

·论著·

# 基于生物信息学分析 F13A1 基因及蛋白质

王 淮,杨健康

(大理大学基础医学院,云南 大理 671000)

**摘要:**目的 利用生物信息学方法对 F13A1 及其编码蛋白质进行分析,为 F13A1 的实验研究提供理论基础和新线索。方法 从 UCSC (<http://genome.ucsc.edu/index.html>) 数据库检索“F13A1”,下载 F13A1 蛋白序列,截取转录起始点上游 2000 bp 碱基序列作为启动子区域进行分析,分析 F13A1 编码蛋白的理化性质、亚细胞定位、信号肽和跨膜区域、蛋白质结构、相互作用网络及 GO 注释。结果 F13A1 包含多个启动区域,不存在 CpG 岛,但存在 SRF、Sp1 和 AP-1 等转录因子结合位点。F13A1 编码蛋白是由 732 个氨基酸组成的无信号肽、无跨膜区域的亲水蛋白,其主要定位于细胞核,二级结构以随机卷曲为主。另外蛋白 F13B、FGA、FGG、F2、FGB、FN1、SERPINF2、VWF、VEGFA、F5 可能与 F13A1 存在相互作用。结论 F13A1 基因存在多个启动子区域,拥有多个转录因子结合位点;F13A1 蛋白与 F13B、FGA、FGG、F2 等多个蛋白质相互作用,其参与组成内质网腔、血小板 α 颗粒内腔以及细胞内细胞器腔等多种细胞成分,同时在血液凝固、纤维蛋白溶解和血小板活化等多种生理过程发挥重要作用。

**关键词:** F13A1; 启动子; 蛋白质; 生物信息学

中图分类号: R394.3

文献标识码: A

DOI: 10.3969/j.issn.1006-1959.2020.11.017

文章编号: 1006-1959(2020)11-0052-06

## Analysis of F13A1 Gene and Protein Based on Bioinformatics

WANG Huai, YANG Jian-kang

(Basic Medical College, Dali University, Dali 671000, Yunnan, China)

**Abstract:** Objective To analyze F13A1 and its encoded protein using bioinformatics methods, and provide theoretical basis and new clues for the experimental study of F13A1. Methods Search "F13A1" from the UCSC (<http://genome.ucsc.edu/index.html>) database, download the F13A1 protein sequence, and intercept the 2000 bp base sequence upstream of the transcription start point as a promoter region for analysis; the physicochemical properties, subcellular localization, signal peptide and transmembrane region, protein structure, interaction network and GO annotation of F13A1 encoded protein were analyzed. Results F13A1 contains multiple promoter regions, there is no CpG island, but there are transcription factor binding sites such as SRF, Sp1 and AP-1. The protein encoded by F13A1 is a hydrophilic protein composed of 732 amino acids with no signal peptide and no transmembrane region. It is mainly located in the nucleus, and the secondary structure is dominated by random coils. In addition, proteins F13B, FGA, FGG, F2, FGB, FN1, SERPINF2, VWF, VEGFA, F5 may interact with F13A1. Conclusion The F13A1 gene has multiple promoter regions with multiple transcription factor binding sites; the F13A1 protein interacts with multiple proteins such as F13B, FGA, FGG, and F2, and is involved in the formation of the endoplasmic reticulum cavity, platelet α particle cavity, and cells. Various cellular components such as the inner organelle cavity play an important role in various physiological processes such as blood coagulation, fibrinolysis and platelet activation.

