

doi: 10.3969/j.issn.1674-1242.2023.04.012

# 基于数量性状转座子数据鉴定 焦虑障碍易感基因的研究

张英涵, 林关宁

(上海交通大学生物医学工程学院, 上海, 200030)

**【摘要】目的** 鉴定焦虑障碍的风险基因。**方法** 基于因果基因集精细映射法, 整合来自英国生物银行关于焦虑障碍疾病的全基因组关联研究数据与来自大样本人脑表达数量性状转座子数据来寻找和焦虑障碍相关的风险基因。**结果和结论** 该研究筛选出了 FAM167A、MSRA、TDH 等 12 个基因作为潜在的可能导致焦虑障碍的风险基因, 证明了它们与焦虑障碍疾病的相关性, 也为后续开发焦虑障碍的临床诊疗靶点提供了可能的方向。

**【关键词】** 焦虑障碍; 全基因组关联研究; 数量性状转座子; FLOT1; 风险基因

**【中图分类号】** Q811.4

**【文献标志码】** A

文章编号: 1674-1242 (2023) 04-0412-05

## Identification of Anxiety Disorder Susceptibility Genes Based on Quantitative Trait Loci Data

ZHANG Yinghan, LIN Guanning

(School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China)

**【Abstract】Objective** To identify risk genes for anxiety disorders. **Methods** Using the Fine-mapping of causal gene sets, the data from the UK Biobank genome-wide association study on anxiety disorders were integrated with quantitative trait locus data from a large sample of human brain expression to find risk genes associated with anxiety disorders. **Results and conclusions** In this study, 12 genes including FAM167A, MSRA and TDH were screened for the potential risk genes that may lead to anxiety disorders, the correlation between them and anxiety disorders is proved, and it also provides a possible direction for the subsequent development of clinical diagnosis and treatment targets for anxiety disorders.

**【Key words】** Anxiety Disorder; Genome-Wide Association Study; Quantitative Trait Locus; FLOT1; Risk Gene

### 0 引言

焦虑症是最普遍的精神疾病之一, 目前全球患病率为 7.3%, 并且被认为是在世界范围内导致残疾的主要原因。焦虑症有多种类型, 如广泛性焦虑症、恐慌症、社交焦虑症、特定恐惧症、广场恐惧症、分离焦

虑症、选择性缄默症, 以及由药物引起的焦虑症<sup>[1]</sup>。焦虑障碍存在明显的家庭聚集现象, 这在很早就引起了研究者的注意。同父同母的兄弟姐妹从出生开始, 对惊吓的反应和产生恐惧情绪的强度就不同。在出生的第一年, 同卵双胞胎对陌生人的恐惧程度比异卵双

收稿日期: 2022-11-29。

基金项目: 国家自然科学基金 (81971292)。

作者简介: 张英涵 (2000—), 男, 吉林人, 硕士研究生, 从事精神疾病遗传学研究, 邮箱 (E-mail): yhzhang@surgery.cuhk.edu.hk。

通信作者: 林关宁, 男, 博士研究生导师, 教授, 邮箱 (E-mail): nickgnlin@sjtu.edu.cn。

胞胎更相似<sup>[2]</sup>。随着过去十几年全基因组关联研究 (Genome-Wide Association Studies, GWAS) 等高通量技术的出现, 对遗传学的研究成倍增加。我们对许多复杂疾病的遗传基础有了更深入的了解, 其中就有包含焦虑症在内的大量精神疾病。研究发现, 可靠的遗传预测因子可能对日常临床实践产生显著影响。目前已经发现的最重要的遗传因子有编码色氨酸羟化酶、儿茶酚 O-甲基转移酶 (COMT)、单胺氧化酶-A (MAOA)、血清素转运蛋白 (5-HTT)、去甲肾上腺素转运蛋白 (NET)、多巴胺转运蛋白的基因, 以及与 DAT、单胺受体 (5-HT1A、5-HT2A、5-HT6、5-HT3A、5-HT3B、 $\beta$ 1-肾上腺素受体)、多巴胺 (DA) 受体、G 蛋白  $\beta$ 3 亚基、促肾上腺皮质激素释放激素 (CRH) 受体 I (CRHR1)、糖皮质激素受体、血管紧张素转换酶、昼夜运动输出周期基因 (CLOCK)、一氧化氮合酶、白细胞介素 (IL) - $\beta$  和脑源性神经营养因子 (BDNF) 有关的基因<sup>[3]</sup>。

