egger 截距测试提示存在定向多效性 ( $P = 1.90E-03$ ), MR-PRESSO 全局检测提示存在水平多效性 ( $P = 3.40E-03$ )。

移除与卒中相关遗传工具后, FGF23 与 CES 的 (随机效应的 MR-IVW 法) 估计结果为 0.84 OR (95% CI 0.62-1.15;  $P = 0.270$ ), 结果无异质性 (Cochran Q 1.45,  $P = 0.694$ ); MR-WME 结果 (0.85 OR, 95% CI 0.58-1.24;  $P = 0.390$ ) 与之一致。留一法提示移除单个 SNP 因果效应的改变量在 -25.7% 到 57.5% 之间。

MR-Egger 截距检测未提示存在定向多效性 ( $P = 0.541$ )，MR-PRESSO 未提示存在水平多效性 ( $P = 0.678$ )。MR-PRESSO 估计结果 (0.84 OR, 95% CI 0.77-0.92;  $P = 0.678$ ) 与 MR-IVW、MR-WME 相一致。移除高血压相关 SNP 后也未提示 FGF23 水平与 CES 存在因果关联 (表 7)。

## 讨论

因果关联的探索有助于优化缺血性卒中的综合预防。本研究使用 MR 框架探索表型之间的关联性质，未发现遗传倾向于升高的血清 25OHD 水平与降低的急性缺血性卒中或其亚型风险有关。MR 结果提示遗传倾向于增高的循环 FGF23 水平与增加的 LAS 风险具有潜在的因果关联。每增加 23 pg/ml (约 1SD) 的循环 FGF23，大动脉粥样硬化性卒中风险增加 74%。

### 1. 与其他研究比较

#### 1.1 缺血性卒中亚型差异

不同的缺血性卒中亚型存在异质性的病理生理机制以及遗传学效应<sup>[90-91]</sup>。大型卒中相关 GWAS 研究指出遗传变异与卒中关联具有亚型特异性<sup>[41,92]</sup>。Harshfield 等人近来发表的文章也指出凝血级联内源性通路及共有通路过程中的因子与心源性脑栓塞性卒中、大动脉粥样硬化性卒中存在因果关联，对小血管性卒中并无遗传学因果效应<sup>[93]</sup>。此外，有研究针对卒中及其亚型的遗传关联结构系统筛选亚型特异性的药物治疗靶点，发现 SCARA5 靶点对心源性脑栓塞性卒中具有潜在的干预价值而且未发现遗传学可能存在的副作用。但是，遗传因果关联证据并不支持靶向 SCARA5 对大动脉粥样硬化性卒中以及小血管性卒中具有潜在的治疗价值<sup>[52]</sup>。进一步理解卒中亚型的遗传关联结构与病理生理机制有助于优化卒中诊疗方案。

我们的结果也发现表型关联效应在缺血性卒中亚型之间存在差异。我们并未发现循环 25OHD 水平和 FGF23 水平与 AIS 风险存在因果关联，但结果提示遗传学倾向于升高的 FGF23 水平与增加的 LAS 风险存在因果关联。使用亚型特异性的工具变量可提高卒中因果关联预测的成功率。

## 1.2 25OHD 相关比较

本研究所得的 25OHD 相关结果与之前的 MR 研究一致<sup>[74,94-95]</sup>。Larsson 用 SUNLIGHT (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits) 研究中循环 25OHD 水平相关的 6 个 SNPs 构建遗传工具, 以 MEGASTROKE 研究中的数据评估维生素 D 与缺血性卒中及其亚型之间的因果关联; 他们的结果显示遗传预测的 1 个 SD 增高的循环 25OHD 水平与 AIS (1.01 OR, 95% CI 0.94-1.08;  $P = 0.84$ ), LAS (0.94 OR, 95% CI 0.80-1.11;  $P = 0.49$ ), SVS (0.95 OR, 95% CI 0.82-1.11;  $P = 0.55$ ), 以及 CES (1.02 OR, 95% CI 0.90-1.16;  $P = 0.74$ ) 的疾病风险未存在因果关联。本研究获得的结果与 Larsson 的结果一致<sup>[94]</sup>, 这也从侧面证实了本研究所设计方案的有效性, 并为该方案在其它表型与缺血性卒中遗传学关联研究 (如循环 FGF23 水平) 中的应用提供了支持性证据。

本研究相较于 Larsson 的研究所选取的 25OHD 相关遗传工具数量更多 (含 SUNLIGHT 研究中的 6 个 SNPs); 此外, 本研究样本来源背景更为均质, 25OHD 水平通过统一的方法检测, 并通过高质控筛选, 数据经过 RINT 转化更符合正态分布。因此, 本研究比 Larsson 的研究具有更高的可信性。本研究与 Larsson 等人基于总结数据的评估结果均未发现遗传增高的循环 25OHD 水平与缺血性卒中或其亚型风险存在因果关联<sup>[94]</sup>。此外, Leong 及其同事 (1.00 OR/SD) 与 Afzal<sup>[74]</sup> (0.96 OR 每 20nmol/L) 等人基于个体水平数据, 分别使用 1 个 (rs2282679) 工具变量和 4 个工具变量 (DHCR7: rs7944926, rs11234027; CYP2R1: rs10741657, rs12794714) 评估 25OHD 水平与缺血性卒中及其亚型风险的遗传关联性质, 也未发现遗传增加的 25OHD 水平与降低的缺血性风险存在因果关联。

在一般人群中的随机对照研究未提示补充维生素 D 增加心血管疾病风险, 也未发现其对卒中具有保护性作用<sup>[26-27]</sup>。VITAL (VITamin D and OmegA-3 Trial) 试验针对普通人群 ( $N = 25,871$ ) 评估每日补充维生素 D3 (2000 IU) 的效应<sup>[26]</sup>, 结果表明补充维生素 D 未增加心血管疾病 (HR 0.96, 95%CI 0.86-1.08) 等疾病风险, 其中卒中风险 HR 为 0.95 (95%CI 0.76-1.20); 此外 VITAL 研究也未发现维生素 D 补充增加心血管疾病相关死亡率 (HR 1.11, 95%CI 0.88-1.40) 及全因死亡风险 (HR 0.99, 95%CI 0.87-1.12), 未发现补充维生素 D 存在明显

的副作用<sup>[26]</sup>。另一项针对社区人群（N = 5108），持续 2.5-4.2 年的高剂量补充维生素 D3 的试验结果显示卒中风险（HR 0.95）与 VITAL 结果相似<sup>[96]</sup>。因此，本研究既不支持也不拒绝在普通人群中补充维生素 D。

循环维生素 D 水平与卒中风险之间所呈现的倒 J 型或 U 型关联可能有助于解释上述现象<sup>[97-98]</sup>。Leung 发现维生素 D 水平与缺血性卒中风险存在倒 J 型的关联，最理想的维生素 D 水平在 70 nmol/L 与 80 nmol/L 之间。与维生素 D 正常（> 70 nmol/L）的人群相比，循环 25OHD 在 43 nmol/L 与 59 nmol/L 间时，维生素 D 与缺血性卒中风险无统计学相关性<sup>[97]</sup>。在 Scragg 的研究中，维生素 D 水平的基线均值为 66.25 nmol/L，标准差为 22.5 nmol/L；该范围与 Leung 等人研究中的阴性范围部分重合，这可能解释在普通人群中维生素 D 补充难以得出其对卒中风险的阳性效应的现象<sup>[96]</sup>。此外，Scragg 的研究中仅 1270 例（24.9%）参与者的维生素 D 水平处于缺乏（25OHD < 50 nmol/L）状态<sup>[96]</sup>。考虑到普通人群中仅仅一部分维生素 D 缺乏<sup>[18-19]</sup>，在该群体中补充维生素 D 难以得出其与卒中或其亚型可能的保护性效应。

### 1.3 FGF23 相关比较

本研究得到的循环 FGF23 水平相关的 LAS 风险与先前报道中（每 1 个 SD 增量的效应）的腰臀比（OR 1.75，95%CI 1.44-2.13）<sup>[99]</sup>，收缩压（OR 1.56，95%CI 1.37-1.78）<sup>[100]</sup>、颈动脉斑块负荷（OR 1.51，95%CI 1.32-1.72）<sup>[100]</sup>，及空腹血糖（OR 1.42，95%CI 1.08-1.85）<sup>[47]</sup>、低密度脂蛋白胆固醇水平（OR 1.28，95%CI 1.07-1.53）<sup>[101]</sup>相似或更高。因此，有必要对 FGF23 在 LAS 的发病发展中作用进一步进行探索。

该研究结果与部分观察性研究相符合<sup>[102-104]</sup>。一项基于社区的研究发现增加 1 个 SD 的 log FGF23，颈动脉斑块出现的风险增加 57%，斑块面积平均增加 0.37 mm<sup>2</sup>，且 FGF23 与颅内动脉粥样硬化的流行病学关联独立于慢性肾脏病<sup>[31]</sup>。在急性缺血性卒中的研究，也存在类似的关联；每增加 100 pg/ml 的循环 FGF23 水平，颅内动脉斑块存在的风险增加 7%，颅内动脉斑块负荷增加 9%<sup>[102]</sup>。慢性肾脏病患者中可观察到更高的循环 FGF23 水平，FGF23 水平与颈动脉内膜中层厚度、斑块风险（3.74 OR/ SD log FGF23）也存在相似的关联<sup>[105]</sup>。这提示血清 FGF23 水平与大动脉粥样硬化性卒中风险呈现正相关，且二者之间的关联机制可能与动脉粥样硬化相关。

目前尚缺乏针对 FGF23 靶点与 LAS 风险的 RCT 研究。一项二次研究旨在通过现有的 RCT 数据探索靶向 FGF23 对心血管疾病潜在的治疗效应。研究结果发现,拟钙剂 cinacalce 可明显降低 FGF23 水平;虽然该研究尚不足以区分是 FGF23 的直接效应与间接效应<sup>[106]</sup>,其结果发现 cinacalce 降低的 FGF23 水平与降低心血管疾病风险(HR 0.66, 95%CI 0.50-0.87)相关。这提示靶向循环 FGF23 可能对改善 LAS 预后是有益的。

## 2. FGF23 相关机制

### 2.1 内皮损伤与免疫炎症

内皮损伤与免疫炎症可能在 FGF23 与大动脉粥样硬化性卒中的致病机制中发挥重要的作用。本研究中, ABO 相关遗传变异 rs2769071 与大动脉粥样硬化性卒中(2.79E-06)等表型具有较紧密的关联性,不满足工具变量假设;但其单变量所估计的 Wald 比率为 6.95OR (2.91E-06),提示免疫系统在 FGF23 与 LAS 相关病理生理相关机制中可能发挥着重要作用。

动脉粥样硬化的病理生理过程包括内皮损伤、脂质沉积、炎性损伤、斑块破裂、血栓形成等<sup>[107]</sup>。FGF23 可加重内皮细胞氧化应激以损伤内皮,该过程通过  $\alpha$ Klotho 非依赖的途径实现<sup>[108-109]</sup>。蛋白  $\alpha$ Klotho 的缺失不影响活性氧的合成,但减少了超氧化物歧化酶等抗氧化物的合成以及活性氧的降解;此外,  $\alpha$ Klotho 的缺失还减少了内皮一氧化氮合酶的合成,不能缓解 FGF23 诱导的血管收缩,从而减少了内皮依赖的血管舒张<sup>[108-109]</sup>。内皮损伤在动脉粥样硬化性血管损伤的病理过程中发挥着重要作用。在炎症因子的诱导下,  $\alpha$ Klotho 表达量降低,这表明免疫炎症在 FGF23 与 LAS 的相关机制中发挥着重要作用。同时, FGF23 也可促进脂质的氧化、沉积等过程<sup>[110-111]</sup>,加快动脉粥样硬化的发展。这进一步提示 FGF23 在大动脉粥样硬化性卒中的发生发展中可能扮演着重要的角色<sup>[107]</sup>。

Biscetti 发现在 2 型糖尿病颈内动脉狭窄群体中,循环 FGF23 水平与斑块不稳定性独立相关,同时炎症分子(白介素 6 等)也增多<sup>[112]</sup>;此外,随着 FGF23 水平增高,血管壁单核细胞粘附分子的表达也增高<sup>[113-114]</sup>。这提示炎症在 LAS 的发生发展中的作用不可忽视。FGF23 与肝细胞表面 FGFR 受体结合可通过 PLC- $\gamma$  信号传导途径促进炎性分子(如 IL-6)的合成<sup>[115]</sup>。循环升高的炎性分子可诱导骨细胞表达 FGF23,进一步增加循环 FGF23 水平<sup>[116-119]</sup>。该过程可能与

低氧诱导因子 1 $\alpha$  (hypoxia-inducible factor-1 alpha, HIF-1 $\alpha$ ) 相关。在炎症反应中, HIF-1 $\alpha$  可通过 NF- $\kappa$ B 等途径被激活; HIF-1 $\alpha$  可与 FGF23 启动子近端的序列结合, 直接刺激 FGF23 分子的转录表达, 进一步促进循环 FGF23 水平的提高<sup>[120]</sup>。此外, FGF23 也可以通过 NF- $\kappa$ B 和 JAK/STAT1 等通路刺激巨噬细胞异位合成 FGF23, 增加局部 FGF23 的表达<sup>[121]</sup>。循环与局部增高的 FGF23 分子可加重炎症损伤。FGF23 通过刺激肿瘤坏死因子  $\alpha$  (TNF- $\alpha$ , Tumor necrosis factor alpha) 的表达, 可诱导 M1 型巨噬细胞极化<sup>[121]</sup>; 也可通过刺激巨噬细胞分泌细胞因子, 直接参与炎症反应, 加重组织的炎症损伤<sup>[122]</sup>。FGF23 与 LAS 发病发展的作用与免疫炎症密切相关 (图 5)。

炎症过程不仅增加了 FGF23 分子的合成 (iFGF23, intact FGF23), 同时也增加了 FGF23 的裂解 (cFGF23, COOH-terminal FGF23), 使得急性炎症中具有生物活性的 iFGF23 保持稳定。在慢性肾脏病中, FGF23 的合成与裂解脱耦联; 大动脉粥样硬化性疾病患者可长期处于慢性炎症反应状态, FGF23 的合成与裂解是否耦联需要进一步实验探索<sup>[123]</sup>。

## 2.2 血管钙化与血管硬化

血管钙化在 FGF23 与 LAS 的病生理过程中可能起到重要作用。首先, FGF23 分子与血管钙化密切相关<sup>[124-125]</sup>。Voigt 等研究者指出 FGF23 在动脉粥样硬化性斑块表达且其表达量在钙化灶显著增高<sup>[124]</sup>。Donate-Correa 等人在不同的动脉组织中也发现相似的现象, 即 FGF23 分子在钙化血管中表达量显著增高。除外局部含量增高, 动脉粥样硬化和血管钙化的热人群中其循环 FGF23 水平也明显增高<sup>[125]</sup>。在慢性肾脏病 (Chronic kidney disease, CKD) 中, FGF23 水平与血管钙化的关联性更为密切。随着 CKD 的进展, 循环 FGF23 水平进行性上升, 血管的钙化也随之增加<sup>[126]</sup>。

血管钙化与心血管疾病风险及其死亡率相关。斑点状钙化在卒中人群中显著增加, 且其与卒中的发病风险密切相关 (OR 1.98, 95%CI 1.42-2.69)<sup>[127]</sup>; 斑点状钙化使动脉斑块易于破裂, 这也解释了其在卒中发病中的部分作用<sup>[128]</sup>。此外, 血管壁钙化的范围增大可导致血管僵硬和顺应性降低, 这与复杂的生物矿化及成骨转化过程相关, 这也有助于解释血管钙化与心血管疾病死亡率高度相关的现象<sup>[129]</sup>。结合本研究中的 FGF23 与 LAS 风险的遗传因果关系, 有理由考虑血管钙化在 FGF23 相关的 LAS 发生发展中起到重要作用。



FGF23 相关血管硬化在 LAS 中起到重要作用。在循环 FGF23 水平增高的群体中,可观察到他们血管的舒张能力存在明显限制<sup>[108]</sup>。FGF23 对血管钙化以及血管硬化的影响与  $\alpha$ Klotho 非依赖的途径有关。蛋白  $\alpha$ Klotho 缺失,血管平滑肌细胞的成骨转化分子标记物 Runx2 表达量增高,可收缩的血管平滑肌标记物 Myocardin, SRF 等分子表达量降低,从而加重血管钙化<sup>[109]</sup>。FGF23 也可以通过组织缺氧等途径诱导平滑肌细胞骨-成骨转化以促进血管硬化<sup>[130]</sup>。动脉硬化可加重组织缺氧缺血<sup>[107]</sup>,组织缺氧可通过刺激 HIF-1 $\alpha$  分子诱导血管平滑肌细胞成骨转化,转而促进血管硬化(图 5)<sup>[129]</sup>。此外,所产生的 HIF-1 $\alpha$  分子可直接诱导 FGF23 表达<sup>[131]</sup>,后者通过抑制促红细胞生成素合成进一步加重组织缺氧,促进血管硬化<sup>[132]</sup>。FGF23 分子与动脉硬化相互作用所形成的恶性循环,可加重组织缺氧,推进 LAS 疾病进程(图 5)。