**Key words:** F13A1; Promoter; Protein; Bioinformatics

人凝血因子 X Ⅲ (Coagulation Factor X Ⅲ, FX Ⅲ) 又称纤维蛋白稳定因子,主要功能是共价交联纤维蛋白纤维,稳定纤维蛋白凝块<sup>[1]</sup>。F13A1 基因编码人凝血因子 X Ⅲ 的 A 链。遗传性凝血因子 X Ⅲ 纯合或复合杂合缺乏症是一种罕见的严重出血性疾病,2016 年 Thomas A 等<sup>[2]</sup>首次报道 1 例覆盖整个外显子 12 的新的纯合子 F13A1 缺失的病例,通过实验确定是由内含子 11 和内含子 12 的 6 bp 微同源序列引起。研究表明<sup>[3]</sup>,F13A1 具有上调血浆 IL-6 水平的功能,在痤疮发病过程中扮演重要角色;F13A1 基因变异与血浆 FX Ⅲ A 水平对急性心肌梗死的长期预后具有显著的作用<sup>[4]</sup>。本文通过生物信息学的方法对 F13A1 基因及其编码蛋白质进行一系列预测分析,期望能为今后的研究提供线索及方向。

### 1 材料与方法

#### 1.1 资料来源 F13A1 基因及编码蛋白的序列获取

基金项目:国家自然科学基金项目(编号:81760565)

作者简介:王淮(1995.8-),女,重庆人,硕士研究生,主要从事人体生理与病理生理学研究

通讯作者:杨健康(1984.1-),男,云南大理人,博士,助理研究员,主要从事分子生物学研究

登录 UCSC (<http://genome.ucsc.edu/index.html>) 数据库,在 GeneSorter 中检索“F13A1”,截取转录起始点上游 2000 bp<sup>[5]</sup>碱基序列作为启动子区域进行分析,同时下载 F13A1 蛋白序列。

1.2 方法 本文运用到的分析软件见表 1,所有软件都应用默认参数分析 F13A1 编码蛋白的理化性质、亚细胞定位、信号肽和跨膜区域、蛋白质结构、相互作用网络及 GO 注释。

### 2 结果

2.1 F13A1 基因特征 人 F13A1 基因总长为 176750 bp,在染色体上的位置是 chr6:6144085-6320834,包含 15 个外显子和 14 个内含子,该基因编码的蛋白质包含 732 个氨基酸残基。

2.2 F13A1 基因启动子分析 Promoter2.0 分析结果显示,F13A1 基因上游存在 3 个启动子区域,分别位于 200 bp,800 bp 和 1900 bp 处,均为临界性预测。NNPP 分析结果显示,F13A1 基因上游存在 5 个启动子区域,见表 2。推测 F13A1 基因至少包含 3 个启动子区域,其中以 1585~1635 bp,420~470 bp 和 1785~1635 bp 可能性最大。

表 1 分析软件

功能	软件	网址
预测启动子	Neural Network Promoter Prediction	<a href="https://www.fruitfly.org/seq_tools/promoter.html">https://www.fruitfly.org/seq_tools/promoter.html</a>
	Promoter2.0	<a href="http://www.cbs.dtu.dk/services/Promoter/">http://www.cbs.dtu.dk/services/Promoter/</a>
预测 CpG 岛	Cpgplot	<a href="https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/">https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/</a>
	MethPrime	<a href="http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi">http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi</a>
预测转录因子结合位点	AliBaba2.1	<a href="http://gene-regulation.com/pub/programs/alibaba2/index.html">http://gene-regulation.com/pub/programs/alibaba2/index.html</a>
	Cister	<a href="https://zlab.bu.edu/~mfrith/cister.shtml">https://zlab.bu.edu/~mfrith/cister.shtml</a>
蛋白质理化性质	ProtParam	<a href="https://web.expasy.org/protparam/">https://web.expasy.org/protparam/</a>
亲疏水性分析	ProtScale	<a href="https://web.expasy.org/protscale/">https://web.expasy.org/protscale/</a>
亚细胞定位	PSORTII	<a href="https://www.genscript.com/psort.html">https://www.genscript.com/psort.html</a>
信号肽预测	SignalP4.1	<a href="http://www.cbs.dtu.dk/services/SignalP-4.1/">http://www.cbs.dtu.dk/services/SignalP-4.1/</a>
跨膜区域分析	TMHMM	<a href="http://www.detaibio.com/tools/transmembrane.html">http://www.detaibio.com/tools/transmembrane.html</a>
二级结构预测	SOPMA	<a href="https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html">https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html</a>
三级结构预测	SWISS-MODEL	<a href="https://swissmodel.expasy.org/">https://swissmodel.expasy.org/</a>
蛋白质拉曼图分析	SAVES v5.0	<a href="https://servicesn.mbi.ucla.edu/SAVES/">https://servicesn.mbi.ucla.edu/SAVES/</a>
蛋白相互作用	STRING	<a href="https://string-db.org/">https://string-db.org/</a>