基因表达变异是一系列疾病易感性的分子原因: 这种表达变异是可遗传的, 并且已经确定与 1 型糖尿病、过敏性哮喘或肥胖等疾病有关。而表达数量性状基因座 (expressed Quantitative Trait Locus, eQTL) 可将种群中的遗传变异与基因表达的变异联系起来, 以查明影响基因表达的多态性遗传区域<sup>[4]</sup>。使用 DNA 微阵列或深度测序, 可以在基因组水平上测量遗传变异和基因表达。因此, 使用 eQTL 方法基本可以研究基因组中所有区域与所有基因表达的关联。本研究通过整合焦虑障碍的 GWAS 数据与大样本人脑 eQTL 数据, 推算出表达量变化会导致焦虑障碍风险变化的基因, 再筛选出关键的易感基因, 并完成对关键基因的通路分析及功能分析, 试图带来对焦虑障碍发病机制的新解释, 对焦虑障碍的早期诊断与治疗有着积极意义。

## 1 方法

### 1.1 FOCUS 算法

FOCUS 算法<sup>[4]</sup>基于贝叶斯方法, 精细映射的贝叶斯方法专门用于关注最有可能成为因果关系的 SNP (Single Nucleotide Polymorphism, 单核苷酸多态性)。SNP 的 PIP (Posterior Inclusion Probability, 后验包含概率) 是在任何模型中包含 SNP 作为因果关系的概率。对于 SNP<sub>*j*</sub>, PIP 由包含 SNP<sub>*j*</sub> 作为因果的所有模型的后验总和计算。

$$\text{PIP}_j = P(c_j = 1 | D) = \sum_{M, c_j=1} P(M | D)$$

FOCUS 算法能够实现的先决条件是, 已知风险区域的全转录组基因研究 (Transcriptome-Wide Association Studies, TWAS) 基因-性状关联统计与预测模型中使用的 SNP 和 eQTL 权重之间的连锁不平衡 (Linkage Disequilibrium, LD) 函数存在相关关系。这类似 GWAS 中的 LD 标记, 其中一个区域内的 LD 在标签 SNP 处诱导关联 (产生传统的曼哈顿图)。即使在单个 SNP 因果影响基因表达的最简单情况下, 也会影响性状, eQTL 预测模型中使用的 SNP 中的 LD 也会在该区域附近的非因果基因处诱导显著的基因-性状关联。在存在多个因果 SNP 和基因的情况下, 标记效应会进一步加剧。尽管由于不同的 eQTL 遗传调控, 基因之间具有不重叠的预测权重, 但具有非零预测权重的 SNP 之间的 LD 会在 TWAS 统计中诱导相关性。估计预测表达的相关结构在附近的基因中, 可以对基因-性状关联进行统计精细映射。

为了在给定 TWAS 统计量  $\mathbf{z}_{\text{TWAS}}$  的情况下推断出哪些基因是因果的, FOCUS 计算任何一组因果基因  $c$  的后验分布为

$$\begin{aligned} \Pr(c | \mathbf{z}_{\text{TWAS}}, \lambda_{\text{SNP}}, \boldsymbol{\Omega}, \mathbf{V}, n\sigma_c^2) \\ = \frac{N(\mathbf{z}_{\text{TWAS}} | \boldsymbol{\Omega}^T \mathbf{V} \lambda_{\text{SNP}}, \mathcal{V} \mathbf{D}_c \mathcal{V} + \mathcal{V}) \Pr(c)}{\sum_{c' \in \mathcal{C}} N(\mathbf{z}_{\text{TWAS}} | \boldsymbol{\Omega}^T \mathbf{V} \lambda_{\text{SNP}}, \mathcal{V} \mathbf{D}_{c'} \mathcal{V} + \mathcal{V}) \Pr(c')} \end{aligned}$$