此外,FGF23 与动脉硬化二者在颅内、颅外 LAS 中相似的差异性也提示 FGF23 相关动脉硬化可能在 LAS 发病发展中起到重要作用。Kim 等人指出动脉硬化与卒中颅内动脉粥样硬化及斑块负荷相关,但并未发现其与颅外脑动脉粥样硬化相关<sup>[133]</sup>;同时 Chang 等人的研究指出 FGF23 与颅内动脉粥样硬化的关联可能比颅外血管更为密切<sup>[102]</sup>。与颅外脑动脉相比,颅内脑动脉具有较少的弹性纤维以及更高比例的胶原纤维含量,这可能有助于解释 FGF23 和动脉硬化在颅内及颅外血管效应的差异<sup>[134]</sup>。颅内动脉粥样硬化性卒中是 LAS 的主要亚型<sup>[135-136]</sup>,因此有理由考虑 FGF23 通过血管硬化在大动脉粥样硬化性卒中的疾病相关机制中发挥重要作用。

### 2.3 维生素 D 相关

FGF23 分子与活性维生素 D 之间存在紧密的相互作用。FGF23 是 1,25-二羟基维生素 D3 的有效抑制剂,且与维生素 D 共同调节钙磷平衡<sup>[137]</sup>。FGF23 可抑制 CYP27B1 的转录以减少 1 $\alpha$ -羟化酶合成,并刺激 CYP24A1 表达促进 24-羟化酶合成;FGF23 通过减少维生素 D 活化并且增加活性维生素 D 降解,降低循环 1,25-二羟基维生素 D3 的水平<sup>[138]</sup>。本研究中 CYP24A1 旁的变异(rs17216707)提供了比其他遗传工具更大的权重,且该遗传工具对循环 FGF23 水平具有最强的解释能力。因此,该 SNP 相关通路可能在 FGF23 对 LAS 的致病机制中扮演着重要的角色。CYP24A1 基因的表达产物可以催化活性维生素 D 的降解。这提示维生素 D 在 FGF23 与 LAS 相关机制中的作用不可忽视。

FGF23 的部分免疫调节作用通过降低活性维生素 D 水平实现。维生素 D 可维持内皮细胞的稳定性,减少内皮损伤<sup>[139-140]</sup>。补充活性维生素 D 可改善系统性红斑狼疮患者的血管内皮功能<sup>[139]</sup>,加强血管细胞间连接以提高血管稳定性,减少内皮损伤<sup>[140]</sup>。内皮损伤机制与维生素 D 缺乏相关的氧化应激及炎症细胞因子表达相关,也与其免疫调节作用存在关联<sup>[141]</sup>。除外经典肾脏活化途径,25OHD 可通过其他组织的血管壁细胞(内皮细胞、平滑肌细胞)以及单核细胞表达的  $1\alpha$  羟化酶活化<sup>[142]</sup>。生理水平的活性维生素 D 可以诱导巨噬细胞向 M2 型转化<sup>[121]</sup>,发挥抗炎<sup>[143]</sup>等免疫调节作用。

维生素 D 不足也可促进血管钙化与血管硬化<sup>[144]</sup>。活性维生素 D 可以减少刺激钙化反应过程的相关分子表达,比如金属蛋白酶 3、金属蛋白酶 9 及血管内皮生长因子等<sup>[145-146]</sup>;骨化三醇也可下调血管壁平滑肌细胞的成骨转化的抑制性通路,如 RUNX2 通路<sup>[147]</sup>。此外,维生素 D 还可减少肿瘤坏死因子  $\alpha$  等促炎因子的表达,缓解血管壁矿化<sup>[148]</sup>。因此,维生素 D 在 FGF23 与 LAS 的病生理机制中可能扮演着重要角色。

除了对血管的直接作用,活性维生素 D 还可通过恢复在炎症因子等作用下下调的  $\alpha$ Klotho 蛋白水平以减少 FGF23 相关血管损伤<sup>[149]</sup>。蛋白  $\alpha$ Klotho 是一种抗衰老蛋白,对血管具有抗炎、抗氧化等作用<sup>[150-151]</sup>,其缺失与心血管相关疾病表型如动脉粥样硬化、血管钙化、内皮功能失调等相关<sup>[152]</sup>。FGF23 对血管钙化的作用与  $\alpha$ Klotho 蛋白的减少有关<sup>[153]</sup>。慢性肾脏病与慢性系统炎症等病生理过程可减少血管壁  $\alpha$ Klotho 分子的表达,减少了  $\alpha$ Klotho 对血管钙化的保护作用,从而推动了血管钙化、血管损伤的进展<sup>[149]</sup>。因此,补充维生素 D 在理论上有助于减少动脉斑块病灶血管损伤,改善 LAS 预后。

虽然目前的证据提供了 FGF23 分子在大动脉粥样硬化性疾病发病发展中的可能关联机制(图 5),但尚不足以明确 FGF23 分子在 LAS 的致病机制中发挥的确切作用<sup>[154]</sup>。本研究通过遗传物理距离相关基因推测的相关机制也不确切。此外,FGF23 分子对 LAS 的作用也需要与钙、磷等相关因素的效应相鉴别<sup>[155]</sup>。因此,需要进一步实验探索 FGF23 与 LAS 之间的具体机制。

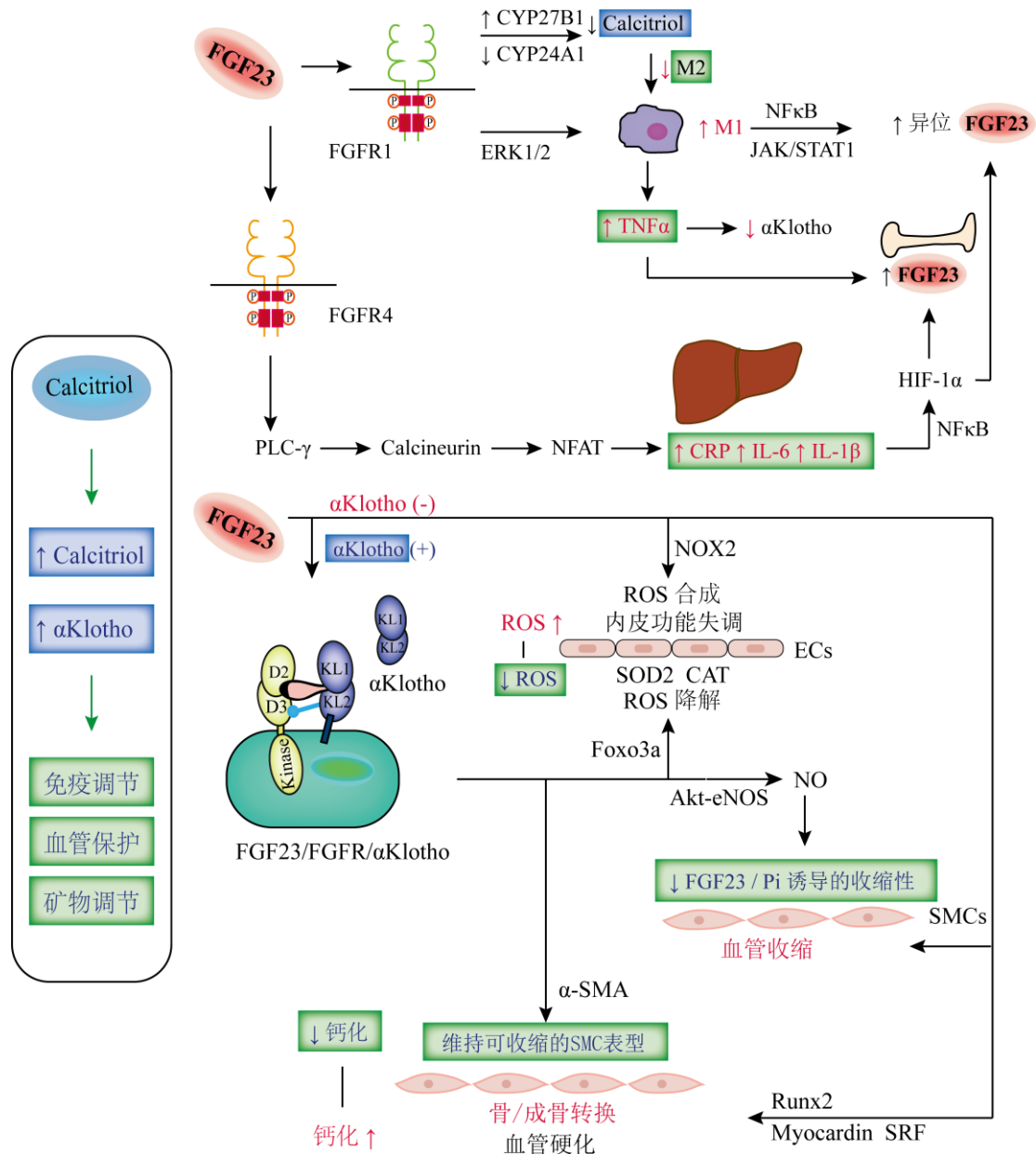


图5 FGF23 与 LAS 的相关机制以及维生素 D 在其中可能的作用

图注：FGF23（Fibroblast growth factor 23，成纤维细胞生长因子）与 LAS（large artery atherosclerotic stroke，大动脉粥样硬化性卒中）相关机制中，免疫炎症及血管钙化可能在其中扮演着重要角色。FGF23 可提高炎症因子的表达量，炎症因子可进一步提高 FGF23 局部和循环的合成量，加重炎症损伤；此外，FGF23 可通过氧化应激损伤血管，减少内皮依赖的血管舒张性，促进血管平滑肌细胞的成骨转换，加重血管钙化损伤。其中的机制可能与 αKlotho 蛋白减少相关。补充维生素 D 可恢复活性维生素 D 与 αKlotho 的生理水平，有助于减少在 LAS 的血管压力下的损伤，提高疾病预后。

### 3. 临床转化

FGF23 与维生素 D 补充。FGF23 对组织的损伤可能部分通过抑制维生素 D 产生<sup>[121]</sup>，且较低的维生素 D 水平与增加的动脉粥样瘤及血管钙化相关<sup>[156]</sup>；此外，血管由于慢性炎症等刺激所下调的，具有抗炎、抗氧化损伤作用的  $\alpha$ Klotho<sup>[109,157]</sup> 可通过与维生素 D 受体激动剂作用而恢复<sup>[149]</sup>。因此，补充维生素 D 理论上有助于改善大动脉粥样硬化性卒中患者预后。

维生素 D 补充相对矛盾的观点在于，上调的活性维生素 D 水平有可能通过与 VDR 结合促进 FGF23 分子的表达<sup>[158]</sup>，转而加重组织损伤。通过文献回顾分析，我们发现对维生素 D 含量不足（20 ng/ml -30 ng/ml）的人群而言，补充维生素 D3 并不明显增高 FGF23 水平<sup>[159-160]</sup>；不仅如此，合理补充维生素 D 有助于恢复  $\alpha$ Klotho 蛋白水平，降低甲状旁腺激素，这对血管具有保护性意义<sup>[161]</sup>。此外，一些干预性研究证据也提示维生素 D 不足群体中补充骨化三醇可改善血管舒张能力，降低内皮粘附因子表达，改善患者预后<sup>[162]</sup>。

对于维生素 D 缺乏（< 20 ng/ml）的人群，FGF23 在维生素 D 补充后的反应可能与 FGF23 水平相关。在肾功能正常人群中，循环 FGF23 水平较低，补充维生素 D3 可以增加 FGF23 的水平<sup>[163-166]</sup>。这可能与强健的 FGF23 负反馈机制相关。如果  $\alpha$ Klotho 水平正常，补充维生素 D 后增加的 FGF23 可能不足以损伤血管<sup>[153]</sup>。对于伴随慢性肾脏病 FGF23 水平增高的人群，短期补充维生素 D 未观察到显著的 FGF23 增高<sup>[165,167]</sup>；但是适当补充维生素 D 可改善高甲状旁腺素血症，并且具有抗炎、调节淋巴细胞功能等作用<sup>[167-168]</sup>。这与 Wolf 对维生素 D 补充对 FGF23 与慢性肾脏病患者预后的影响的理论相符<sup>[169]</sup>。对维生素 D 缺乏且 FGF23 高反应的人群，维生素 D 补充可以结合拟钙剂等使用，以降低 FGF23 的水平，改善预后<sup>[106,170]</sup>。

补充维生素 D 需要依据不同维生素 D 和 FGF23 水平及不同疾病状态。考虑到本研究中所发现的 FGF23 与 LAS 风险可能存在的遗传因果关联，以及维生素 D 在其中可能扮演的角色，在颅内动脉粥样硬化等血管压力下的 LAS 高风险人群中适当补充骨化三醇可能有助于预防卒中事件的发生。一项小样本的非盲法 RCT 研究发现维生素 D 补充可以改善缺血性卒中患者改善三月后的预后<sup>[171]</sup>，这也为在缺血性卒中人群中补充提供了支持性证据。

#### 4. 优势与局限

本研究通过迄今为止最大的 GWAS 数据评估循环 25OHD 和 FGF23 水平与缺血性卒中及其亚型之间的因果关联，研究可信性高。此外，本研究优化 Larsson<sup>[63]</sup>等人的研究方案且结合多种方法评估表型之间的因果关联，得到循环 25OHD 水平与之前的孟德尔随机化的研究结果高度一致<sup>[94]</sup>，这从侧面也证明了该方案的有效性。在此基础上我们得到遗传倾向于增高的循环 FGF23 与增加的大动脉粥样硬化性卒中风险存在可能的因果关联，结果较为可信。

本研究有几处局限。第一，本研究所使用 FGF23 表型数据是通过不同检测方式评估的。在肾脏功能不全患者中，iFGF23 与 cFGF23 不匹配，这可能影响遗传工具对 FGF23 水平评估的准确性。但是，本研究选用的 FGF23 相关 GWAS 数据通过估计的肾小球滤过率筛选以保证纳入人群的肾脏功能相对正常，这尽可能保证了大多数血清样本的 iFGF23/cFGF23 相耦联<sup>[132]</sup>。此外，每个队列内部进行固定效应的 GWAS 分析，log 转换后的 FGF23 水平呈正态分布，在此基础上进行队列间的荟萃分析，可对 FGF23 水平的相对变化进行评估。

第二，可能存在未检出的多效性。本研究中，通过筛选的 3 个遗传变异使得 PRESSO 法无法应用，并且 Egger 截距测试  $I_{GX}^2$  提示截距存在一定稀释，虽然截距量与 0 相当，且遗传工具未提示存在定向多效性，但可能存在未检出的水平多效性。但是，本研究通过三阶段的方法评估并调整了潜在的混杂或多效性影响。对于全部的遗传工具而言，FGF23 与 LAS 风险存在可能的因果关联（MR-WME 法）。在 FGF23 的研究中，筛选出 1 个工具变量与卒中存在关联，1 个变异与缺血性卒中的经典危险因素（高血压）存在多效性关联；在此基础上，MR-Egger 评估无定向多效性，因果方向与 MR-IVW 一致，大小相近。并且，MR-WME 的结果与 MR-IVW 一致。这些证据提示观察到的关联不太可能由潜在的混杂因素解释。

第三，遗传变异 rs17216707 提供了比其他遗传工具更大的权重。通过筛选的 3 个遗传工具中另外 2 个与 FGF23 相关的遗传变异估计的效应量方向一致，每增加约 23 pg/ml 的 FGF23，LAS 风险增加 71%。尽管略去 rs17216707 后精确度下降，但是总体估计大小一致。

第四，孟德尔随机化研究的结果转化局限于特定遗传背景以及因果关联的

阶段。首先, MR 依赖三个关键假设, 这些假设与特定的遗传背景相关。我们使用欧洲裔相关的 GWAS 数据获得遗传倾向于增高的 FGF23 与 LAS 风险增加相关。在不同种族的人群中如印度裔、亚洲裔、非洲裔中, 遗传结构可能存在差异, FGF23 与 LAS 的因果关联也可能不同<sup>[172]</sup>。目前非欧洲裔基因组研究的数据集个体多样并且大型公开 GWAS 数据的缺乏, 导致其代表性不足, GenomAsia 100K 项目可能有效解决这一问题<sup>[173]</sup>。此外, MR 不能区分在特定阶段的关联性质。本研究针对普通人群评估终生暴露的 FGF23 水平与 LAS 的因果关联, 对于 FGF23 与 LAS 在特定疾病状态也需要进一步探讨<sup>[174]</sup>。

第五, FGF23 研究队列可能与 MEGASTROKE 队列人群存在交叉。虽然无明显的证据表明 FGF23 研究队列与 MEGASTROKE 研究队列存在交叉(表 1、表 2), 但不能除外 FGF23 研究内队列如社区动脉粥样硬化风险研究(ARIC, the Atherosclerosis Risk in Communities Study)与心血管健康研究(CHS, the Cardiovascular Health Study)内部分人群同时参与到 MEGASTROKE 卒中研究的队列中。但是, MEGASTROKE 研究样本量远大于 FGF23 研究, 并且可能的交叉人群若是出现在 MEGASTROKE 对照组则不会增加 1 型错误率<sup>[175]</sup>。

第六, 目前尚缺乏与 MEGASTROKE 相互独立的大型 GWAS 数据集作为重复性研究数据, 也缺乏独立的大动脉粥样硬化性卒中 GWAS 研究数据集评估 FGF23 与 LAS 的效应。

## 5. 进一步研究方向

我们的研究提供了遗传增加的 FGF23 水平增加大动脉粥样硬化性卒中风险的初步证据, 并未发现 25OHD 与缺血性卒中或其亚型存在遗传学因果关联。但考虑到可能存在的不足, 还需要更多的研究探索 FGF23 在 LAS 中的作用。

作为一个近年来发现的心血管独立危险因素, FGF23 分子在肾外疾病(如大动脉粥样硬化性卒中)中的作用以及机制还需要进一步探索, 维生素 D 与  $\alpha$ Klotho 蛋白在 FGF23 与 LAS 的相关机制中所扮演的角色也需要进一步的实验评估。