表 2 NNPP 分析 F13A1 基因启动子区域结果

起始位点	结束位点	得分	序列
21	71	0.90	cctgaaattgtatgaatgctgaggggaagcagtgctagacatcttaggg
420	470	0.96	gaagggtttttaaaaatactggagacctcaaattgtatgtagctgaggg
1297	1347	0.80	gggagggtttttaaaaatactgaagagacctgaaattgcatgaatgctg
1585	1635	0.98	tttcttcttataacattgggagattgagtgccagggcagaaatttga
1785	1635	0.94	actaaaaatacaaaaattagcgggcatggtggcatgcacctgtagtcc

2.3 F13A1 基因 CpG 岛预测 在观察值/预期值>0.6, G%+C%>50%, 长度>100 bp 条件下, 运用在线软件 Cpgplot 和 MethPrime 对 F13A1 基因进行 CpG 岛预测, 均未发现甲基化岛, 推测 F13A1 基因启动子区域不包含 CpG 岛, 见图 1、图 2。

2.4 F13A1 基因转录因子结合位点预测 AliBaba2.1

的预测结果共有 255 个, 主要包含 Sp1、Ap-1、SRF、NF-1、USF 等。Cister 预测 F13A1 基因共有 27 个转录因子结合位点, 表 3 为概率最大的前 5 个。推测 F13A1 基因转录因子结合位点至少包含 SRF、Sp1 和 AP-1, 此为两个软件的共同预测结果, 相对可靠。

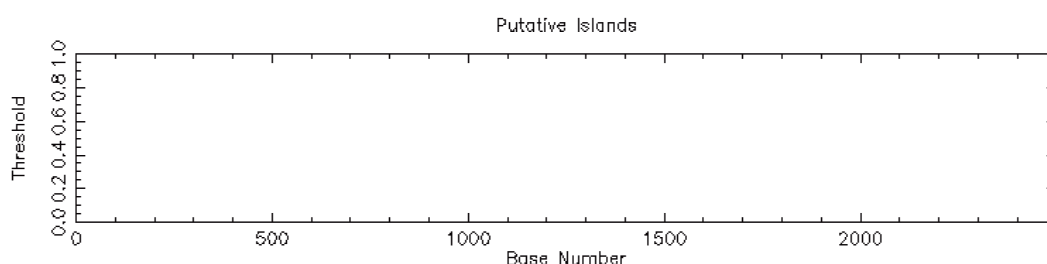


图 1 Cpgplot 预测 F13A1 基因启动子区的 CpG 岛结果

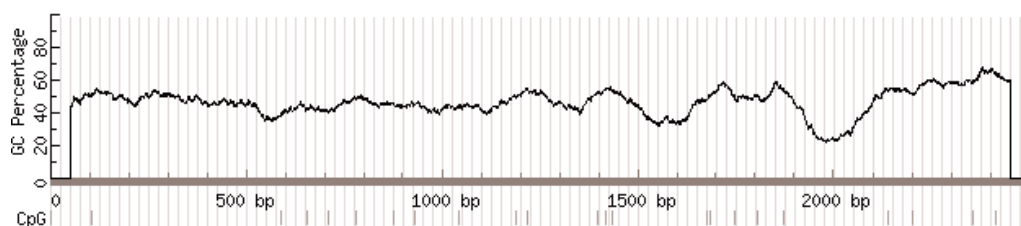


图 2 MethPrime 预测 F13A1 基因启动子区的 CpG 岛结果

表 3 Cister 预测 F13A1 基因转录因子结合位点结果

因子类型	结合位点	正负链	结合序列	可能性
ERE	2233-2246	-	tgtctgtgtgacct	0.84
SRF	2180-2192	-	gccttatttgacac	0.73
Sp1	2264-2276	-	ccccctcctgcc	0.48
Ets	2366-2376	-	gacttcctggc	0.47
Myf	2283-2294	+	cagtagcagctg	0.41