式中,  $c$  是长度为  $m$  的所有二进制向量的集合。假设每个因果指标  $c_i \sim \text{Bern}(p)$  都有一个伯努利先验。在实践中, 设置  $p = 1 \times 10^{-3}$ 。计算第  $i$  个基因的边际 PIP 为

$$\begin{aligned} \text{PIP}(c_i = 1 | \mathbf{z}_{\text{TWAS}}, \lambda_{\text{SNP}}, \boldsymbol{\Omega}, \mathbf{V}, n\sigma_c^2) \\ = \sum_{c' \in \mathcal{C}; c'_i=1} \Pr(c' | \mathbf{z}_{\text{TWAS}}, \lambda_{\text{SNP}}, \boldsymbol{\Omega}, \mathbf{V}, n\sigma_c^2) \end{aligned}$$

### 1.2 GWAS 数据

我们使用了来自英国生物银行关于焦虑障碍的 GWAS 数据<sup>[5]</sup>。对得到的共 11 832 122 个 SNP 进行清洗, 未发现插补后信息分数  $\leq 0.9$ , 次要等位基因频率  $\text{MAF} \leq 0.01$ ,  $p$  值越界或具有重复 rs 编号的 SNP, 删除 1 811 540 个不是 SNP 或链模糊的变体数据。

对清洗后的数据做统计分析, 通过最小二乘法获得各基因主要等位基因的效应值大小 BETA, 中值为  $1.590 14 \times 10^{-5}$ , 符合统计规律。平均卡方值为 1.415, 最大卡方值为 108.619。再用基因组控制方法来校正卡方统计, 得出  $\text{GC} - \lambda_{0.5} = 1.254$ ,  $\text{GC} - \lambda_{0.1} = 1.361$ ,

$GC - \lambda_{0.01} = 1.605$ ,  $GC - \lambda_{0.001} = 2.692$ 。其  $\lambda$  值相对较高, 表现出了较高的膨胀值与假阳性。

最后将清洗后数据的染色体序号、SNP 的 rs 编号、效应等位基因、备选等位基因、碱基对、SNP 系数 (BETA) 及  $p$  值做标注并生成适用于 FOCUS 的 GWAS 摘要文件。此文件将用于后续的 TWAS 精细映射分析。

### 1.3 eQTL 数据

为了尽量增加 eQTL 的样本量, 并且可以与 SMR 方法使用相同的参考面板, 本研究将来自 PrediXcan 模型的最新脑 GTEX\_v8<sup>[6]</sup>权重与来自 TWAS FUSION 软件的 METSIM、NTR、YFS 和 CMC 模型的脑区权重组合到一个用于 FOCUS 的单一可用数据库中。由于来自 PrediXcan 模型<sup>[7]</sup>的脑 GTEX\_v8 权重使用 GRCh38 (hg38) 作为参考基因组版本, 而本研究的 GWAS 数据及后续用于 SMR 的 eQTL 数据均采用 GRCh37 (hg19) 作为标注, 故使用 LiftOver 软件完成了将脑 GTEX\_v8 的 GRCh38 (hg38) 向 GRCh37 (hg19) 的基因组坐标的转换, 并且将来自 1 000 Genomes

Project Phase3<sup>[8]</sup>的 LD 信息作为参考。

## 2 结果

结合 FOCUS 的结果, 对得到的 GWAS SNP 进行分析, 得到了来自 13 个基因的 18 个 SNP, 如表 1 所示。

如图 1 所示, 根据 PIP 值对得到的 13 个基因进行因果排序, 可以看出明显的分层现象, FOCUS 为 FAM86B3P、TDH、FLOT1、MSRA、RP11-753C18.8、FAM167A、EIF3KP1 及 PCCB 基因分配了较高的后验概率。