考虑到 FGF23 与维生素 D 之间的密切关联, 以及 FGF23 分子在老龄化等的过程中呈现表达上升的趋势, 在特定人群中适当补充维生素 D 可能是有效的, 需要进一步的研究证据评估其有效性以及安全性。

## 结论

本研究未发现遗传倾向于增高的 25OHD 水平与降低的缺血性卒中或其亚型风险存在关联,但研究结果提示遗传倾向于增高的循环 FGF23 水平与增加的大动脉粥样硬化性卒中风险存在潜在的因果关联。

我们的研究与之前的循环维生素 D 水平相关孟德尔随机化研究以及在普通人群中的维生素 D 补充试验结果一致,未发现循环 25OHD 与缺血性卒中风险存在因果关联。考虑到维生素 D 补充在大型试验中未出现明显的副作用,我们的研究既不支持也不反对在普通人群中补充维生素 D。

FGF23 相关的 LAS 风险与之前研究报道的高血压、血糖等相关指标的 LAS 风险相比更高,这提示 FGF23 对 LAS 发病过程起到重要作用。考虑到 FGF23 与活性维生素 D 水平之间存在的紧密互作,以及 CYP24A1 旁的遗传变异在 FGF23 水平与 LAS 风险关联中的高权重,活性维生素 D 在 FGF23 与 LAS 相关机制中的作用不可忽视。

FGF23 对 LAS 发病的机制可能与内皮损伤和免疫炎症有关,也可能与血管钙化和血管硬化有关。FGF23 对血管的损伤部分与降低的维生素 D 水平相关,也与  $\alpha$ Klotho 蛋白减少相关。考虑到活性维生素 D 可恢复在血管应激状态下所下调的  $\alpha$ Klotho 蛋白表达,结合本研究中所观察到的 FGF23 与 LAS 的遗传因果关联,在大动脉粥样硬化性卒中危险人群(比如颅内大动脉粥样硬化)中适当补充维生素 D 可能有助于减少卒中事件的发生。

维生素 D 是否进行补充不仅需要根据不同疾病状态,也需要考虑维生素 D 水平。在维生素 D 含量不足(20 ng/ml - 30 ng/ml)人群中适当补充维生素 D 可能并不明显提高 FGF23 水平,但可改善血管  $\alpha$ Klotho 表达、血管硬化、内皮炎症等状况。对补充维生素 D 后 FGF23 水平增高的人群,结合拟钙剂使用有望降低 FGF23 水平,进一步改善预后。

本研究为在 LAS 危险人群中补充维生素 D 提供了初步的支持性证据。需要进一步的高质量研究评估补充的有效性。此外,作为一个近年来新发现的心血管独立危险因素,FGF23 在 LAS 发病发展的机制也需要进一步的实验探索,并评估  $\alpha$ Klotho、维生素 D 等分子在 FGF23 与 LAS 相关机制中扮演的角色。

## 参考文献

- [1] Johnson C O, Nguyen M, Roth G A, et al. Global, regional, and national burden of stroke, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016[J]. *Lancet Neurol.* 2019, 18(5): 439-458.
- [2] Zhou M, Wang H, Zeng X, et al. Mortality, morbidity, and risk factors in China and its provinces, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017[J]. *Lancet.* 2019, 394(10204): 1145-1158.
- [3] Mozaffarian D, Benjamin E J, Go A S, et al. Executive summary: Heart disease and stroke statistics--2016 Update: A report from the American Heart Association[J]. *Circulation.* 2016, 133(4): 447-454.
- [4] Network G B o D C. Global burden of disease study 2017 (GBD 2017) results[J]. 2018. <http://ghdx.healthdata.org/gbd-results-tool>.
- [5] Hankey G J. Stroke[J]. *Lancet.* 2017, 389(10069): 641-654.
- [6] Avan A, Digaleh H, Di Napoli M, et al. Socioeconomic status and stroke incidence, prevalence, mortality, and worldwide burden: an ecological analysis from the Global Burden of Disease Study 2017[J]. *BMC Med.* 2019, 17(1): 191.
- [7] Xia X, Yue W, Chao B, et al. Prevalence and risk factors of stroke in the elderly in Northern China: data from the National Stroke Screening Survey[J]. *J Neurol.* 2019, 266(6): 1449-1458.
- [8] Markidan J, Cole J W, Cronin C A, et al. Smoking and risk of ischemic stroke in young men[J]. *Stroke.* 2018, 49(5): 1276-1278.
- [9] Price AJ, Wright FL, Green J, et al. Differences in risk factors for 3 types of stroke: UK prospective study and meta-analyses[J]. *Neurology.* 2018, 90: e298-e306.
- [10] Li Y, Huang Z, Jin C, et al. Longitudinal change of perceived salt intake and stroke risk in a Chinese population[J]. *Stroke.* 2018, 49(6): 1332-1339.
- [11] Kivioja R, Pietila A, Martinez-Majander N, et al. Risk factors for early-onset ischemic stroke: A case-control study[J]. *J Am Heart Assoc.* 2018, 7(21): e009774.
- [12] Wassertheil-Smoller S, Qi Q, Dave T, et al. Polygenic risk for depression increases risk of ischemic stroke: from the stroke genetics network study[J]. *Stroke.* 2018, 49(3): 543-548.



- [13] Wang G L, Zhang R, Zhou Y T, et al. Combined effects of a body shape index and serum C-reactive protein on ischemic stroke incidence among mongolians in China[J]. *Biomed Environ Sci*. 2019, 32(3): 169-176.
- [14] Jenny N S, Callas P W, Judd S E, et al. Inflammatory cytokines and ischemic stroke risk: The REGARDS cohort[J]. *Neurology*. 2019, 92(20): e2375-e2384.
- [15] Cainzos-Achirica M, Enjuanes C, Greenland P, et al. The prognostic value of interleukin 6 in multiple chronic diseases and all-cause death: The Multi-Ethnic Study of Atherosclerosis (MESA)[J]. *Atherosclerosis*. 2018, 278: 217-225.
- [16] Georgakis M K, Gill D, Rannikmae K, et al. Genetically determined levels of circulating cytokines and risk of stroke[J]. *Circulation*. 2019, 139(2): 256-268.
- [17] Fu Y, Liu Q, Anrather J, et al. Immune interventions in stroke[J]. *Nat Rev Neurol*. 2015, 11(9): 524-535.
- [18] Herrick K A, Storandt R J, Afful J, et al. Vitamin D status in the United States, 2011-2014[J]. *Am J Clin Nutr*. 2019, 110(1): 150-157.
- [19] 仓桢, 王宁荐, 李琴, 等. 华东地区自然人群血清维生素 D 水平调查[J]. *中华高血压杂志*. 2018, 26(05): 500.
- [20] Christakos S, Dhawan P, Verstuyf A, et al. Vitamin D: metabolism, molecular mechanism of action, and pleiotropic effects[J]. *Physiol Rev*. 2016, 96(1): 365-408.
- [21] Medrano M, Carrillo-Cruz E, Montero I, et al. Vitamin D: effect on haematopoiesis and immune system and clinical applications[J]. *Int J Mol Sci*. 2018, 19(9): 2663.
- [22] Zhou R, Wang M, Huang H, et al. Lower vitamin D status Is associated with an increased risk of ischemic stroke: A systematic review and meta-analysis[J]. *Nutrients*. 2018, 10(3): 277.
- [23] Qiu H, Wang M, Mi D, et al. Vitamin D status and the risk of recurrent stroke and mortality in ischemic Stroke patients: data from a 24-month follow-up study in China[J]. *J Nutr Health Aging*. 2017, 21(7): 766-771.
- [24] Ji W, Zhou H, Wang S, et al. Low serum levels of 25-hydroxyvitamin D are associated with stroke recurrence and poor functional outcomes in patients with ischemic stroke[J]. *J Nutr Health Aging*. 2017, 21(8): 892-896.
- [25] Wajda J, Swiat M, Owczarek A J, et al. Severity of vitamin D deficiency

- predicts mortality in ischemic stroke patients[J]. *Dis Markers*. 2019, 2019: 3652894.
- [26] Manson J E, Cook N R, Lee I M, et al. Vitamin D supplements and prevention of cancer and cardiovascular disease[J]. *N Engl J Med*. 2019, 380(1): 33-44.
- [27] Barbarawi M, Kheiri B, Zayed Y, et al. Vitamin D supplementation and cardiovascular disease risks in more than 83000 individuals in 21 randomized clinical trials: A meta-analysis[J]. *JAMA Cardiol*. 2019, 4(8): 765-775.
- [28] Lutsey P L, Selvin E, Misialek J R, et al. Abstract P222: Serum fibroblast growth factor-23 and risk of incident stroke: The Atherosclerosis Risk in Communities Study (ARIC)[J]. *Circulation*. 2014, 129(suppl\_1): AP222.
- [29] Panwar B, Judd S E, Wadley V G, et al. Association of fibroblast growth factor 23 With risk of incident coronary heart disease in community-living adults[J]. *JAMA Cardiol*. 2018, 3(4): 318-325.
- [30] Wright C B, Shah N H, Mendez A J, et al. Fibroblast growth factor 23 is associated with subclinical cerebrovascular damage: the Northern Manhattan study[J]. *Stroke*. 2016, 47(4): 923-928.
- [31] Shah N H, Dong C, Elkind M S, et al. Fibroblast growth factor 23 is associated with carotid plaque presence and area: The Northern Manhattan study[J]. *Arterioscler Thromb Vasc Biol*. 2015, 35(9): 2048-2053.
- [32] Yao X Y, Li S, Zhang L G, et al. Higher serum fibroblast growth factor-23 levels and the risk of stroke and its subtypes: evidence from a meta-analysis of prospective studies[J]. *J Stroke Cerebrovasc Dis*. 2018, 27(11): 3076-3083.
- [33] Panwar B, Jenny N S, Howard V J, et al. Fibroblast growth factor 23 and risk of incident stroke in community-living adults[J]. *Stroke*. 2015, 46(2): 322-328.
- [34] Wright C B, Dong C, Stark M, et al. Plasma FGF23 and the risk of stroke: The Northern Manhattan study (NOMAS)[J]. *Neurology*. 2014, 82(19): 1700-1706.
- [35] Seiler S, Reichart B, Roth D, et al. FGF-23 and future cardiovascular events in patients with chronic kidney disease before initiation of dialysis treatment[J]. *Nephrol Dial Transplant*. 2010, 25(12): 3983-3989.
- [36] Vervloet M. Renal and extrarenal effects of fibroblast growth factor 23[J]. *Nat Rev Nephrol*. 2019, 15(2): 109-120.
- [37] Andrukhova O, Slavic S, Odorfer K I, et al. Experimental myocardial infarction

- upregulates circulating fibroblast growth factor-23[J]. *J Bone Miner Res.* 2015, 30(10): 1831-1839.
- [38] Revez J A, Lin T, Qiao Z, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration[J]. *Nat Commun.* 2020, 11: 1647.
- [39] Robinson-Cohen C, Bartz T M, Lai D, et al. Genetic variants associated with circulating fibroblast growth factor 23[J]. *J Am Soc Nephrol.* 2018, 29(10): 2583-2592.
- [40] Quan H, Zhang J, Zhou D, et al. Estimation of standard deviation for a log-transformed variable based on summary statistics in the original scale[J]. *Stat Biopharm Res.* 2018, 10(1): 30-38.
- [41] Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes[J]. *Nat Genet.* 2018, 50(4): 524-537.
- [42] Auton A, Brooks L D, Durbin R M, et al. A global reference for human genetic variation[J]. *Nature.* 2015, 526(7571): 68-74.
- [43] Machiela M J and Chanock S. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants[J]. *Bioinformatics.* 2015, 31(21): 3555-3557.
- [44] Choi K W, Chen C-Y, Stein M B, et al. Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample Mendelian randomization study[J]. *JAMA psychiatry.* 2019, 76(4): 399-408.
- [45] Katan M B. Apolipoprotein E isoforms, serum cholesterol, and cancer[J]. *Lancet.* 1986, 1(8479): 507-508.
- [46] Sun L, Clarke R, Bennett D, et al. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults[J]. *Nat Med.* 2019, 25(4): 569-574.
- [47] Larsson S C, Scott R A, Traylor M, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study[J]. *Neurology.* 2017, 89(5): 454-460.
- [48] Emdin C A, Khera A V and Kathiresan S. Mendelian randomization[J]. *JAMA.*

2017, 318(19): 1925-1926.

[49] van Rheenen W, Peyrot W J, Schork A J, et al. Genetic correlations of polygenic disease traits: from theory to practice[J]. *Nat Rev Genet.* 2019, 20(10): 567-581.

[50] Ference B A, Kastelein J J P, Ginsberg H N, et al. Association of genetic variants related to CETP inhibitors and statins with lipoprotein levels and cardiovascular risk[J]. *JAMA.* 2017, 318(10): 947-956.

[51] Ference B A, Ray K K, Catapano A L, et al. Mendelian randomization study of ACLY and cardiovascular disease[J]. *N Engl J Med.* 2019, 380(11): 1033-1042.

[52] Chong M, Sjaarda J, Pigeyre M, et al. Novel drug targets for ischemic stroke identified through Mendelian randomization analysis of the blood proteome[J]. *Circulation.* 2019, 140(10): 819-830.

[53] Taylor K, Davey Smith G, Relton C L, et al. Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization[J]. *Genome Med.* 2019, 11(1): 6.

[54] Davies N M, Holmes M V and Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians[J]. *BMJ.* 2018, 362: k601.

[55] Bowden J and Holmes M V. Meta-analysis and Mendelian randomization: A review[J]. *Res Synth Methods.* 2019, 10(4): 486-496.

[56] Yuan S, Jiang X, Michaelsson K, et al. Genetic prediction of serum 25-hydroxyvitamin D, calcium, and parathyroid hormone levels in relation to development of type 2 diabetes: A Mendelian randomization study[J]. *Diabetes Care.* 2019, 42(12): 2197-2203.

[57] Hemani G. MRCIEU/TwoSampleMR: WellcomeOpen[J]. 2019. <https://github.com/MRCIEU/TwoSampleMR>.

[58] Yavorska O O and Burgess S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data[J]. *Int J Epidemiol.* 2017, 46(6): 1734-1739.

[59] Verbanck M, Chen C Y, Neale B, et al. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases[J]. *Nat Genet.* 2018, 50(5): 693-698.

- [60] Myers T A, Chanock S J and Machiela M. LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations[J]. *Front Genet.* 2020, 11: 157.
- [61] Wickham H, Francois R, Henry L, et al. dplyr: A grammar of data manipulation. R package version 0.4. 3[J]. *R Found Stat Comput, Vienna.* 2015. <https://CRAN.R-project.org/package=dplyr>.
- [62] Conway J R, Lex A and Gehlenborg N. UpSetR: An R package for the visualization of intersecting sets and their properties[J]. *Bioinformatics.* 2017, 33(18): 2938-2940.
- [63] Larsson S C, Burgess S and Michaelsson K. Association of genetic variants related to serum calcium levels with coronary artery disease and myocardial infarction[J]. *JAMA.* 2017, 318(4): 371-380.
- [64] Kamat M A, Blackshaw J A, Young R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations[J]. *Bioinformatics.* 2019, 35(22): 4851-4853.
- [65] Staley J R, Blackshaw J, Kamat M A, et al. PhenoScanner: a database of human genotype-phenotype associations[J]. *Bioinformatics.* 2016, 32(20): 3207-3209.
- [66] Willer C J, Schmidt E M, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels[J]. *Nat Genet.* 2013, 45(11): 1274-1283.
- [67] Scott R A, Lagou V, Welch R P, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways[J]. *Nat Genet.* 2012, 44(9): 991-1005.
- [68] Mahajan A, Go M J, Zhang W, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility[J]. *Nat Genet.* 2014, 46(3): 234-244.
- [69] Scott R A, Scott L J, Magi R, et al. An expanded genome-wide association study of type 2 diabetes in Europeans[J]. *Diabetes.* 2017, 66(11): 2888-2902.
- [70] Locke A E, Kahali B, Berndt S I, et al. Genetic studies of body mass index yield new insights for obesity biology[J]. *Nature.* 2015, 518(7538): 197-206.
- [71] Turcot V, Lu Y, Highland H M, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in

- obesity[J]. *Nat Genet.* 2018, 50(1): 26-41.
- [72] Ehret GB, Munroe PB, Rice KM, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk[J]. *Nature.* 2011, 478: 103-109.
- [73] Churchhouse C, Neale B M, Abbott L, et al. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK biobank[J]. *Neale Lab.* 2017.
- [74] Afzal S and Nordestgaard B G. Vitamin D, hypertension, and ischemic stroke in 116 655 individuals from the general population: A genetic study[J]. *Hypertension.* 2017, 70(3): 499-507.
- [75] Bowden J, Davey Smith G and Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression[J]. *Int J Epidemiol.* 2015, 44(2): 512-525.
- [76] Bowden J, Del Greco M F, Minelli C, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption[J]. *Int J Epidemiol.* 2019, 48(3): 728-742.
- [77] Burgess S and Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method[J]. *Eur J Epidemiol.* 2017, 32: 377-389.
- [78] Bowden J, Davey Smith G, Haycock P C, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a Weighted Median Estimator[J]. *Genet Epidemiol.* 2016, 40(4): 304-314.
- [79] Hemani G, Tilling K and Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data[J]. *PLoS Genet.* 2017, 13(11): e1007081.
- [80] Brion M J, Shakhbazov K and Visscher P M. Calculating statistical power in Mendelian randomization studies[J]. *Int J Epidemiol.* 2013, 42(5): 1497-1501. <https://shiny.cnsgenomics.com/mRnd/>.
- [81] Bowman K, Jones L, Pilling L C, et al. Vitamin D levels and risk of delirium: A Mendelian randomization study in the UK Biobank[J]. *Neurology.* 2019, 92(12): e1387-e1394.
- [82] Bowden J, Hemani G and Davey Smith G. Invited commentary: Detecting individual and global horizontal pleiotropy in Mendelian randomization-a job for the

- humble heterogeneity statistic?[J]. *Am J Epidemiol*. 2018, 187(12): 2681-2685.
- [83] Thompson W D, Tyrrell J, Borges M C, et al. Association of maternal circulating 25(OH)D and calcium with birth weight: A Mendelian randomisation analysis[J]. *PLoS Med*. 2019, 16(6): e1002828.
- [84] Bowden J, Del Greco M F, Minelli C, et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I<sup>2</sup> statistic[J]. *Int J Epidemiol*. 2016, 45(6): 1961-1974.
- [85] Hernán M A and Cole S R. Invited commentary: Causal diagrams and measurement bias[J]. *Am J Epidemiol*. 2009, 170(8): 959-962; discussion 963-954.
- [86] Richmond R C and Davey Smith G. Commentary: Orienting causal relationships between two phenotypes using bidirectional Mendelian randomization[J]. *Int J Epidemiol*. 2019, 48(3): 907-911.
- [87] Park J-H, Wacholder S, Gail M H, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries[J]. *Nat genet*. 2010, 42(7): 570-575.
- [88] Nowak C and Arnlov J. A Mendelian randomization study of the effects of blood lipids on breast cancer risk[J]. *Nat Commun*. 2018, 9(1): 3957.
- [89] Dichgans M, Malik R, König I R, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants[J]. *Stroke*. 2014, 45(1): 24-36.
- [90] Zerna C, Thomalla G, Campbell B C V, et al. Current practice and future directions in the diagnosis and acute treatment of ischaemic stroke[J]. *Lancet*. 2018, 392(10154): 1247-1256.
- [91] Markus HS and Bevan S. Mechanisms and treatment of ischaemic stroke--insights from genetic associations[J]. *Nat Rev Neurol*. 2014, 10: 723-730.
- [92] NINDS-SiGN and ISGC. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study[J]. *Lancet Neurol*. 2016, 15(2): 174-184.
- [93] Harshfield E L, Sims M C, Traylor M, et al. The role of haematological traits in risk of ischaemic stroke and its subtypes[J]. *Brain*. 2020, 143(1): 210-221.
- [94] Larsson SC, Traylor M, Mishra A, et al. Serum 25-hydroxyvitamin D concentrations and ischemic stroke and its subtypes[J]. *Stroke*. 2018, 49: 2508-2511.