**2.5 F13A1 蛋白质的理化性质分析** ProtParam 分析结果显示,人 F13A1 蛋白质共有 732 个氨基酸残基,分子质量为 83267.30 Da,理论等电点值为 5.75,属酸性蛋白质。分子式为 C<sub>3711</sub>H<sub>5743</sub>N<sub>1015</sub>O<sub>1110</sub>S<sub>29</sub>,原子总数为 11608。在 732 个氨基酸残基中带负电荷的氨基酸残基(Asp+Glu)的数量为 95,带正电荷的氨基酸残基(Arg+Lys)的数量为 83。F13A1 蛋白质的不稳定系数为 37.30,属于稳定蛋白质,脂肪系数为 79.00,总的平均亲水性为 -0.391,属于亲水性蛋白。ProtScale 分析结果显示,位于 201 位的谷氨酸(E)亲水性最强,分值为 -3.411;位于 631 位的脯氨酸(P)疏水性最强,分值为 2.244。可知 F13A1 蛋白的疏水区域少于亲水区域,属于亲水蛋白,该结果与 ProtParam 结果一致,可信性高,见图 3。

**2.6 F13A1 蛋白的亚细胞定位、信号肽和跨膜结构域预测** PSORTII 对 F13A1 蛋白的亚细胞定位的预

测结果显示,位于细胞核的可能性为 34.8%,位于线粒体的可能性为 26.1%,位于细胞质的可能性为 26.1%,位于分泌系统囊泡的可能性为 13.0%。SignalP4.1 的预测结果显示,C、Y、S 的最大值分别为 0.109、0.108、0.125,S-mean 的值为 0.107,D 值为 0.108,可知 F13A1 蛋白不存在剪切位点,不是分泌蛋白,见图 4。TMHMM 预测 F13A1 蛋白的跨膜区域结果显示,该蛋白不存在跨膜区域,不属于膜蛋白,见图 5。

**2.7 F13A1 蛋白二级结构与三级结构预测** SMOPA 预测 F13A1 蛋白质的二级结构结果显示,随机卷曲共有 358 个,占 48.91%;延伸链共有 204 个,占 27.87%;α 螺旋共有 131 个,占 17.90%;β 转角有 39 个,占 5.33%。其中大部分氨基酸处于随机卷曲状态,F13A1 蛋白整体处于比较松散的状态,见图 6。