其中, FAM167A 与 EIF3KP1 组织中有多个 SNP 且 PIP 值较高, 为最可能导致焦虑障碍的因果基因。

对 FAM167A 与 EIF3KP1 所在的基因座进行优先排序, 得出如下结果。

图 2 显示的是 8 号染色体上的 FAM167A 基因座与 5 号染色体上的 EIF3KP1 基因座的结果。在图 2 中, 灰色点表示研究采用的是最新 GWAS 分析的 SNP 的  $p$  值, 菱形表示来自 SMR 测试的探针的  $p$  值。

表 1 FOCUS 结果  
Tab.1 FOCUS Results

探针 ID	基因名称	组织	最相关 SNP
ENSG00000244731	C4A	小脑半球	rs1265905
ENSG00000244731	C4A	小脑半球	rs2647019
ENSG00000185294	SPPL2C	小脑	rs17573607
ENSG00000185294	SPPL2C	小脑	rs2243967
ENSG00000175749	EIF3KP1	小脑半球	rs26431
ENSG00000175749	EIF3KP1	小脑半球	rs34809
ENSG00000175749	EIF3KP1	小脑半球	rs158402
ENSG00000154319	FAM167A	背外侧前额叶皮层	rs4841466
ENSG00000154319	FAM167A	背外侧前额叶皮层	rs11782823
ENSG00000173295	FAM86B3P	小脑半球	rs2980441
ENSG00000235266	RP11-753C18.8	小脑半球	rs10883816
ENSG00000175806	MSRA	脊髓颈椎 c1 段	rs7833387
ENSG00000154316	TDH	脊髓颈椎 c1 段	rs12682349
ENSG00000251402	FAM90A25P	尾状基地神经节	rs2618443
ENSG00000080854	IGSF9B	小脑	rs688500
ENSG00000114054	PCCB	背外侧前额叶皮层	rs66691851
ENSG00000106392	C1GALT1	尾状基地神经节	rs10246303
ENSG00000137312	FLOT1	小脑	rs12527415