- [95] Leong A, Rehman W, Dastani Z, et al. The causal effect of vitamin D binding protein (DBP) levels on calcemic and cardiometabolic diseases: A Mendelian randomization study[J]. *PLoS Med.* 2014, 11(10): e1001751.
- [96] Scragg R, Stewart A W, Waayer D, et al. Effect of monthly high-dose vitamin D supplementation on cardiovascular disease in the vitamin D assessment study : A randomized clinical trial[J]. *JAMA Cardiol.* 2017, 2(6): 608-616.
- [97] Leung R Y, Han Y, Sing C W, et al. Serum 25-hydroxyvitamin D and the risk of stroke in Hong Kong Chinese[J]. *Thromb Haemost.* 2017, 117(1): 158-163.
- [98] Durup D, Jorgensen H L, Christensen J, et al. A reverse J-shaped association between serum 25-hydroxyvitamin D and cardiovascular disease mortality: The CopD study[J]. *J Clin Endocrinol Metab.* 2015, 100(6): 2339-2346.
- [99] Marini S, Merino J, Montgomery B E, et al. Mendelian randomization study of obesity and cerebrovascular disease[J]. *Ann Neurol.* 2020, 87(4): 516-524.
- [100] Parish S, Arnold M, Clarke R, et al. Assessment of the role of carotid atherosclerosis in the association between major cardiovascular risk factors and ischemic stroke subtypes[J]. *JAMA Netw Open.* 2019, 2(5): e194873-e194873.
- [101] Hindy G, Engstrom G, Larsson S C, et al. Role of blood lipids in the development of ischemic stroke and its subtypes: A Mendelian randomization study[J]. *Stroke.* 2018, 49(4): 820-827.
- [102] Chang Y, Kim J, Woo H G, et al. Plasma fibroblast growth factor 23 concentration is associated with intracranial cerebral atherosclerosis in acute ischemic stroke patients[J]. *J Clin Neurol.* 2020, 16(1): 29-36.
- [103] Hu X, Ma X, Luo Y, et al. Contribution of fibroblast growth factor 23 to Framingham risk score for identifying subclinical atherosclerosis in Chinese men[J]. *Nutr Metab Cardiovasc Dis.* 2017, 27(2): 147-153.
- [104] Rodríguez-Ortiz M E, Alcalá-Díaz J F, Canalejo A, et al. Fibroblast growth factor 23 predicts carotid atherosclerosis in individuals without kidney disease. The CORDIOPREV study[J]. *Eur J Intern Med.* 2019.
- [105] Yan J, Zhang M, Ni Z, et al. Associations of serum fibroblast growth factor 23 with dyslipidemia and carotid atherosclerosis in chronic kidney disease stages 3-5D[J]. *Clin Exp Med.* 2017, 10(9): 13588-13597.



- [106] Moe S M, Chertow G M, Parfrey P S, et al. Cinacalcet, fibroblast growth factor-23, and cardiovascular disease in hemodialysis: the evaluation of cinacalcet HCl therapy to lower cardiovascular events (EVOLVE) trial[J]. *Circulation*. 2015, 132(1): 27-39.
- [107] Libby P, Buring J E, Badimon L, et al. Atherosclerosis[J]. *Nat Rev Dis Primers*. 2019, 5(1): 56.
- [108] Six I, Okazaki H, Gross P, et al. Direct, acute effects of Klotho and FGF23 on vascular smooth muscle and endothelium[J]. *PLoS One*. 2014, 9(4): e93423.
- [109] Richter B, Haller J, Haffner D, et al. Klotho modulates FGF23-mediated NO synthesis and oxidative stress in human coronary artery endothelial cells[J]. *Pflugers Arch*. 2016, 468(9): 1621-1635.
- [110] Pirillo A and Catapano A L. Pitavastatin and HDL: Effects on plasma levels and function(s)[J]. *Atheroscler Suppl*. 2017, 27: e1-e9.
- [111] Mirza M A, Alsio J, Hammarstedt A, et al. Circulating fibroblast growth factor-23 is associated with fat mass and dyslipidemia in two independent cohorts of elderly individuals[J]. *Arterioscler Thromb Vasc Biol*. 2011, 31(1): 219-227.
- [112] Biscetti F, Straface G, Porreca C F, et al. Increased FGF23 serum level is associated with unstable carotid plaque in type 2 diabetic subjects with internal carotid stenosis[J]. *Cardiovasc Diabetol*. 2015, 14: 139-139.
- [113] De Oliveira Neves F M, Araújo C B, de Freitas D F, et al. Fibroblast growth factor 23, endothelium biomarkers and acute kidney injury in critically-ill patients[J]. *J Transl Med*. 2019, 17(1): 121.
- [114] Wallquist C, Mansouri L, Norrbäck M, et al. Associations of fibroblast growth factor 23 with markers of inflammation and leukocyte transmigration in chronic kidney disease[J]. *Nephron*. 2018, 138(4): 287-295.
- [115] Singh S, Grabner A, Yanucil C, et al. Fibroblast growth factor 23 directly targets hepatocytes to promote inflammation in chronic kidney disease[J]. *Kidney Int*. 2016, 90(5): 985-996.
- [116] McKnight Q, Jenkins S, Li X, et al. IL-1 $\beta$  drives production of FGF-23 at the onset of chronic kidney disease in mice[J]. *J Bone Miner Res*. 2020. [Online ahead of print].

- [117] Egli-Spichtig D, Imenez Silva P H, Glaudemans B, et al. Tumor necrosis factor stimulates fibroblast growth factor 23 levels in chronic kidney disease and non-renal inflammation[J]. *Kidney Int.* 2019, 96(4): 890-905.
- [118] Durlacher-Betzer K, Hassan A, Levi R, et al. Interleukin-6 contributes to the increase in fibroblast growth factor 23 expression in acute and chronic kidney disease[J]. *Kidney Int.* 2018, 94(2): 315-325.
- [119] Feger M, Hase P, Zhang B, et al. The production of fibroblast growth factor 23 is controlled by TGF- $\beta$ 2[J]. *Sci Rep.* 2017, 7(1): 4982.
- [120] Zhang Q, Doucet M, Tomlinson R E, et al. The hypoxia-inducible factor-1 $\alpha$  activates ectopic production of fibroblast growth factor 23 in tumor-induced osteomalacia[J]. *Bone Res.* 2016, 4: 16011.
- [121] Han X, Li L, Yang J, et al. Counter-regulatory paracrine actions of FGF-23 and 1,25(OH) $_2$  D in macrophages[J]. *FEBS letters.* 2016, 590(1): 53-67.
- [122] Masuda Y, Ohta H, Morita Y, et al. Expression of FGF23 in activated dendritic cells and macrophages in response to immunological stimuli in mice[J]. *Biol Pharm Bull.* 2015, 38(5): 687-693.
- [123] Czaya B and Faul C. FGF23 and inflammation-a vicious coalition in CKD[J]. *Kidney Int.* 2019, 96(4): 813-815.
- [124] Voigt M, Fischer D C, Rimpau M, et al. Fibroblast growth factor (FGF)-23 and fetuin-A in calcified carotid atheroma[J]. *Histopathology.* 2010, 56(6): 775-788.
- [125] Mirza M A, Larsson A, Lind L, et al. Circulating fibroblast growth factor-23 is associated with vascular dysfunction in the community[J]. *Atherosclerosis.* 2009, 205(2): 385-390.
- [126] Schlieper G, Schurgers L, Brandenburg V, et al. Vascular calcification in chronic kidney disease: an update[J]. *Nephrol Dial Transplant.* 2016, 31(1): 31-39.
- [127] Zhang F, Yang L, Gan L, et al. Spotty calcium on cervicocerebral computed tomography angiography associates with increased risk of ischemic stroke[J]. *Stroke.* 2019, 50(4): 859-866.
- [128] Ruiz J L, Weinbaum S, Aikawa E, et al. Zooming in on the genesis of atherosclerotic plaque microcalcifications[J]. *J Physiol.* 2016, 594(11): 2915-2927.
- [129] Durham A L, Speer M Y, Scatena M, et al. Role of smooth muscle cells in

- vascular calcification: implications in atherosclerosis and arterial stiffness[J]. *Cardiovasc Res*. 2018, 114(4): 590-600.
- [130] Dai B, David V, Martin A, et al. A comparative transcriptome analysis identifying FGF23 regulated genes in the kidney of a mouse CKD model[J]. *PLoS One*. 2012, 7(9): e44161.
- [131] Francis C and David V. Inflammation regulates fibroblast growth factor 23 production[J]. *Curr Opin Nephrol Hypertens*. 2016, 25(4): 325-332.
- [132] Edmonston D and Wolf M. FGF23 at the crossroads of phosphate, iron economy and erythropoiesis[J]. *Nat Rev Nephrol*. 2020, 16(1): 7-19.
- [133] Kim J, Cha M J, Lee D H, et al. The association between cerebral atherosclerosis and arterial stiffness in acute ischemic stroke[J]. *Atherosclerosis*. 2011, 219(2): 887-891.
- [134] Gade PS, Tulamo R, Lee KW, et al. Calcification in human intracranial aneurysms is highly prevalent and displays both atherosclerotic and nonatherosclerotic types[J]. *Arterioscler Thromb Vasc Biol*. 2019, 39: 2157-2167.
- [135] Suemoto C K, Grinberg L T, Leite R E P, et al. Morphometric measurements of extracranial and intracranial atherosclerotic disease: A population-based autopsy study[J]. *Atherosclerosis*. 2018, 270: 218-223.
- [136] Yaghi S, Prabhakaran S, Khatri P, et al. Intracranial atherosclerotic disease[J]. *Stroke*. 2019, 50(5): 1286-1293.
- [137] Quarles L D. Skeletal secretion of FGF-23 regulates phosphate and vitamin D metabolism[J]. *Nat Rev Endocrinol*. 2012, 8(5): 276-286.
- [138] Degirolamo C, Sabba C and Moschetta A. Therapeutic potential of the endocrine fibroblast growth factors FGF19, FGF21 and FGF23[J]. *Nat Rev Drug Discov*. 2016, 15(1): 51-69.
- [139] Reynolds J, Ray D, Alexander M Y, et al. Role of vitamin D in endothelial function and endothelial repair in clinically stable systemic lupus erythematosus[J]. *Lancet*. 2015, 385 Suppl 1: S83.
- [140] Vila Cuenca M, Ferrantelli E, Meinster E, et al. Vitamin D attenuates Endothelial dysfunction in uremic rats and maintains human endothelial stability[J]. *J Am Heart Assoc*. 2018, 7(17): e008776.

- [141] Kim D H, Meza C A, Clarke H, et al. Vitamin D and endothelial Function[J]. *Nutrients*. 2020, 12(2): 575.
- [142] Demer L L, Hsu J J and Tintut Y. Steroid hormone vitamin D: implications for cardiovascular disease[J]. *Circ Res*. 2018, 122(11): 1576-1585.
- [143] David V, Francis C and Babitt J L. Ironing out the cross talk between FGF23 and inflammation[J]. *Am J Physiol Renal Physiol*. 2017, 312(1): f1-f8.
- [144] Wang J, Zhou J J, Robertson G R, et al. Vitamin D in vascular calcification: a double-edged sword?[J]. *Nutrients*. 2018, 10(5): 652.
- [145] Nakagawa K, Sasaki Y, Kato S, et al. 22-Oxa-1alpha,25-dihydroxyvitamin D3 inhibits metastasis and angiogenesis in lung cancer[J]. *Carcinogenesis*. 2005, 26(6): 1044-1054.
- [146] Bao B Y, Yeh S D and Lee Y F. 1alpha,25-dihydroxyvitamin D3 inhibits prostate cancer cell invasion via modulation of selective proteases[J]. *Carcinogenesis*. 2006, 27(1): 32-42.
- [147] Drissi H, Pouliot A, Koolloos C, et al. 1,25-(OH)<sub>2</sub>-vitamin D3 suppresses the bone-related Runx2/Cbfa1 gene promoter[J]. *Exp Cell Res*. 2002, 274(2): 323-333.
- [148] Aoshima Y, Mizobuchi M, Ogata H, et al. Vitamin D receptor activators inhibit vascular smooth muscle cell mineralization induced by phosphate and TNF- $\alpha$ [J]. *Nephrol Dial Transplant*. 2012, 27(5): 1800-1806.
- [149] Lim K, Lu T S, Molostvov G, et al. Vascular Klotho deficiency potentiates the development of human artery calcification and mediates resistance to fibroblast growth factor 23[J]. *Circulation*. 2012, 125(18): 2243-2255.
- [150] Kuro-o M, Matsumura Y, Aizawa H, et al. Mutation of the mouse klotho gene leads to a syndrome resembling ageing[J]. *Nature*. 1997, 390(6655): 45-51.
- [151] Xu Y and Sun Z. Molecular basis of Klotho: from gene to function in aging[J]. *Endocr Rev*. 2015, 36(2): 174-193.
- [152] Navarro-González J F, Donate-Correa J, Muros de Fuentes M, et al. Reduced Klotho is associated with the presence and severity of coronary artery disease[J]. *Heart*. 2014, 100(1): 34-40.
- [153] Komaba H and Fukagawa M. The role of FGF23 in CKD--with or without Klotho[J]. *Nat Rev Nephrol*. 2012, 8(8): 484-490.