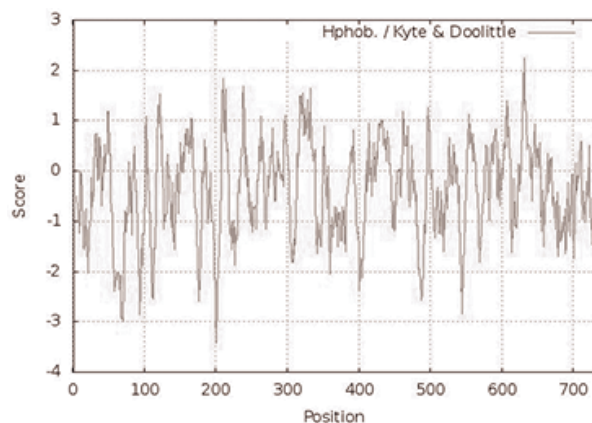


图 3 ProtScale 分析 F13A1 蛋白亲疏水性结果

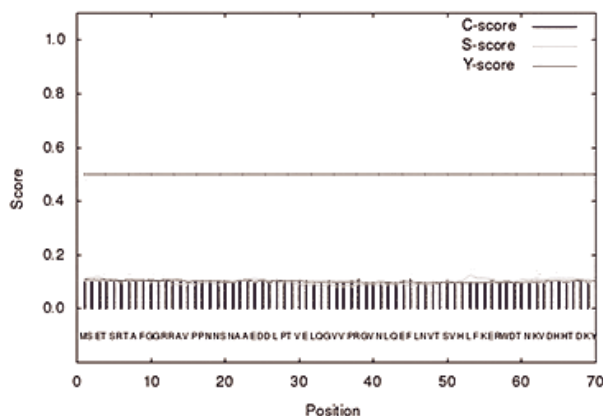


图 4 SignalP4.1 预测 F13A1 信号肽结果

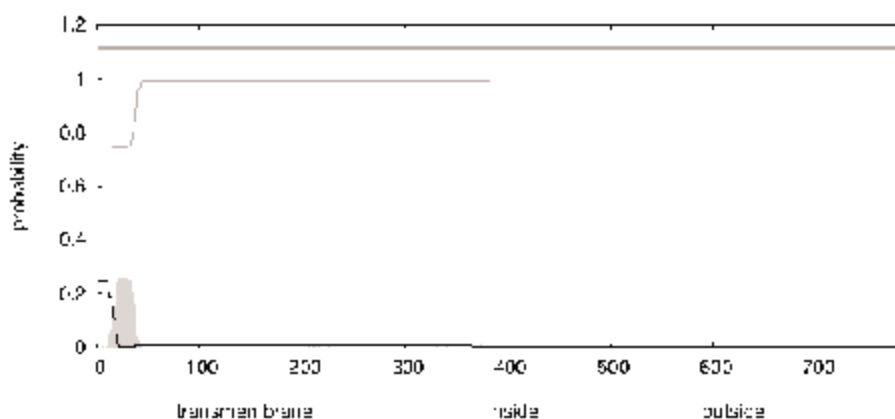


图 5 预测 F13A1 跨膜区域结果

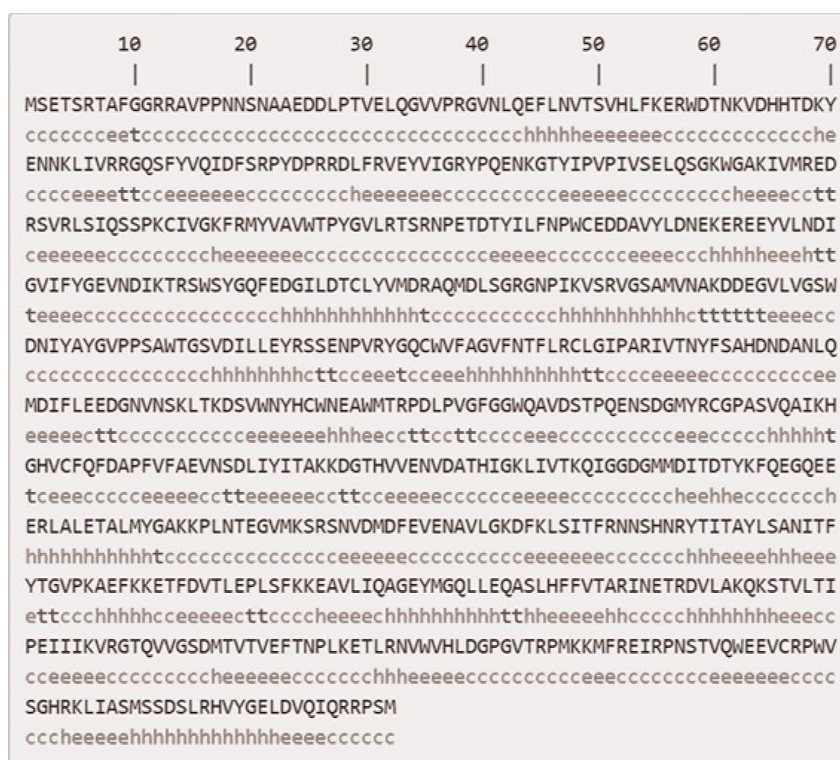


图 6 F13A1 蛋白二级结构预测结果

蛋白质的高级结构决定其生理功能,采用同源建模法预测蛋白质的三级结构,将 F13A1 蛋白序列提交至 SWISS-MODEL,得到其高级结构见图 7。其中预测模型 A 采用的模板为 1f13.1.A,覆盖率为 100%,序列相似度为 63%;预模型 B 采用的模板为 3ohx.1.B,覆盖率为 15%,序列相似度为 26%;预测模型 C 采用的模板为 2ice.1.aA,覆盖率为 9%,序列相似度为 30%。综合比较,预测模型 A 的可信度最高