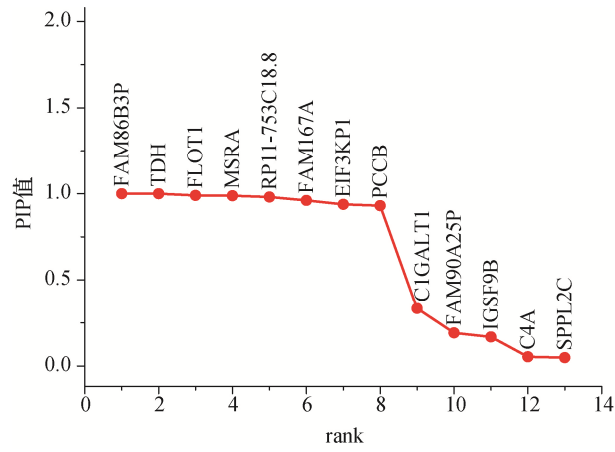


图 1 FOCUS 的 PIP 结果  
Fig.1 PIP results of FOCUS

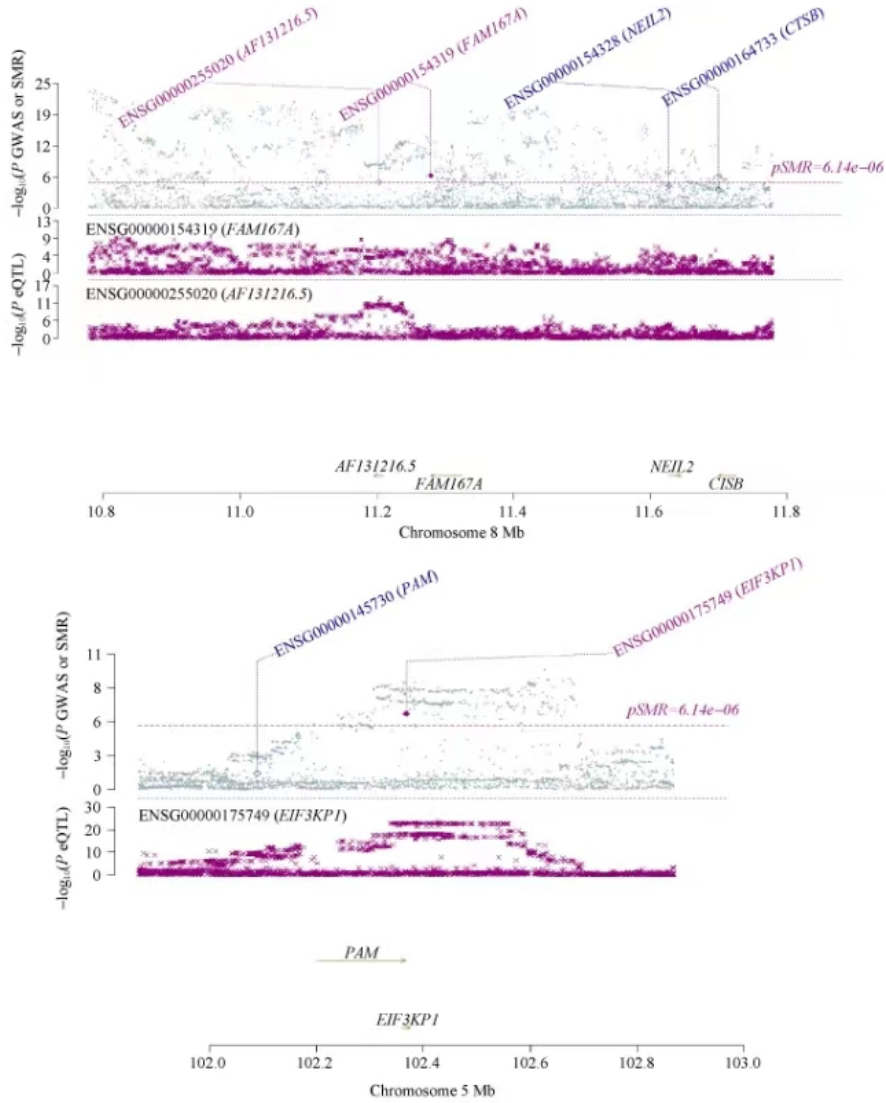


图 2 GWAS 与 eQTL 共定位情况  
Fig. 2 Co-localization of GWAS and eQTL signals



### 3 讨论

焦虑障碍是最普遍的精神障碍之一，其终生患病率可超过 20%。早在现代基因组学出现之前，科学家就通过家族谱系的研究提出基因是导致焦虑障碍的一个重要因素。迄今为止，大多数对焦虑障碍的关联研究仍集中在候选基因上，根据有限的先验假设选择基因，重点关注它们在染色体的连锁峰下的位置，然后在动物模型中确定可能导致焦虑特征的基因组区域，这样挑选的研究区域就会存在潜在的偏倚性。GWAS 方法可以无偏差地获得整个基因组中常见变异的关联性图谱，并且结合 eQTL 数据，将遗传变异与基因表达的变异联系起来，以查明基因表达的多态性遗传区域，以此解释基因与疾病之间复杂的关联性。本研究采用 FOCUS 算法，整合了来自英国生物银行关于焦虑障碍的二元性状的 GWAS 数据及来自 PredixCan 模型的最新脑 GTEx-v8 的 13 个脑区的 eQTL 数据，并使用了来自 fastGLMM 算法所整合的 GWAS 分析数据，其对本研究的二元性状展现出了良好的性能。同时运行 SMR 和 HEIDI 测试，结果报告了包括 FAM85B、RP11-62H7.2、FAM86B3P、RP11-481A20.10、FAM66A、ITIH4 等基因在内的风险基因，这为临床上焦虑障碍的治疗与药物的开发提供了潜在方向。本研究还发现 FOCUS 报告包括 SOX7、AREL1、FLOT1 在内的导致焦虑障碍的风险基因。这些基因同样为后续的研究提供了可能的方向。