- [154] Takashi Y and Fukumoto S. FGF23 beyond phosphotropic hormone[J]. *Trends Endocrinol Metab.* 2018, 29(11): 755-767.
- [155] El-Abbadi M M, Pai A S, Leaf E M, et al. Phosphate feeding induces arterial medial calcification in uremic mice: role of serum phosphorus, fibroblast growth factor-23, and osteopontin[J]. *Kidney Int.* 2009, 75(12): 1297-1307.
- [156] Muscogiuri G, Annweiler C, Duval G, et al. Vitamin D and cardiovascular disease: From atherosclerosis to myocardial infarction and stroke[J]. *Int J Cardiol.* 2017, 230: 577-584.
- [157] Rizzo B, Maltese G, Paraskevi M P, et al. Induction of antioxidant genes by sulforaphane and klotho in human aortic smooth muscle cells[J]. *Free Radic Biol Med.* 2014, 75 Suppl 1: S14-15.
- [158] Charoenngam N, Rujirachun P, Holick M F, et al. Oral vitamin D3 supplementation increases serum fibroblast growth factor 23 concentration in vitamin D-deficient patients: a systematic review and meta-analysis[J]. *Osteoporos Int.* 2019, 30(11): 2183-2193.
- [159] Trummer C, Schwetz V, Pandis M, et al. Effects of vitamin D supplementation on FGF23: a randomized-controlled trial[J]. *Eur J Nutr.* 2019, 58(2): 697-703.
- [160] Alvarez JA, Law J, Coakley KE, et al. High-dose cholecalciferol reduces parathyroid hormone in patients with early chronic kidney disease: a pilot, randomized, double-blind, placebo-controlled trial[J]. *Am J Clin Nutr.* 2012, 96: 672-679.
- [161] Linhartova K, Veselka J, Sterbakova G, et al. Parathyroid hormone and vitamin D levels are independently associated with calcific aortic stenosis[J]. *Circ J.* 2008, 72(2): 245-250.
- [162] Chitalia N, Ismail T, Tooth L, et al. Impact of vitamin D supplementation on arterial vasomotion, stiffness and endothelial biomarkers in chronic kidney disease patients[J]. *PLoS One.* 2014, 9(3): e91363.
- [163] De Niet S, Coffiner M, Da Silva S, et al. A randomized study to compare a monthly to a daily administration of vitamin D(3) supplementation[J]. *Nutrients.* 2018, 10(6): 659.
- [164] Kamelian T, Saki F, Jeddi M, et al. Effect of cholecalciferol therapy on serum

- FGF23 in vitamin D deficient patients: a randomized clinical trial[J]. *J Endocrinol Invest.* 2018, 41(3): 299-306.
- [165] Alshayeb H, Showkat A, Wall B M, et al. Activation of FGF-23 mediated vitamin D degradative pathways by cholecalciferol[J]. *J Clin Endocrinol Metab.* 2014, 99(10): E1830-1837.
- [166] Nygaard B, Frandsen N E, Brandi L, et al. Effects of high doses of cholecalciferol in normal subjects: a randomized double-blinded, placebo-controlled trial[J]. *PLoS One.* 2014, 9(8): e102965.
- [167] Carvalho JTG, Schneider M, Cuppari L, et al. Cholecalciferol decreases inflammation and improves vitamin D regulatory enzymes in lymphocytes in the uremic environment: A randomized controlled pilot trial[J]. *PLoS One.* 2017, 12(6): e0179540.
- [168] Ojeda López R, Esquivias de Motta E, Carmona A, et al. Correction of 25-OH-vitamin D deficiency improves control of secondary hyperparathyroidism and reduces the inflammation in stable haemodialysis patients[J]. *Nefrologia.* 2018, 38: 41-47.
- [169] Wolf M. Update on fibroblast growth factor 23 in chronic kidney disease[J]. *Kidney Int.* 2012, 82(7): 737-747.
- [170] Raggi P, Chertow GM, Torres PU, et al. The ADVANCE study: a randomized study to evaluate the effects of cinacalcet plus low-dose vitamin D on vascular calcification in patients on hemodialysis[J]. *Nephrol Dial Transplant.* 2011, 26(4): 1327-1339.
- [171] Narasimhan S and Balasubramanian P. Role of vitamin D in the outcome of ischemic stroke- s randomized controlled trial[J]. *J Clin Diagn Res.* 2017, 11: cc06-cc10.
- [172] Brown B C, Ye C J, Price A L, et al. Transethnic genetic-correlation estimates from summary statistics[J]. *Am J Hum Genet.* 2016, 99(1): 76-88.
- [173] Consortium. G K. The GenomeAsia 100K Project enables genetic discoveries across Asia[J]. *Nature.* 2019, 576(7785): 106-111.
- [174] Mokry L E, Ross S, Ahmad O S, et al. Vitamin D and risk of multiple sclerosis: A Mendelian randomization study[J]. *PLoS Med.* 2015, 12(8): e1001866.

- [175] Burgess S, Davies N M and Thompson S G. Bias due to participant overlap in two-sample Mendelian randomization[J]. *Genet Epidemiol.* 2016, 40(7): 597-608.

## 附录

附表 1 血清 25OHD 显著相关遗传工具（移除）

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta	SE	P	Chose
rs11591147	PCSK9	1: 55505647	G/T	0.98	-0.045	0.007	1.64E-09	0
rs2131925	DOCK7	1: 63025942	G/T	0.36	0.023	0.002	3.61E-28	0
rs61816761 <sup>d</sup>	FLG-AS1,FLG	1: 152285861	G/A	0.98	-0.123	0.008	7.35E-53	0
rs6672758	GALNT2	1: 230303512	C/T	0.20	-0.018	0.003	2.40E-12	0
rs7528419	CELSR2	1: 109817192	A/G	0.78	-0.020	0.002	1.35E-16	0
rs867772	43891	1: 220972343	A/G	0.32	0.015	0.002	1.14E-11	0
rs1047891	CPS1	2: 211540507	C/A	0.68	0.015	0.002	1.18E-12	0
rs1260326	GCKR	2: 27730940	T/C	0.39	-0.021	0.002	4.42E-24	0
rs541041	APOB	2: 21294975	G/A	0.18	0.015	0.003	2.27E-09	0
rs727857	LINC01122	2: 58981967	G/A	0.39	0.014	0.002	1.05E-11	0
rs7604788	RP11-116D2.1	2: 21190024	C/T	0.97	-0.034	0.006	1.32E-09	0
rs6782190	CADM2	3: 85639672	G/A	0.35	0.017	0.002	1.45E-16	0
rs11732896	HSD17B11	4: 88287993	G/A	0.70	0.016	0.002	1.79E-13	0
rs1229984	ADH1B	4: 100239319	T/C	0.02	0.045	0.006	1.53E-12	0
rs78649910	DOK7	4: 3482213	T/A	0.89	0.021	0.003	7.15E-11	0
rs10070734	LINC00461	5: 87940026	T/C	0.29	-0.013	0.002	1.82E-09	0
rs2248551	MED23	6: 131924689	G/A	0.83	0.023	0.003	3.04E-18	0
rs78151190	LRRC16A	6: 25619007	A/C	0.87	0.017	0.003	1.39E-08	0
rs9476310	RP11-325M4.2	6: 57767576	C/T	0.49	-0.012	0.002	4.21E-09	0
rs10085881	DNAH11	7: 21577960	T/C	0.72	0.015	0.002	7.83E-11	0
rs75741381	VGF	7: 100809458	C/G	0.85	0.017	0.003	4.15E-09	0
rs28692966	EBF2	8: 25892919	G/A	0.75	-0.015	0.002	1.13E-10	0
rs34290760	RP11-115J16.1	8: 9185179	C/G	0.97	0.033	0.006	1.70E-08	0
rs4738684	CYP7A1	8: 59393273	A/G	0.33	-0.012	0.002	4.41E-09	0



附表 1 血清 25OHD 显著相关遗传工具 (移除)

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta	SE	P	Chose
rs804281	GATA4	8: 11611865	A/G	0.42	-0.013	0.002	4.72E-11	0
rs13284054	ABCA1	9: 107669073	T/C	0.88	-0.018	0.003	2.07E-08	0
rs4418728	CYP26A1	10: 94839724	G/T	0.55	-0.011	0.002	4.24E-08	0
rs77532868	GRID1	10: 88081438	C/T	0.95	-0.027	0.004	1.56E-09	0
rs12798050 <sup>c</sup>	S100A11P3	11: 71223256	C/T	0.17	-0.110	0.003	0.00E+00	0
rs613808 <sup>d</sup>	APOA1-AS	11: 116710968	A/G	0.28	-0.026	0.002	3.40E-32	0
rs61891388	RP11-867G23.13	11: 66079818	T/G	0.54	-0.013	0.002	4.06E-10	0
rs78168201 <sup>c</sup>	SHANK2	11: 70971149	C/T	0.99	-0.089	0.009	5.18E-25	0
rs964184	ZNF259	11: 116648917	G/C	0.13	-0.043	0.003	1.09E-48	0
rs10859995	HAL	12: 96375682	T/C	0.42	0.040	0.002	1.06E-88	0
rs12372115	RMST	12: 97982701	G/T	0.93	0.022	0.004	1.93E-08	0
rs12881545	DLK1	14: 101176212	G/C	0.33	-0.012	0.002	2.91E-08	0
rs1800588	ALDH1A2, LIPC	15: 58723675	C/T	0.78	0.033	0.002	4.38E-42	0
rs261291	ALDH1A2	15: 58680178	T/C	0.64	0.027	0.002	2.50E-39	0
rs55829990	USP3	15: 63790642	T/C	0.66	0.019	0.002	9.12E-19	0
rs62007299	PEAK1	15: 77711719	G/A	0.29	0.013	0.002	1.32E-09	0
rs11076175	CETP	16: 57006378	A/G	0.82	-0.023	0.003	9.48E-19	0
rs17231506	CETP	16: 56994528	C/T	0.68	0.018	0.002	5.45E-18	0
rs77924615	PDILT	16: 20392332	G/A	0.81	0.017	0.003	7.11E-11	0
rs10454087	FAM134C	17: 40735641	C/T	0.72	0.014	0.002	8.70E-10	0
rs2952289	RP11-120M18.2	17: 66464414	C/T	0.20	-0.018	0.002	1.18E-12	0
rs4121823	LIPG	18: 47144223	T/A	0.15	0.019	0.003	3.83E-12	0
rs590215	RP11-795H16.2	18: 57904088	C/T	0.73	0.013	0.002	1.04E-08	0

附表 1 血清 25OHD 显著相关遗传工具 (移除)

SNP	Near Gene	Chr: Pos	A1/A2	EAF	Beta	SE	P	Chose
rs11606 <sup>a</sup>	CNOT3	19: 54658102	C/G	0.57	-0.012	0.002	4.41E-09	0
rs142158911	LDLR	19: 11190534	G/A	0.89	-0.026	0.003	4.78E-16	0
rs187429064 <sup>c</sup>	AC138430.4, TM6SF2	19: 19380513	A/G	0.99	-0.065	0.009	7.95E-12	0
rs7412	APOE	19: 45412079	C/T	0.92	-0.030	0.004	1.36E-16	0
rs2207132	MAFB	20: 39142516	G/A	0.97	0.035	0.006	5.56E-10	0
rs2229742	NRIP1	21: 16339172	G/C	0.90	0.025	0.003	1.48E-14	0

注: 该数据集来自 Joana A Revez 等人的 25OHD (25-羟基维生素 D) 大型 GWAS (Genome-wide association study) 统计摘要数据。该数据集包含人群为欧洲生物库中的欧洲裔 (N = 417,580)。Selected 表示表型筛选中移除 (0) 的遗传工具。遗传变异为基于秩的逆正态转换标准化的 25OHD 浓度 (nmol/L)。SNP、Chr: Pos、EAF、P、SE 分别表示: 单核苷酸多态性、遗传变异定位、等位基因 A1 频率、标准误、P 值。遗传变异定位基于人类参考基因组 hg19。右上标 c 表示在 MEGASTROKE 中无法找到匹配或者替代工具匹配; d 表示连锁不平衡 ( $r^2 > 0.1$ ) 移除的遗传变异; e 表示该 SNP (rs117576073) 在小血管性卒中和心源性脑栓塞性卒中相关 GWAS 数据中不存在。

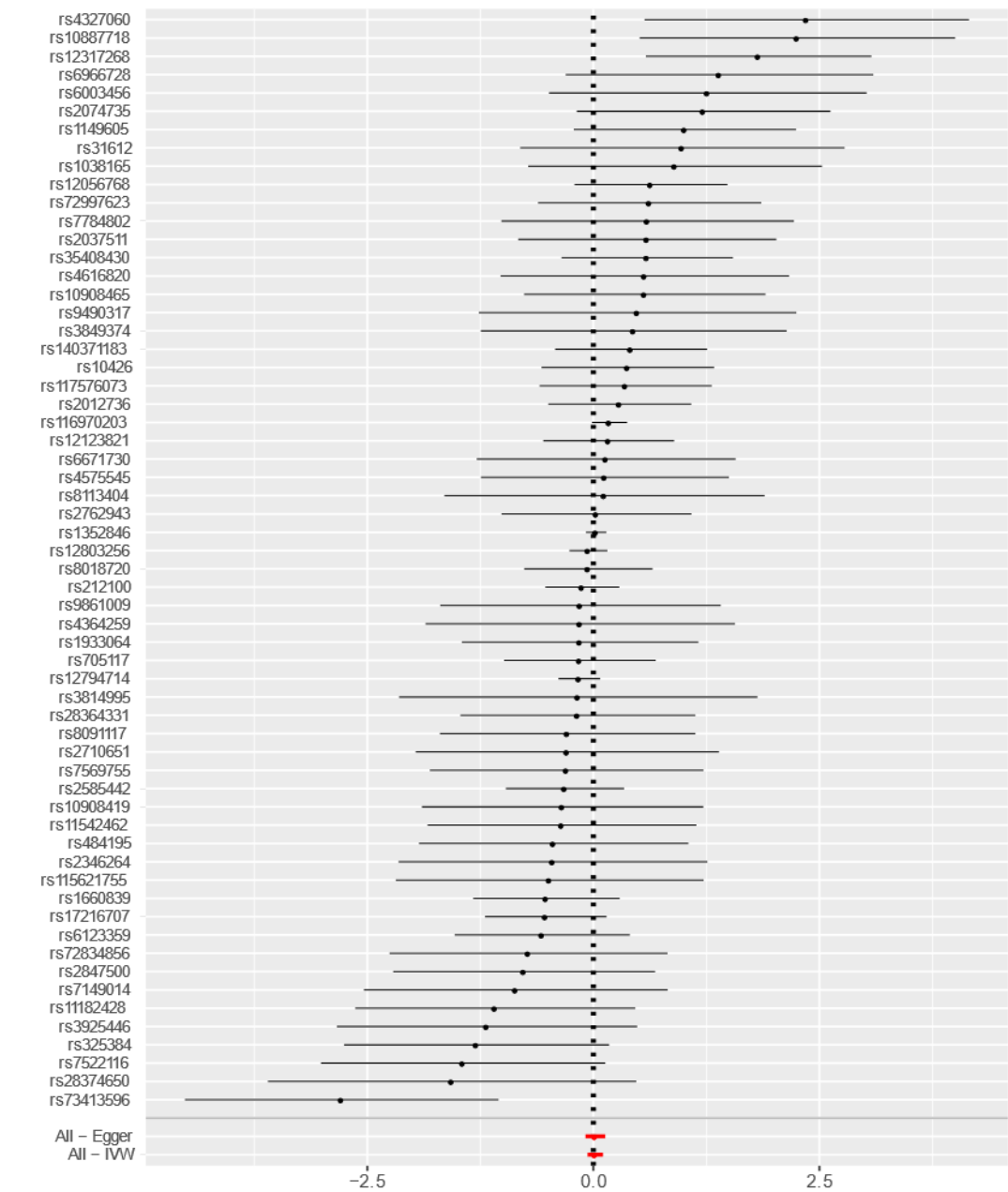
附表 2 MEGASTROKE 缺血性卒中及其亚型匹配的 FGF23 关联效应

Trait	SNP	Near Gene	Chr: Pos	A1/A2	Beta	SE	P
AIS	rs17216707	CYP24A1	20: 52732362	T/C	-0.0198	0.0129	1.23E-01
	rs2769071	ABO	9: 136145974	G/A	0.0412	0.0116	3.84E-04
	rs11741640	RGS14	5: 176792743	G/A	0.0016	0.0110	8.83E-01
	rs17479566	LINC01506	9: 71198013	T/C	0.0012	0.0133	9.30E-01
	rs9925837	LINC01229	16: 79927303	G/A	0.0014	0.0139	9.20E-01
LAS	rs17216707	CYP24A1	20: 52732362	T/C	0.0569	0.0323	7.80E-02
	rs2769071	ABO	9: 136145974	G/A	0.1333	0.0285	2.79E-06
	rs11741640	RGS14	5: 176792743	G/A	-0.0088	0.0270	7.45E-01
	rs17479566	LINC01506	9: 71198013	T/C	0.0332	0.0301	2.70E-01
	rs9925837	LINC01229	16: 79927303	G/A	0.0323	0.0348	3.53E-01
SVS	rs17216707	CYP24A1	20: 52732362	T/C	-0.0061	0.0298	8.38E-01
	rs2769071	ABO	9: 136145974	G/A	-0.0068	0.0267	8.00E-01
	rs11741640	RGS14	5: 176792743	G/A	-0.0137	0.0254	5.89E-01
	rs17479566	LINC01506	9: 71198013	T/C	0.0076	0.0280	7.85E-01
	rs9925837	LINC01229	16: 79927303	G/A	-0.0286	0.0327	3.82E-01
CES	rs17216707	CYP24A1	20: 52732362	T/C	-0.0303	0.0249	2.24E-01
	rs2769071	ABO	9: 136145974	G/A	0.0866	0.0227	1.34E-04
	rs11741640	RGS14	5: 176792743	G/A	0.0048	0.0212	8.20E-01
	rs17479566	LINC01506	9: 71198013	T/C	-0.0028	0.0234	9.07E-01
	rs9925837	LINC01229	16: 79927303	G/A	-0.0288	0.0272	2.90E-01
CAD	rs17216707	CYP24A1	20: 52732362	T/C	-0.0179	0.0109	1.00E-01
	rs2769071	ABO	9: 136145974	G/A	0.0441	0.0095	3.28E-06
	rs11741640	RGS14	5: 176792743	G/A	-0.0217	0.0105	3.87E-02
	rs17479566	LINC01506	9: 71198013	T/C	-0.0179	0.0109	1.00E-01
	rs9925837	LINC01229	16: 79927303	G/A	0.0021	0.0128	8.70E-01