为验证模型 A 的可信度,通过蛋白质拉曼图在线分析软件 SAVES v5.0 对预测模型的氨基酸的二面角进行分析,判断预测模型的可靠性,得到结果见图 8。预测结构中所涉及的 1279 个氨基酸,其中 1149 个处于最佳区域,即红色区域;126 个处于允许区域,即黄色区域;另有 4 个处于允许区域范围之

外,说明该预测模型中几乎所有的氨基酸均形成合理的二面角,构成的蛋白质高级结构稳定可信。

**2.8 F13A1 蛋白相互作用网络分析** F13A1 蛋白质的潜在功能 通过 STRING 预测 F13A1 蛋白的相互作用网络,结果显示,F13B、FGA、FGG、F2、FGB、FN1、SERPINF2、VWF、VEGFA、F5 共 10 个蛋白质与 F13A1 蛋白相互作用,分值均大于 0.9,见图 9。对以上 10 个蛋白质编码基因进行 GO 注释发现,这些蛋白质主要参与血液凝固、纤维蛋白凝块形成和纤维蛋白溶解生理过程,见图 10。同时发现,这些蛋白质具有与信号受体、蛋白酶、肝素以及细胞粘附分子等结合的功能。提示 F13A1 蛋白可通过与其他蛋白质相互作用从而参与凝血以及伤口愈合等生理过程。