其中，FLOT1 基因的编码 Flotillin-1，是一种定位于小窝的蛋白质，小窝是内细胞膜上的小区域。这种蛋白质在囊泡运输和细胞形态中发挥作用。两项研究将 FLOT1 基因及其编码 Flotillin-1 蛋白作为阿尔茨海默病<sup>[9]</sup>和重度抑郁症<sup>[10]</sup>的风险基因。目前 FLOT1 的功能尚未完全确定。然而，先前的研究表明，Flotillin-1 可能在中枢神经系统中起关键作用<sup>[11]</sup>。从果蝇到人类，FLOT1 的蛋白质序列高度保守<sup>[12]</sup>，表明了 FLOT1 的重要性。据报道，FLOT1 参与神经元分化、轴突生长和再生，并被证明在果蝇大脑中大量表达<sup>[12]</sup>。这些证据证明了 FLOT1 在中枢神经系统中的重要性。除了多巴胺转运蛋白和谷氨酸转运蛋白，一项最近的研究表示，Flotillin-1 通过与血清素转运蛋白相互作用调节血清素能神经传递<sup>[13]</sup>。考虑到 5-羟色胺转运蛋白、多巴胺转运蛋白和谷氨酸转运蛋白在重度抑郁症中的重要

作用，FLOT1 可能通过调节多巴胺转运蛋白和谷氨酸转运蛋白 EAAT2 的内化参与焦虑障碍。但需要更多的工作来研究 FLOT1 在焦虑障碍发病机制中的确切作用。

#### 参考文献

- [1] RICHARD J M. Anxiety and its disorders: the nature and treatment of anxiety and panic[J]. **2nd ed. Am. J. Psychiatry**, 2002, 159(8): 1453-1453.
- [2] MARKS I M. Genetics of fear and anxiety disorders[J]. **Br. J. Psychiatry**, 1986, 149(Oct. ): 406-418.
- [3] LESCH K P, BENGEL D, HEILS A, *et al.* Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region[J]. **Science**, 1996, 274(5292): 1527-1531.
- [4] MANCUSO N, FREUND M K, JOHNSON R, *et al.* Probabilistic fine-mapping of transcriptome-wide association studies[J]. **Nat. Genet.**, 2019, 51(4): 675-682.
- [5] NAGEL M, JANSEN P R, STRINGER S, *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449, 484 individuals identifies novel genetic loci and pathways[J]. **Nat. Genet.**, 2018, 50(7): 920-927.
- [6] AGUET F, BARBEIRA A N, BONAZZOLA R, *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues[J]. **Science (80-. )**, 2020, 369(6509): 1318-1330.
- [7] BARBEIRA A, DICKINSON S P, BONAZZOLA R, *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics[J]. **Nat. Commun.**, 2018, 9(1): 1-20.
- [8] 1000 Genomes Project Consortium. A global reference for human genetic variation[J]. **Nature**, 2015, 526: 68-74.
- [9] MOHAMMAD A, NORIYUKI K, HIROYASU A, *et al.* Flotillin is a novel diagnostic blood marker of alzheimer's disease[J]. **J. Alzheimers. Dis.**, 2019, 72(4): 1165-1176.
- [10] ZHONG J M, LI S, ZENG W L, *et al.* Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder[J]. **Neuropsychopharmacology**, 2019, 44(9): 1542-1551.
- [11] CREMONA M L, MATTHIES H J G, PAU K, *et al.* Flotillin-1 is essential for PKC-triggered endocytosis and membrane microdomain localization of DAT[J]. **Nat. Neurosci.**, 2011, 14(4): 469-477.
- [12] GALBIATI F, VOLONTÉ D, GOLTZ J S, *et al.* Identification, sequence and developmental expression of invertebrate flotillins from *Drosophila melanogaster*[J]. **Gene**, 1998, 210(2): 229-237.
- [13] REISINGER S N, KONG E, MOLZ B. Flotillin-1 interacts with the serotonin transporter and modulates chronic corticosterone response[J]. **Genes. Brain. Behav.**, 2019, 18(2).