附表 3 FGF23 相关遗传变异可能的水平多效性

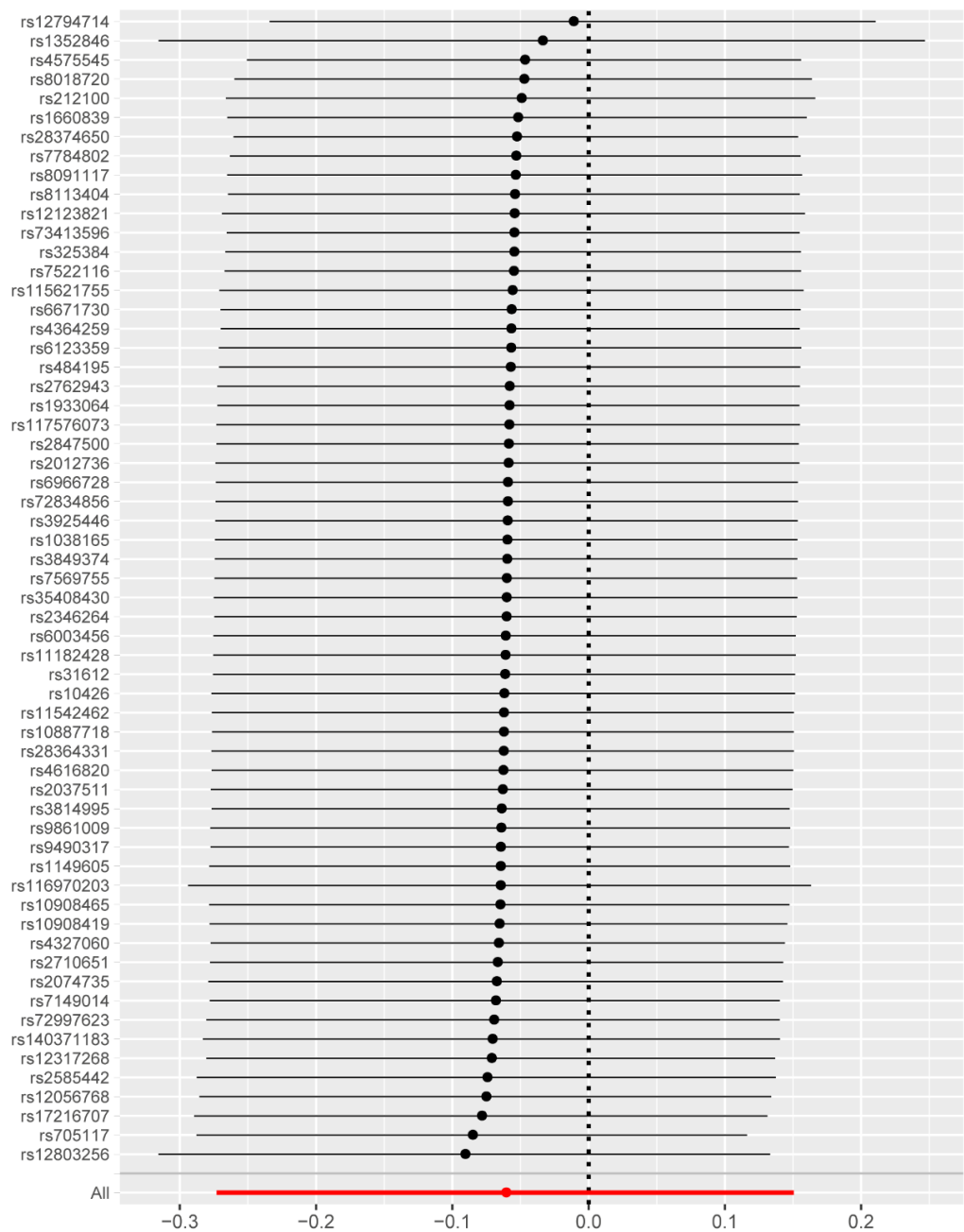
SNP	Trait	Beta	SE	<i>p</i>	Study	PMID	year	N
<b>rs2769071</b> A/G <sup>†</sup> (ABO) <sup>‡</sup>	High density lipoprotein	-0.0199	0.0050	$2.61 \times 10^{-04}$	GLGC	24097068	2013	187,167
	Low density lipoprotein	-0.0353	0.0054	$3.06 \times 10^{-10}$	GLGC	24097068	2013	173,082
	Total cholesterol	-0.0383	0.0051	$7.48 \times 10^{-13}$	GLGC	24097068	2013	187,365
	Type 2 diabetes adjusted for BMI	-0.0500	0.0150	$1.20 \times 10^{-03}$	DIAGRAM	28566273	2017	159,208
	Type 2 diabetes	-0.0540	0.0130	$2.30 \times 10^{-05}$	DIAGRAM	28566273	2017	159,208
	Self-reported high cholesterol	-0.0053	0.0009	$4.26 \times 10^{-10}$	Neale B	UKBB	2017	337,159
	Diabetes diagnosed by doctor	-0.0020	0.0006	$4.54 \times 10^{-04}$	Neale B	UKBB	2017	336,473
	Self-reported diabetes	-0.0017	0.0005	$8.69 \times 10^{-04}$	Neale B	UKBB	2017	337,159
	Diastolic blood pressure	0.0167	0.0026	$2.80 \times 10^{-10}$	Neale B	UKBB	2017	317,756
	Self-reported hypertension	0.0059	0.0011	$2.85 \times 10^{-07}$	Neale B	UKBB	2017	337,159
	Vascular or heart problems diagnosed by doctor: high blood pressure	0.0065	0.0012	$2.48 \times 10^{-08}$	Neale B	UKBB	2017	336,683
<b>rs11741640</b> A/G <sup>†</sup> (RGS14) <sup>‡</sup>	Self-reported hypertension	0.0045	0.0012	$2.22 \times 10^{-04}$	Neale B	UKBB	2017	337,159
	Vascular or heart problems diagnosed by doctor: high blood pressure	0.0042	0.0012	$5.66 \times 10^{-04}$	Neale B	UKBB	2017	336,683
	Alcohol intake frequency	0.0117	0.0040	$3.74 \times 10^{-03}$	Neale B	UKBB	2017	336,965

所有多效性关联均为欧洲裔数据。缩写 FGF23、SE、SNPs、PMID、GLGC、DIAGRAM、Neale B 分别表示成纤维细胞生长因子、标准误、单核苷酸多态性位点、Pubmed 检索号、全球脂质遗传协会、糖尿病遗传重复和荟萃分析研究，以及 Neale 生物库数据；标志 <sup>†</sup> 表示等位基因以及其它基因，<sup>‡</sup> 表示基因组物理距离最近的基因。



附图 1 循环 25OHD 与急性缺血性卒中单遗传工具因果估计

图注：此图表示移除可能与卒中相关混杂因素存在多效性的遗传变异后保留的遗传工具的单遗传工具估计效应（N = 60）。MR、25OHD、AIS 分别表示孟德尔随机化研究、循环 25-羟基维生素 D 以及急性缺血性卒中。



附图 2 循环 25OHD 与 LAS 留一法分析

图注：此图表示移除可能与卒中相关混杂因素存在多效性的遗传变异后保留的遗传工具的留一法异质性检验（N = 60）。MR、25OHD、LAS 分别表示孟德尔随机化研究、循环 25-羟基维生素 D 以及大动脉粥样硬化性卒中。

## 综述

### 基于序列优化缺血性卒中干预靶点

#### 摘要

优化缺血性卒中干预靶点有望缓解卒中高负荷与治疗低成功率之间的矛盾，药物开发的高投入、低成功率也对更为精准地定位合适的干预靶点提出了要求。解析遗传变异对卒中相关表型的影响有助于筛选药物干预靶点。测序技术的快速发展带来了大量生物序列数据，为探索卒中及卒中相关表型的分子决定因素提供了基础。同时，这也对遗传信息的解析方法提出了更高的要求。因此，本文从全基因组关联研究出发，结合分子数量性状位点，到因果关联及遗传网络结构分析，回顾了潜在干预靶点研究理念与方法。同时，本文也探讨了深度学习框架对解析遗传信息，优化潜在干预靶点的应用。综合多组学数据和多种方法有助于弥补当前方法学的缺陷，解析复杂的遗传调控网络中潜藏的因果效应；深度学习框架有望更为精准、全面定位卒中潜在干预靶点。

#### 1. 概述

缺血性卒中是我国致死、致残的首要病因，随着卒中危险因素普遍性暴露，其疾病负担呈爆发式增长<sup>[1-2]</sup>。卒中后常伴有神经系统的不可逆性损伤，如失语、瘫痪，部分患者还可出现癫痫等症状<sup>[3-5]</sup>。目前缺血性卒中主要通过早期再灌注进行治疗，包括静脉溶栓与机械取栓；然而其有限的治疗窗以及高出血转化风险大大降低了卒中的再通率<sup>[5-7]</sup>。据报道，约有一半的人群在接受血栓切除术后仍存在不同程度的残疾<sup>[8]</sup>。优化卒中干预靶点，有望获缓解卒中负荷与治疗成功率之间的矛盾。

作为常规药物开发的第一步，发现更为精准的疾病关键靶点至关重要，这是后续药物平台筛选、临床前以及临床研究评估的基础<sup>[9]</sup>。据研究报道，1000多项卒中相关靶点药物干预实验和100多项临床试验的药物并未最终落实于临床<sup>[10]</sup>。十分之一或更低的转化预期，数十年的时间预期和数十亿美元的资金投入预期，也对更快、更精确地定位合适的干预靶点提出了要求<sup>[11]</sup>。

生物的遗传密码隐含着卒中相关病理生理信息。理解基因对卒中以及卒中后表型的影响,有助于建立疾病分子分型,促进个体化医疗<sup>[12-14]</sup>;更为深入地解析这些信息也有助于优化潜在的干预靶点,建立新的治疗策略,提高疾病预后<sup>[15-16]</sup>。测序技术的快速发展及成本的下降增加了大规模序列数据的可获取性,促进了影响基因表达的遗传变异的识别,为研究疾病相关表型的分子决定因素提供了前所未有的机遇<sup>[17-18]</sup>。测序技术的更新有助于提高序列检测的准确性,提高细胞、组织的分辨率<sup>[18-20]</sup>,并有助于综合多组学解析遗传信息对卒中及卒中相关表型的影响<sup>[21]</sup>。涌现的大量生物序列数据对解析遗传对表型影响的方法提出了更进一步的要求。

长久以来,研究者们尝试从遗传信息挖掘出具有生物意义,驱动疾病病因学关键途径的潜在干预靶点。从表型的关联性到因果关系的推断,基于基因组探索潜在干预靶点的研究理念不断更新<sup>[22-23]</sup>。本文对以遗传变异探索潜在干预靶点的研究理念及方法进行讨论<sup>[24-28]</sup>,以期进一步优化卒中潜在干预靶点。

## 2. 关联研究

### 2.1 全基因组关联研究

卒中潜在干预靶点可以通过遗传变异与疾病之间的关联性筛选。遗传变异与表型的关联提示可能的致病基因以及机制,有助于提高我们对复杂表型的理解,优化分子分型,并提供可能的治疗干预策略<sup>[22]</sup>。

全基因组关联研究(GWAS, Genome-wide association Studies)基于遗传变异的连锁不平衡原理,利用复杂性状与基因型之间的相关性,选取关联显著性最强的 SNP(单核苷酸多态性, Single-nucleotide polymorphism)位点选取易感基因,剖析可能的生物学通路<sup>[24]</sup>。近十年来, GWAS 通过性状的遗传关联研究为脑血管病提供了许多可能的易感基因和机制<sup>[29-31]</sup>。例如, GWAS 发现的 JCAD 基因在后续研究中被证实其在血管内皮功能失调与动脉粥样硬化的病理过程中起到重要作用,这为心血管疾病的干预提供了研究方向<sup>[32]</sup>。由于疾病状态与健康时的遗传对表型的效应存在差异,有研究者还利用 GWAS, 针对卒中预后寻找相关遗传变异<sup>[33-34]</sup>。Mola-Caminal M 利用改良的 Rankin 评分评估卒中发病 3 月后的临床表现变化,通过 GWAS 发现 PATJ 基因与卒中预后存在关联;这为揭示潜在治疗靶点效应,并为其潜在机制的体外、体内研究提供了初步证据<sup>[34]</sup>。



但是, GWAS 所获取的易感基因不一定具有生物学意义。GWAS 通过统计学显著性定位遗传变异。由于连锁不平衡模糊了可能与疾病具有因果关联的遗传变异, 所获取的大多数信号与疾病或相关表型不存在直接生物学关联<sup>[24]</sup>。此外, 遗传变异依据距离锚定的基因并不一定介导了遗传变异对表型的影响<sup>[35]</sup>。

## 2.2 全转录组关联研究

通过整合基因表达信息, 如全转录组关联研究 (Transcriptome-wide association studies, TWAS)<sup>[36-37]</sup>、共定位<sup>[38-39]</sup>等, 可以针对 GWAS 存在的问题进一步缩小候选基因的范围。我们针对 TWAS 方法对该理念进行讨论。

TWAS 基于遗传变异通过调节基因表达对疾病或相关表型产生影响的假设, 即利用表达数量性状位点 (expression quantitative trait loci, eQTL) 来识别基因与表型间的关联<sup>[25,40]</sup>。TWAS 首先获取每个基因一定范围内 (如 500kb, 1Mb) 的 SNPs 所预测的基因表达量 (顺式表达数量性状位点, cis eQTL); 在此基础上预测个体的基因表达, 并建立遗传预测的基因表达与表型的关联<sup>[25,40]</sup>。TWAS 相较于 GWAS 具有更高的效能, 可以进一步优化靶点基因并识别可能的机制<sup>[41-42]</sup>。同时, 表型相关的基因表达具有细胞、组织特异性<sup>[43-44]</sup>。通过不同细胞、组织的基因表达参考数据, 可筛选特异性的致病基因<sup>[45-46]</sup>。

然而, TWAS 并不支持因果关系的分析。TWAS 为统计学关联研究, 筛选出的候选基因可能由于遗传变异或基因之间存在共调节的相互作用, 影响了 eQTL 与性状的关联<sup>[47-48]</sup>。此外, 基因表达由遗传、环境共同决定, 遗传决定的表达由顺式 eQTL 以及距离较远的反式 eQTL (trans eQTL) 共同作用<sup>[49]</sup>。目前广泛使用的常见 cis eQTL 仅解释 10% 的基因表达的遗传方差<sup>[50]</sup>。相反, trans eQTL 占据了约 70% 的遗传方差<sup>[23]</sup>。因此, 仅通过 cis eQTL 筛选也存在不足。

## 3. 因果关联

为寻找与表型存在因果关系的靶点, 孟德尔随机化 (Mendelian Randomization, MR) 框架是一个经济有效的选择<sup>[26]</sup>。MR 框架基于孟德尔遗传定律 (即在配子形成时等位基因独立分离), 假设群体中亲代等位基因随机分配给子代, 是一种天然的随机对照设计<sup>[51]</sup>。此外, 由于个体遗传信息在合子形成时期就已确定, MR 框架有效避免了混杂因素的干扰和可能的逆向因果关联<sup>[51]</sup>。

近年来, 孟德尔随机化框架被广泛用于探索与卒中可能存在因果关联的性状, 比如血脂、血压、血糖、肥胖、血钙等<sup>[52-56]</sup>。Georgakis MK 等人综合分析了循环细胞因子和生长因子水平与卒中的相关性性质, 结果发现遗传倾向于更高水平的单核细胞趋化因子 1 (Monocyte chemoattractant protein-1, MCP-1) 与增加的卒中风险存在因果关联, 其中以大动脉粥样硬化性卒中关联性最强, 每增加 1 个标准差的 log MCP-1 水平卒中风险增加 19%<sup>[57]</sup>, 是可能的干预靶点。

孟德尔随机化框架可以优化分子数量性状位点与表型间的因果关系<sup>[58-59]</sup>。分子数量性状位点 (molecular quantitative trait loci, mol QTL) 包括 eQTL, 蛋白数量性状位点 (protein QTL, pQTL), 以及染色体标记如甲基化数量性状位点 (methylation QTL, mQTL) 等<sup>[60-61]</sup>。作为基因与性状的中间表型, mol QTL 为疾病的遗传关联提供了功能性的证据, 可进一步优化潜在干预靶点<sup>[17,20]</sup>。

此外, MR 框架还具有筛选潜在靶点的可能副作用, 定位再利用药物靶点的价值<sup>[16]</sup>。Chong 以 MR 框架系统筛选循环蛋白与缺血性卒中的关联性状并分析蛋白可能的副作用; 结果发现清道夫受体类 A5 蛋白 (scavenger receptor class A5, SCARA5) 对心源性脑栓塞性卒中和蛛网膜下腔出血具有保护性的作用, 且并无明显的副作用; 因此, 可考虑 SCARA5 蛋白为一个新的潜在干预靶点<sup>[16]</sup>。MR 框架还可用于比较药物疗效<sup>[62-64]</sup>。Brian A. Ference 所在团队以药物干预靶点所在区域内所有血脂特征独立且相关的遗传变异位点构建遗传分数, 并通过 MR 框架准确预测不同降脂药物的疗效, 分析了潜在的治疗机制<sup>[62-64]</sup>。

孟德尔随机化框架在鉴别水平多效性的前提下可获得较为准确的因果估计结果<sup>[26]</sup>。但是一个 mol QTL 常与多个基因相互作用<sup>[65]</sup>, 基因组中大于 2/3 的 1Mb 阅读窗可对表型产生影响<sup>[66]</sup>。对表型有贡献的遗传变异在基因组中广泛分布, 这对 MR 因果变异推断提出了挑战<sup>[44]</sup>。Zhu 等人通过 SMR (Summary based Mendelian randomization) 法结合 HEIDI (heterogeneity in dependent instruments) 异质性测试除外连锁干扰, 缩小了潜在因果靶点的范围<sup>[67]</sup>; Porcu 等提出同时使用多个遗传变异作为工具变量, 多个基因表达性状作为暴露因素, 以评估基因对表型的因果关系, 提高了筛选的准确性<sup>[59]</sup>; 此外, 概率孟德尔随机化 Egger 法 (probabilistic Mendelian randomization -Egger, PMR-Egger) 参照 MR-Egger 法, 考虑水平多效性的前提下的因果效应, 进一步优化了潜在靶点<sup>[68]</sup>。尽管如此, 但水平多效性的问题仍然难以解决, 原因之一是在于遗传调控网络的复杂性<sup>[23]</sup>。