图 7 F13A1 蛋白三级结构预测结果

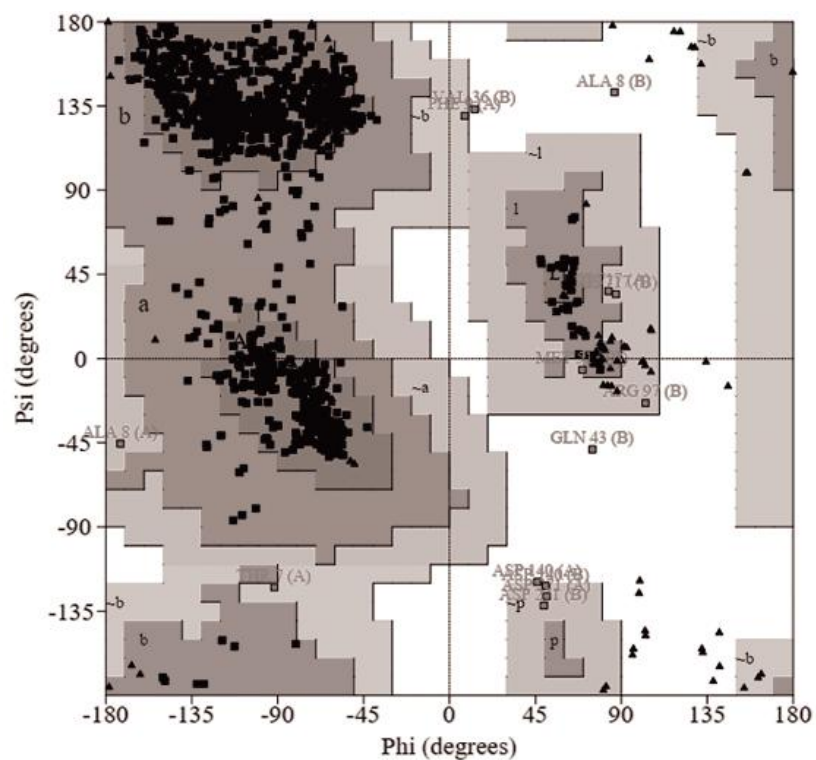


图 8 F13A1 蛋白质预测三级结构模型的拉曼图分析

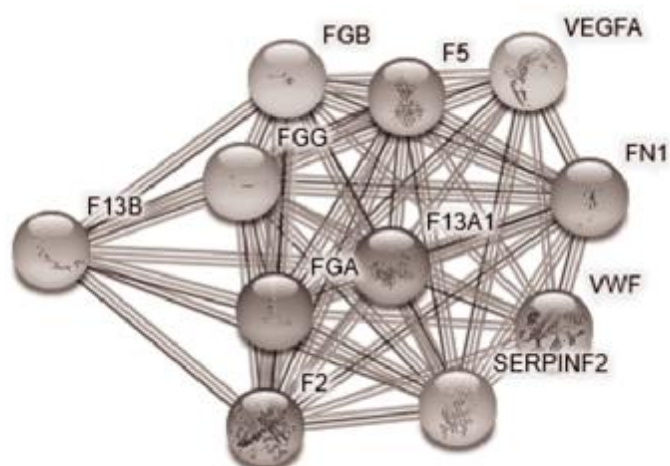
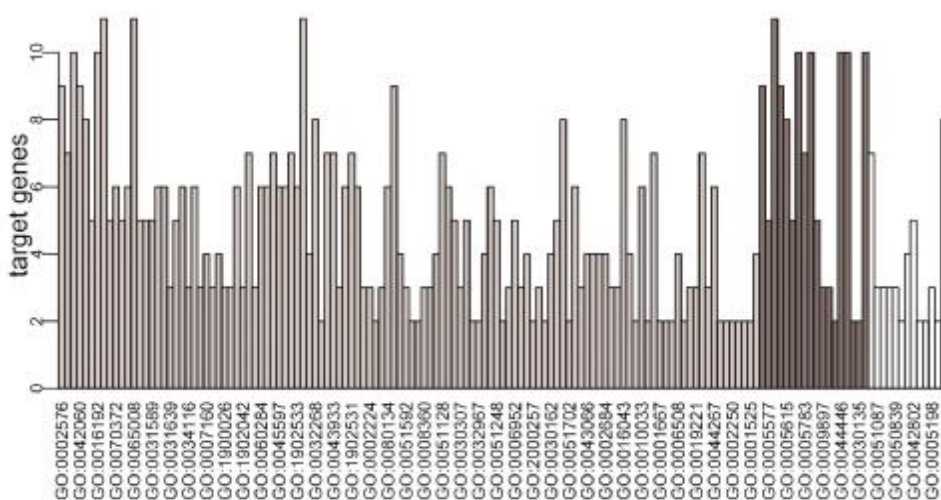


图 9 F13A1 蛋白相互作用网络





注:灰色:生物过程;黑色:细胞成分;白色:分子功能

图 10 与 F13A1 相互作用蛋白质的 GO 注释结果

### 3 讨论

F13A1 蛋白作为转谷氨酰胺超家族的一员,参与凝血、伤口愈合、血管生成<sup>[6]</sup>和血小板脱粒等众多生理过程,其功能缺陷或编码基因突变将会导致一系列疾病的发生。比如 F13A1 基因变异导致的习惯性流产<sup>[7]</sup>、先天性凝血因子 XIII 缺乏症<sup>[8]</sup>,有研究表明检测 F13A1 在血液中的表达能够辅助判断口腔鳞状细胞癌的发生及转移<sup>[9]</sup>,也可作为腹腔粘连或肠梗阻的预后指标<sup>[10]</sup>。目前的研究主要针对 F13A1 的表达水平、参与疾病发生与生理过程,系统分析 F13A1 的启动子及其编码蛋白等相关信息有助于深入研究其在正常生理过程和疾病发生发展中起到的作用。