#### 4. 网络分析

Fisher 所提出的无穷小模型为理解遗传模式、开展 GWAS 等研究提供了理论基础<sup>[69-70]</sup>。但是，随着人们对遗传调节机制的深入研究，Fisher 模型不足以回答多少遗传变异为因果变异，哪些遗传变异对表型具有关键调节作用等问题；为此，Liu、Pritchard 等人提出全基因模型框架（图 R1），纳入遗传变异对基因表达的影响，以理解遗传结构对表型的调节机制<sup>[23-44]</sup>。该模型将基因分为核心基因与周围基因，遗传对表型的影响由核心基因的直接效应和周围基因的反式作用间接效应共同组成<sup>[23]</sup>。研究表明，大多数基因组都解释了一部分遗传方差<sup>[44]</sup>，大量弱遗传效应的 trans eQTL 综合解释了大部分的表型方差<sup>[23]</sup>，而仅有 0.1%-1% 常见 SNP 位点具有因果效应<sup>[71]</sup>。在复杂的遗传调控网络之中，部分周围基因调节了多个下游的核心基因，被称为周围主要调节基因；这些基因可能也是转录因子等调节分子的靶点，具有潜在的干预意义<sup>[23]</sup>。

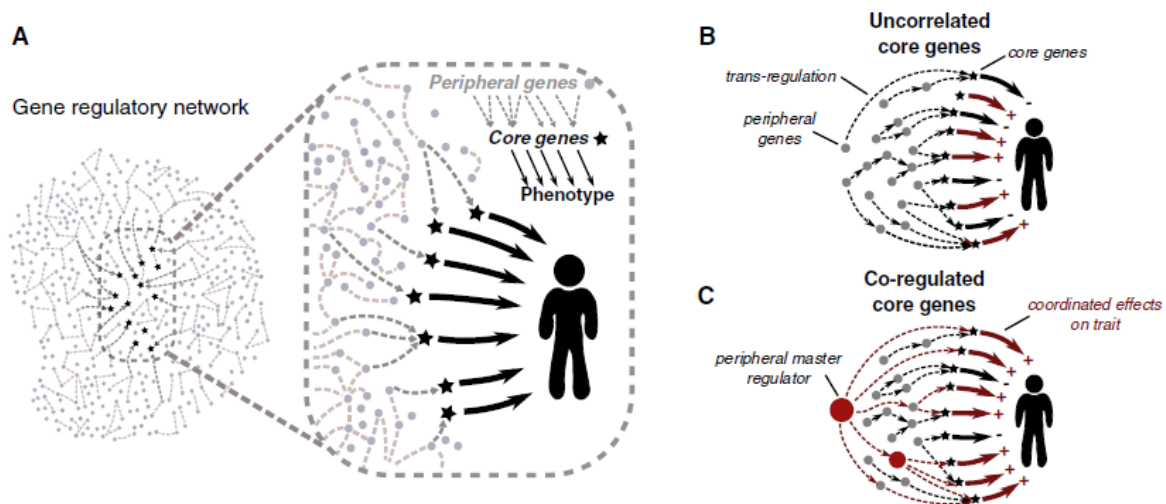


图 R1 遗传结构的全基因模型。

A. 基因调控网络中，表型由核心基因和周围基因组成，核心基因直接影响表型，周围基因通过调节其它基因对表型产生影响。B. 大多数周围基因的顺式调控变异是核心基因的弱效反式作用变异，其对核心基因的效应值方向不同；因此，大量典型的弱效周围变异解释了大部分的遗传方差。C. 基因调控网络中，部分周围基因驱动了多个核心基因的协同调控，即周围主要调控基因。引自 Cell. 2019;177(4):1022–1034.e6. <sup>[23]</sup>。

基因共表达网络将基因间关联性相似的基因聚类, 假设同一模块中的基因有相同或者相近的生物学通路<sup>[27]</sup>。差异共表达网络可识别不同条件下的表达模块<sup>[72]</sup>, 有助于识别细胞、组织、时间及疾病等相关的关键模块<sup>[73-75]</sup>, 并优化关键基因<sup>[76]</sup>。若是一个模块包含多个共同变化的核心基因 (图 R1.C), 该模块与疾病表型的变化密切相关, 可以作为关键模块<sup>[77]</sup>; 其中驱动核心基因变化的基因 (周围主要调节基因) 和多个基因高度连接, 即 hub 基因<sup>[23]</sup>, 有进一步探索的价值<sup>[78-79]</sup>。

加权基因相关网络分析 (Weighted Gene Correlation Network Analysis, WGCNA) 是目前使用最广泛的差异共表达网络分析方法<sup>[80-81]</sup>。WGCNA 法假定表达网络符合无尺度分布, 将基因间表达量的相关系数进行 N 次幂运算, 放大相关性的强弱差异, 将其转换为无尺度网络; 然后, WGCNA 将表达模式相似的基因进行等级聚类, 采用软阈值的策略得到不同的子模块。共表达网络中的节点表示基因, 节点与节点之间的边表示共表达相关性, 模块内的基因存在相似性, 可能为共调控、协同作用或者处于相同通路<sup>[80]</sup>。通过识别和表型有最高相关性的子模块, 再对其进行功能注释, 最终可得到研究对象的表达模式, 筛选功能相关的关键基因<sup>[82-83]</sup>。

经典的 WGCNA 的探索模式是利用 RNA (Ribose Nucleic Acid, 核糖核酸) 数据探索不同条件下的差异共表达网络结构, 通过匹配 mol QTL 数据可进一步评估具有遗传倾向的潜在关键干预靶点<sup>[84-85]</sup>。WGCNA 还可直接以 eQTL 等数据研究遗传预测的基因表达<sup>[35,75,86]</sup>。Gamazon 通过 WGCNA 尝试以遗传预测的共表达网络解析精神疾病遗传倾向性的可能生物机制, 优化了具有潜在功能意义的致病基因的选择<sup>[75]</sup>。Meng 利用外周血 eQTL 通过 WGCNA 探索新的基因 ASB16-AS1 与其它基因之间的联系, 发现 ASB16-AS1 与多种之前研究所证实的骨密度相关基因共表达; 体外研究证实该基因可调节成骨细胞生成相关基因 BMP2 和 ALPL<sup>[35]</sup>。这表明结合 WGCNA 有助于识别具有生物意义的调节靶点。此外, 差异共表达网络分析可整合多组学数据更全面地解析生物学的关联<sup>[87-90]</sup>。

然而, 差异共表达方法的结果受到数据质量以及数量的影响<sup>[91]</sup>, 测序技术的快速发展和数据质量的提高, 这种影响有望逐渐减小。需要注意的是, 虽然差异共表达网络可解析基因之间的网络调控结构, 但该方法不提供确切的因果关联。多个分析框架相结合有助于减少单个方法的不足, 综合筛选可能的干预靶点<sup>[27,92]</sup>。此外, 利用更优的方法, 如深度学习, 有望进一步优化干预靶点。

## 5. 深度学习

基因组学是一门数据科学，通过从基因组尺度的数据探索新的特性，从而发现疾病相关分子标志以及干预靶点<sup>[93]</sup>。基因组数据的复杂性使得以一定假设和专业基础知识为基础的方法不足以完全破译遗传变异与基因表达和表型的关联；相反，数据驱动的深度学习框架有望进一步解析基因表达的调控密码<sup>[28]</sup>。深度学习可分为监督学习和非监督学习两大类，其中监督学习按照基本结构可分为全连接神经网络、卷积神经网络、循环神经网络等，而非监督学习包括自编码器<sup>[94]</sup>。单层神经网络与逻辑回归、线性回归等模型相似，而多层神经网络（又称为深度神经网络）具有远优于前者的分类、预测能力<sup>[94]</sup>。

深度学习模型可以识别序列中的数据特征，通过大规模数据的训练以识别因果 eQTL 的优先顺序并推断疾病相关的因果变异<sup>[95-97]</sup>。深度学习模型可基于 DNA（脱氧核糖核苷酸，deoxyribonucleic acid）序列从头计算，预测遗传变异组织、细胞特异性的表达效应，包括罕见变异及之前从关联研究中未观察到的变异的效应；该方法还可从关联中抽提因果关系，获取遗传对表达的影响<sup>[97]</sup>。此外，深度学习还可进一步识别疾病相关调控网络中的驱动成分<sup>[98]</sup>。

深度学习成本低、准确性高，是研究表达调控的定量遗传学和实验方法的重要补充。生物数据数量和种类的增长及基于序列的深度学习模型的开源促进了深度学习在基因组学的应用<sup>[99-101]</sup>，迁移学习使得以小规模数据预测因果变异（eQTL，SNP）成为可能<sup>[102-103]</sup>。目前，深度学习尚处于起步阶段。在不久的将来，深度学习框架有望成为遗传表达调控研究的重要组成成分，并促进疾病分子标志以及潜在干预靶点的识别，并推进个体化医疗的进程。

## 6. 展望与小结

遗传学与精准医学的一个关键挑战是破译基因表达的调控密码<sup>[93]</sup>。结合多组学信息可以更为全面地解析生物调控机制，寻找可能的干预靶点。疾病相关 SNPs 富集于组织特异的活性染色质区域<sup>[104-106]</sup>。结合多层次信息，如甲基化等对基因表达的效应，可能会产生更多新的生物学机制相关的解释<sup>[27]</sup>。这有助于了解疾病的累积风险，提高临床风险预测<sup>[107]</sup>；这也有助于理解遗传调节机制，寻找具有因果效应的变异<sup>[108-109]</sup>。此外，测序技术的发展可以促进非编码 RNA

调控靶点的识别。考虑到多数表型相关 SNP 位点位于非编码区<sup>[110]</sup>，并且越来越多的研究发现 RNA 可作为干预靶点<sup>[111-115]</sup>；联合遗传变异与非编码 RNA 可促进二者与表达调控之间联系的解析，寻找可能的 RNA 干预靶点<sup>[116-117]</sup>。

目前探索具有潜在因果效应的干预靶点的方法存在不足。人们期望从遗传调控网络中解析出在生物相关机制中发挥关键因果效应的变异<sup>[58-59]</sup>。遗传调控网络的复杂性，如连锁不平衡、多效性等，使得这一期望难以实现。联合孟德尔随机化框架、差异共表达网络分析等模型有助于解决这一问题<sup>[75]</sup>。深度学习模型的发展与应用有望弥补目前计算方法的不足。结合分子数量性状位点，可进一步探索遗传变异与表型之间的关系，更为全面准确地破译遗传调控网络。这有助于筛选卒中相关表型的分子标志，获取疾病更为准确的关键干预靶点，推进精准医疗的进程。

## 综述参考文献

- [1] Johnson C O, Nguyen M, Roth G A, et al. Global, regional, and national burden of stroke, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016[J]. *Lancet Neurol*. 2019, 18(5): 439-458.
- [2] 王陇德, 刘建民, 杨弋, 等. 我国脑卒中防治仍面临巨大挑战——《中国脑卒中防治报告 2018》概要[J]. *中国循环杂志*. 2019, 34(02): 105-119.
- [3] Chen Z, Churilov L, Chen Z, et al. Association between implementation of a code stroke system and poststroke epilepsy[J]. *Neurology*. 2018, 90(13): e1126-e1133.
- [4] Shimamura N, Katagai T, Kakuta K, et al. Rehabilitation and the neural network after stroke[J]. *Transl Stroke Res*. 2017, 8(6): 507-514.
- [5] Hankey G J. Stroke[J]. *Lancet*. 2017, 389(10069): 641-654.
- [6] 中华医学会神经病学分会和中华医学会神经病学分会脑血管病学组. 中国急性缺血性脑卒中诊治指南 2018[J]. *中华神经科杂志*. 2018, 51(9): 666-682.
- [7] Thiebaut A M, Gauberti M, Ali C, et al. The role of plasminogen activators in stroke treatment: Fibrinolysis and beyond[J]. *Lancet Neurol*. 2018, 17(12): 1121-1132.
- [8] Nogueira R G, Jadhav A P, Haussen D C, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct[J]. *N Engl J Med*. 2018, 378(1): 11-21.
- [9] Fogel D B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review[J]. *Contemp Clin Trials Commun*. 2018, 11: 156-164.
- [10] Patel R A G and McMullen P W. Neuroprotection in the treatment of acute ischemic stroke[J]. *Prog Cardiovasc Dis*. 2017, 59(6): 542-548.
- [11] Mullin R. Cost to develop new pharmaceutical drug now exceeds \$2.5 B[J]. *Sci Am*. 2014, 24.
- [12] Eichler E E. Genetic variation, comparative genomics, and the diagnosis of disease[J]. *N Engl J Med*. 2019, 381(1): 64-74.

- [13] Peterson J F, Roden D M, Orlando L A, et al. Building evidence and measuring clinical outcomes for genomic medicine[J]. *Lancet*. 2019, 394(10198): 604-610.
- [14] Wise A L, Manolio T A, Mensah G A, et al. Genomic medicine for undiagnosed diseases[J]. *Lancet*. 2019, 394(10197): 533-540.
- [15] Wilson M R, Sample H A, Zorn K C, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis[J]. *N Engl J Med*. 2019, 380(24): 2327-2340.
- [16] Chong M, Sjaarda J, Pigeyre M, et al. Novel drug targets for ischemic stroke identified through Mendelian randomization analysis of the blood proteome[J]. *Circulation*. 2019, 140(10): 819-830.
- [17] Neumeyer S, Hemani G and Zeggini E. Strengthening causal inference for complex disease using molecular quantitative trait loci[J]. *Trends Mol Med*. 2020, 26(2): 232-241.
- [18] Stark R, Grzelak M and Hadfield J. RNA sequencing: the teenage years[J]. *Nat Rev Genet*. 2019, 20(11): 631-656.
- [19] Potter S S. Single-cell RNA sequencing for the study of development, physiology and disease[J]. *Nat Rev Nephrol*. 2018, 14(8): 479-492.
- [20] Taylor K, Davey Smith G, Relton C L, et al. Prioritizing putative influential genes in cardiovascular disease susceptibility by applying tissue-specific Mendelian randomization[J]. *Genome Med*. 2019, 11(1): 6.
- [21] Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits[J]. *Nat Commun*. 2018, 9(1): 918.
- [22] van Rheenen W, Peyrot W J, Schork A J, et al. Genetic correlations of polygenic disease traits: from theory to practice[J]. *Nat Rev Genet*. 2019, 20(10): 567-581.
- [23] Liu X, Li Y I and Pritchard J K. Trans effects on gene expression can drive omnigenic inheritance[J]. *Cell*. 2019, 177(4): 1022-1034.e1026.
- [24] Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies[J]. *Nat Rev Genet*. 2019, 20(8): 467-484.
- [25] Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies[J]. *Nat Genet*. 2016, 48(3): 245-252.



- [26] Emdin C A, Khera A V and Kathiresan S. Mendelian randomization[J]. *JAMA*. 2017, 318(19): 1925-1926.
- [27] Van Dam S, Vosa U, van der Graaf A, et al. Gene co-expression analysis for functional classification and gene-disease predictions[J]. *Brief Bioinform*. 2018, 19(4): 575-592.
- [28] Eraslan G, Avsec Ž, Gagneur J, et al. Deep learning: new computational modelling techniques for genomics[J]. *Nat Rev Genet*. 2019, 20(7): 389-403.
- [29] Chung J, Marini S, Pera J, et al. Genome-wide association study of cerebral small vessel disease reveals established and novel loci[J]. *Brain*. 2019, 142(10): 3176-3189.
- [30] Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes[J]. *Nat Genet*. 2018, 50(4): 524-537.
- [31] Kunkle B W, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing[J]. *Nat Genet*. 2019, 51(3): 414-430.
- [32] 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.
- [33] Torres-Aguila N P, Carrera C, Giese A K, et al. Genome-wide association study of white blood cell counts in patients with ischemic stroke[J]. *Stroke*. 2019, 50(12): 3618-3621.
- [34] Mola-Caminal M, Carrera C, Soriano-Tarraga C, et al. PATJ low frequency variants are associated with worse ischemic stroke functional outcome[J]. *Circ Res*. 2019, 124(1): 114-120.
- [35] Meng X H, Chen X D, Greenbaum J, et al. Integration of summary data from GWAS and eQTL studies identified novel causal BMD genes with functional predictions[J]. *Bone*. 2018, 113: 41-48.
- [36] Thériault S, Gaudreault N, Lamontagne M, et al. A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis[J]. *Nat Commun*. 2018, 9(1): 988.

- [37] Yang J, Yan B, Fan Y, et al. Integrative analysis of transcriptome-wide association study and gene expression profiling identifies candidate genes associated with stroke[J]. *PeerJ*. 2019, 7: e7435.
- [38] Hormozdiari F, van de Bunt M, Segrè A V, et al. Colocalization of GWAS and eQTL signals detects target genes[J]. *Am J Hum Genet*. 2016, 99(6): 1245-1260.
- [39] Franceschini N, Giambartolomei C, de Vries P S, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes[J]. *Nat Commun*. 2018, 9(1): 5141.
- [40] Wainberg M, Sinnott-Armstrong N, Mancuso N, et al. Opportunities and challenges for transcriptome-wide association studies[J]. *Nat Genet*. 2019, 51(4): 592-599.
- [41] Lindstrom S, Wang L, Smith E N, et al. Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism[J]. *Blood*. 2019, 134(19): 1645-1657.
- [42] Li Y I, Wong G, Humphrey J, et al. Prioritizing Parkinson's disease genes using population-scale transcriptomic data[J]. *Nat Commun*. 2019, 10(1): 994.
- [43] Ongen H, Brown A A, Delaneau O, et al. Estimating the causal tissues for complex traits and diseases[J]. *Nat Genet*. 2017, 49(12): 1676-1683.
- [44] Boyle E A, Li Y I and Pritchard J K. An expanded view of complex traits: From polygenic to omnigenic[J]. *Cell*. 2017, 169(7): 1177-1186.
- [45] Mancuso N, Gayther S, Gusev A, et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions[J]. *Nat Commun*. 2018, 9(1): 4079.
- [46] Ratnapriya R, Sosina O A, Starostik M R, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration[J]. *Nat Genet*. 2019, 51(4): 606-610.
- [47] Yao C, Joehanes R, Johnson A D, et al. Dynamic role of trans regulation of gene expression in relation to complex traits[J]. *Am J Hum Genet*. 2017, 100(4): 571-580.
- [48] Brynedal B, Choi J, Raj T, et al. Large-scale trans-eQTLs affect hundreds of transcripts and mediate patterns of transcriptional co-regulation[J]. *Am J Hum Genet*. 2017, 100(4): 581-591.

- [49] Lloyd-Jones L R, Holloway A, McRae A, et al. The genetic architecture of gene expression in peripheral blood[J]. *Am J Hum Genet.* 2017, 100(2): 228-237.
- [50] Grundberg E, Small K S, Hedman A K, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins[J]. *Nat Genet.* 2012, 44(10): 1084-1089.
- [51] Davies N M, Holmes M V and Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians[J]. *BMJ.* 2018, 362: k601.
- [52] Sun L, Clarke R, Bennett D, et al. Causal associations of blood lipids with risk of ischemic stroke and intracerebral hemorrhage in Chinese adults[J]. *Nat Med.* 2019, 25(4): 569-574.
- [53] Parish S, Arnold M, Clarke R, et al. Assessment of the role of carotid atherosclerosis in the association between major cardiovascular risk factors and ischemic stroke subtypes[J]. *JAMA Netw Open.* 2019, 2(5): e194873-e194873.
- [54] Larsson S C, Scott R A, Traylor M, et al. Type 2 diabetes, glucose, insulin, BMI, and ischemic stroke subtypes: Mendelian randomization study[J]. *Neurology.* 2017, 89(5): 454-460.
- [55] Marini S, Merino J, Montgomery B E, et al. Mendelian randomization study of obesity and cerebrovascular disease[J]. *Ann Neurol.* 2020, 87(4): 516-524.
- [56] Larsson S C, Traylor M, Burgess S, et al. Serum magnesium and calcium levels in relation to ischemic stroke: Mendelian randomization study[J]. *Neurology.* 2019, 92(9): e944-e950.
- [57] Georgakis M K, Gill D, Rannikmae K, et al. Genetically determined levels of circulating cytokines and risk of stroke[J]. *Circulation.* 2019, 139(2): 256-268.
- [58] Qi T, Wu Y, Zeng J, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood[J]. *Nat Commun.* 2018, 9(1): 2282.
- [59] Porcu E, Rueger S, Lepik K, et al. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits[J]. *Nat Commun.* 2019, 10(1): 3300.
- [60] Hormozdiari F, Gazal S, van de Geijn B, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits[J]. *Nat Genet.* 2018, 50(7): 1041-1047.

- [61] Suhre K, Arnold M, Bhagwat A M, et al. Connecting genetic risk to disease end points through the human blood plasma proteome[J]. *Nat Commun*. 2017, 8: 14357.
- [62] Ference B A, Kastelein J J P, Ginsberg H N, et al. Association of genetic variants related to CETP inhibitors and statins with lipoprotein levels and cardiovascular risk[J]. *JAMA*. 2017, 318(10): 947-956.
- [63] Ference B A, Ray K K, Catapano A L, et al. Mendelian randomization study of ACLY and cardiovascular disease[J]. *N Engl J Med*. 2019, 380(11): 1033-1042.
- [64] Ference B A, Robinson J G, Brook R D, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes[J]. *N Engl J Med*. 2016, 375(22): 2144-2153.
- [65] Richardson T G, Hemani G, Gaunt T R, et al. A transcriptome-wide Mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome[J]. *Nat commun*. 2020, 11(1): 185-185.
- [66] Loh P R, Tucker G, Bulik-Sullivan B K, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts[J]. *Nat Genet*. 2015, 47(3): 284-290.
- [67] Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets[J]. *Nat Genet*. 2016, 48: 481-487.
- [68] Yuan Z, Zhu H, Zeng P, et al. Testing and controlling for horizontal pleiotropy with the probabilistic Mendelian randomization in transcriptome-wide association studies[J]. *bioRxiv*. 2019, 691014: 691014.
- [69] Fisher R A. The correlation between relatives on the supposition of Mendelian inheritance[J]. *Trans R Soc Edinb Earth Sci*. 1918, 52(2): 399-433.
- [70] Barton N H, Etheridge A M and Véber A. The infinitesimal model: Definition, derivation, and implications[J]. *Theor Popul Biol*. 2017, 118: 50-73.
- [71] Khera A V, Chaffin M, Aragam K G, 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.
- [72] de la Fuente A. From 'differential expression' to 'differential networking' - identification of dysfunctional regulatory networks in diseases[J]. *Trends Genet*. 2010, 26(7): 326-333.

- [73] Schafer S T, Paquola A C M, Stern S, et al. Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons[J]. *Nat Neurosci.* 2019, 22(2): 243-255.
- [74] Tasic B, Yao Z, Graybuck L T, et al. Shared and distinct transcriptomic cell types across neocortical areas[J]. *Nature.* 2018, 563(7729): 72-78.
- [75] Gamazon E R, Zwinderman A H, Cox N J, et al. Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits[J]. *Nat Genet.* 2019, 51(6): 933-940.
- [76] Bhuva D D, Cursons J, Smyth G K, et al. Differential co-expression-based detection of conditional relationships in transcriptional data: comparative analysis and application to breast cancer[J]. *Genome Biol.* 2019, 20(1): 236.
- [77] Rangaraju S, Dammer E B, Raza S A, et al. Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease[J]. *Mol Neurodegener.* 2018, 13(1): 24.
- [78] Gandal M J, Haney J R, Parikshak N N, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap[J]. *Science.* 2018, 359(6376): 693-697.
- [79] Lorsch Z S, Hamilton P J, Ramakrishnan A, et al. Stress resilience is promoted by a Zfp189-driven transcriptional network in prefrontal cortex[J]. *Nat Neurosci.* 2019, 22(9): 1413-1423.
- [80] Langfelder P and Horvath S. WGCNA: An R package for weighted correlation network analysis[J]. *BMC bioinformatics.* 2008, 9(1): 559.
- [81] Kakati T, Bhattacharyya D K, Barah P, et al. Comparison of methods for differential co-expression analysis for disease biomarker prediction[J]. *Comput Biol Med.* 2019, 113: 103380.
- [82] Chowdhury H A, Bhattacharyya D K and Kalita J K. Differential co-expression analysis of gene expression: A survey of best practices[J]. *IEEE/ACM Trans Comput Biol Bioinform.* 2019.
- [83] Li H, Wang X, Lu X, et al. Co-expression network analysis identified hub genes critical to triglyceride and free fatty acid metabolism as key regulators of age-related vascular dysfunction in mice[J]. *Aging (Albany NY).* 2019, 11(18): 7620-7638.

- [84] Walker R L, Ramaswami G, Hartl C, et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms[J]. *Cell*. 2019, 179(3): 750-771.e722.
- [85] Kia D A, Zhang D, Guelfi S, et al. Integration of eQTL and Parkinson's disease GWAS data implicates 11 disease genes[J]. *bioRxiv*. 2019: 627216.
- [86] Liu B H. Differential coexpression network analysis for gene expression data[J]. *Methods Mol Biol*. 2018, 1754: 155-165.
- [87] Zhang X, Sun R and Liu L. Potentially critical roles of TNPO1, RAP1B, ZDHHC17, and PPM1B in the progression of coronary atherosclerosis through microarray data analysis[J]. *J Cell Biochem*. 2019, 120(3): 4301-4311.
- [88] Duan C, Cao Z, Tang F, et al. miRNA-mRNA crosstalk in myocardial ischemia induced by calcified aortic valve stenosis[J]. *Aging (Albany NY)*. 2019, 11: 448-466.
- [89] Wang Q, Roy B and Dwivedi Y. Co-expression network modeling identifies key long non-coding RNA and mRNA modules in altering molecular phenotype to develop stress-induced depression in rats[J]. *Transl Psychiatry*. 2019, 9(1): 125.
- [90] Pei G, Chen L and Zhang W, WGCNA application to proteomic and metabolomic data analysis, in: *Methods Enzymol*, Elsevier, 2017, pp. 135-158.
- [91] Ballouz S, Verleyen W and Gillis J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers[J]. *Bioinformatics*. 2015, 31(13): 2123-2130.
- [92] Yurko R, G'Sell M, Roeder K, et al. A selective inference approach for FDR control using multi-omics covariates yields insights into disease risk[J]. *bioRxiv*. 2020: 806471.
- [93] Hieter P and Boguski M. Functional genomics: It's all how you read it[J]. *Science*. 1997, 278(5338): 601-602.
- [94] LeCun Y, Bengio Y and Hinton G. Deep learning[J]. *Nature*. 2015, 521: 436-444.
- [95] Zhou J and Troyanskaya O G. Predicting effects of noncoding variants with deep learning-based sequence model[J]. *Nat Methods*. 2015, 12(10): 931-934.
- [96] Liu Y, Wang D, He F, et al. Phenotype prediction and genome-wide association study using deep convolutional neural network of soybean[J]. *Front Genet*. 2019, 10:

1091.

[97] Zhou J, Theesfeld C L, Yao K, et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk[J]. *Nat Genet.* 2018, 50(8): 1171-1179.

[98] Camacho D M, Collins K M, Powers R K, et al. Next-generation machine learning for biological networks[J]. *Cell.* 2018, 173(7): 1581-1592.

[99] Zampieri G, Vijayakumar S, Yaneske E, et al. Machine and deep learning meet genome-scale metabolic modeling[J]. *PLoS Comput Biol.* 2019, 15(7): e1007084.

[100] Chen K M, Cofer E M, Zhou J, et al. Selene: A PyTorch-based deep learning library for sequence data[J]. *Nat Methods.* 2019, 16(4): 315-318.

[101] Zhou H, Sinsheimer J S, Bates D M, et al. OPENMENDEL: A cooperative programming project for statistical genetics[J]. *Hum Genet.* 2020, 139(1): 61-71.

[102] Künzel S R, Stadie B C, Vemuri N, et al. Transfer learning for estimating causal effects using neural networks[J]. *arXiv preprint arXiv.* 2018: 1808.07804.

[103] Zhang Z, Zhao Y, Liao X, et al. Deep learning in omics: a survey and guideline[J]. *Brief Funct Genomics.* 2019, 18(1): 41-57.

[104] Finucane H K, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics[J]. *Nat Genet.* 2015, 47(11): 1228-1235.

[105] Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes[J]. *Nature.* 2015, 518(7539): 317-330.

[106] Farh K K, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants[J]. *Nature.* 2015, 518(7539): 337-343.

[107] Westerman K, Sebastiani P, Jacques P, et al. DNA methylation modules associate with incident cardiovascular disease and cumulative risk factor exposure[J]. *Clin Epigenetics.* 2019, 11(1): 142.

[108] Taylor D L, Jackson A U, Narisu N, et al. Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle[J]. *Proc Natl Acad Sci U S A.* 2019, 116(22): 10883-10888.

[109] Hannon E, Gorrie-Stone T J, Smart M C, et al. Leveraging DNA-methylation quantitative-trait Loci to characterize the relationship between methylomic variation,

- gene expression, and complex traits[J]. *Am J Hum Genet.* 2018, 103(5): 654-665.
- [110] Pickrell J K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits[J]. *Am J Hum Genet.* 2014, 94(4): 559-573.
- [111] Lu D and Thum T. RNA-based diagnostic and therapeutic strategies for cardiovascular disease[J]. *Nat Rev Cardiol.* 2019, 16(11): 661-674.
- [112] Eken S M, Jin H, Chernogubova E, et al. MicroRNA-210 enhances fibrous cap stability in advanced atherosclerotic lesions[J]. *Circ Res.* 2017, 120(4): 633-644.
- [113] Han B, Zhang Y, Zhang Y, et al. Novel insight into circular RNA HECTD1 in astrocyte activation via autophagy by targeting MIR142-TIPARP: implications for cerebral ischemic stroke[J]. *Autophagy.* 2018, 14(7): 1164-1184.
- [114] Wang Y, Luo Y, Yao Y, et al. Silencing the lncRNA Mac1pil in pro-inflammatory macrophages attenuates acute experimental ischemic stroke via LCP1 in mice[J]. *J Cereb Blood Flow Metab.* 2020, 40(4): 747-759.
- [115] McNeill E M, Warinner C, Alkins S, et al. The conserved microRNA miR-34 regulates synaptogenesis via coordination of distinct mechanisms in presynaptic and postsynaptic cells[J]. *Nat Commun.* 2020, 11(1): 1092.
- [116] Liu Z, Ran Y, Tao C, et al. Detection of circular RNA expression and related quantitative trait loci in the human dorsolateral prefrontal cortex[J]. *Genome Biol.* 2019, 20(1): 99.
- [117] Xia W, Zhu X W, Mo X B, et al. Integrative multi-omics analysis revealed SNP-lncRNA-mRNA (SLM) networks in human peripheral blood mononuclear cells[J]. *Hum Genet.* 2017, 136(4): 451-462.



## 致谢

七年，如梭似箭，感谢我亲爱的家人、朋友一路陪伴。经历七年光阴，初入大学殿堂的懵懂稚子逐渐成长为一名合格的医生。曾经一次次迷茫，一次次彷徨，也曾抬头望天，试问人生何往。感谢你们的陪伴，让我逐渐明晰了前进的方向；也因为有你们的陪伴，在跌倒中我积蓄了内心的力量。

感谢我的益友良师，郝峻巍教授。是他，将我的视野引至科学前沿；是他，将我的脚踪导向临床一线。曾经我以为他定如巍巍高山，遂战战兢兢，亦步亦趋；但他温润的性情和坚定的信念，让我在追随的同时辨明内心所求。他用谦逊渲染了我们的人生态度，用笃实、勤恳锻造了我们的工作风格，又用精致、严谨雕琢了我们探索真理的拳拳赤子之心。一次次的交流如今仍然历历在目，或幽默，或严肃，带我走出迷雾；每一次的反馈都使我更进一步，让我逐渐学会聆听他人的声音，逐渐学会丈量自己的脚步。

感谢刘桂友老师、蒋薇师姐在完善论文期间所给予的帮助。刘老师拥有着深厚的生信积淀，令人信服；感谢他认真回答我的问题，解开了困扰我良久的疑惑。感谢刘老师细致地检查论文，耐心指出其中的问题以及修改方向。感谢蒋薇师姐在繁忙的临床工作之余，还抽出时间收集并检查我们的论文。同时，还要感谢 Stephen Burgess 教授为我解疑，感谢 Vince Forgetta 博士以及 Jiang Xia 博士一丝不苟地回答问题。此外，还要感谢 John McGrath 教授，Cassianne Robinson-Cohen 教授和 MEGASTROKE 协会无私地提供数据。你们的支持让我的硕士论文工作能够顺利推进。

感谢实验室的师兄师姐们对我的指导以及关照。崔盼师姐对生活充满热爱，她自律的生活方式吸引着我追逐的步伐，她热情的相交逐渐消除了我内心怯懦的畏惧。感谢崔盼师姐无私指导我的研究方向，且让我在临床研究中有所尝试，小小满足了我心中对临床研究的渴望；感谢她对我的信任与放手，让我独立完成实验操作。感谢韩冉冉师姐、李赫师姐、李千师姐、李道静师姐为我提供了实验技能的学习，感谢蒋薇师姐、李敏淑老师、刘晔老师细致地指导我实验技术操作。此外，感谢王俊杰师姐以其踏实的处事风格平静了我浮躁的内心，感谢郑凯师兄在为人处事方面给我的提点。感谢李赫师姐、董晓柳师姐 22 楼的陪伴，感谢我亲爱的同门对我的热情相助。

杨丽主任、张超主任、董晓柳师姐扎实的临床知识、丰富的临床经验给了我非常大的帮助。司家瑞老师的指导提高了我在深度学习领域的学识，这一过程也提高了我略显木讷的情商。张平教授的教导为我坚守初心埋下了种子。感谢留守七年小伙伴们的抱团取暖，感谢天津同伴们的相互扶持，也感谢我亲爱的家人对我的放手与支持。你们的指导、陪伴与信任，我感恩于心，铭记于心。

最后，感谢答辩委员会的评审专家在百忙之中花费时间为我们提出宝贵的指导意见。

医路漫漫，也许有幸继续在这个团队中成长，也许无缘与你们继续相伴；不论如何，我会带着你们的祝福与期望一直前进。我相信，我虽非人中菱角，虽暂无鲁班之工，但一定会走出自己的价值，不辜负郝老师的培养。

## 个人简历

姓名：林灵敏

性别：女

出生年月：1993 年 11 月

籍贯：浙江

### 主要学习和工作经历：

2013 年 8 月-2020 年 7 月

天津医科大学 临床医学（本科硕士连续培养）专业 获硕士学位

### 在学期间发表论文情况：

Si J R, **Lin L M**, Chen H L, et al. Dynamic Glucose Prediction and Health Management of Diabetes Mellitus Based on Deep Learning[J]. *Basic Clin Pharmacol Toxicol*. 2020, 126: 47.

### 在学期间参加科研情况：

2017-2019, 基于深度学习的糖尿病血糖预测以及健康管理, TMUROP2017-03 天津医科大学大学生科研创新项目, 1000, 第一承担人