本文先对 F13A1 基因进行分析,发现该基因存在多个启动子区域,其中以 1585-1635bp、420-470bp 和 1785-1635bp 可能性最大,但是启动子区中未发现 CpG 岛的存在。另外,该基因拥有多个转录因子结合位点,其中以 SRF、Sp1 和 AP-1 最为常见。对 F13A1 蛋白质进行分析,发现 F13A1 蛋白是由 732 个氨基酸组成的亲水蛋白,等电点为 5.75。F13A1 二级结构以随机卷曲为主。本文还对 F13A1 的分子功能进行分析,发现 10 个与其相互作用的蛋白质。通过 GO 注释发现,F13A1 蛋白参与组成内质网腔、血小板 α 颗粒内腔以及细胞内细胞器腔等多种细胞成分,同时在血液凝固、纤维蛋白溶解和血小板活化等多种生理过程发挥重要作用。

综上所述,本文用生物信息学方法分析了 F13A1 基因的启动子、CpG 岛及转录因子结合位点情况,以及相应蛋白质的理化性质、亚细胞定位、信号肽、跨膜区域、高级结构和分子功能等情况,旨在为后续实验研究 F13A1 基因功能及表达调控提供理论依据和重要线索。

### 参考文献:

[1]Ma SY,Chen CM,Liang Q,et al.Phenotype and genotype of

FXIII deficiency in two unrelated probands: identification of a novel F13A1 large deletion mediated by complex rearrangement [J].Orphanet Journal of Rare Diseases,2019,14(1):182.

[2]Thomas A,Ivaskevicius V,Zawadzki C,et al.Characterization of a novel large deletion caused by double-stranded breaks in 6-bp microhomologous sequences of intron 11 and 12 of the F13A1 gene[J].Human Genome Variation,2016(3):15059.

[3]Ya XY,Wu WJ,Peng MS,et al.Identity-by-Descent analysis reveals susceptibility loci for severe acne in Chinese Han cohort [J].The Journal of Investigative Dermatology,2019,139(9):2049-2051.

[4]Anasani L,Marchesini J,Pestelli G,et al.F13A1 gene variant (V34L) and residual circulating FXIII levels predict short- and long-term mortality in acute myocardial infarction after coronary angioplasty [J].International Journal of Molecular Sciences, 2018,19(9):2766.

[5]鲍思元,金科华,陈红霞,等.人 IDE 基因启动子生物信息学分析[J].华中师范大学学报(自然科学版),2018,52(4):525-531.

[6]Mitchell JL,Mutch NJ.Let's cross-link: diverse functions of the promiscuous cellular transglutaminase factor XIII-A[J].Journal of Thrombosis and Haemostasis,2019,17(1):19-30.

[7]Xu ZY,Zhang Y,Liu W,et al.Polymorphisms of F2, PROC, PROZ, and F13A1 genes are associated with recurrent spontaneous abortion in Chinese Han women [J].Clinical and Applied Thrombosis Hemostasis,2018,24(6):894-900.

[8]Ivaskevicius V,Biswas A,Garly ML,et al.Comparison of F13A1 gene mutations in 73 patients treated with recombinant FXIII-A2[J].Haemophilia,2017,23(3):e194-e203.

[9]刘俊君.F13A1 在口腔鳞状细胞癌中表达及功能的初步研究[D].广西医科大学,2019.

[10]Gemmati D, Occhionorelli S, Tisato V, et al. Inherited genetic predispositions in F13A1 and F13B genes predict abdominal adhesion formation: identification of gender prognostic indicators [J].Scientific Reports,2018,8(1):16916.

收稿日期:2020-03-26;修回日期:2020-04-06

编辑/成森