

# hRFT2 基因单核苷酸多态性的生物信息学分析

谢云飞<sup>1</sup> 解博红<sup>2</sup> 杨子善<sup>1</sup>

(<sup>1</sup> 新乡医学院生命科学技术系 河南省遗传性疾病与分子靶向药物高校重点实验室培育基地, 新乡 453003 ;

<sup>2</sup> 新乡医学院基础医学院, 新乡 453003 )

**摘要 :** 旨在筛选可能与人类疾病有关的 hRFT2 基因单核苷酸多态性 (single nucleotide polymorphisms, nsSNPs) 和突变位点, 从 SNP 数据库中检索并筛选出 395 个有效的 hRFT2 基因 SNPs, 其中包括 30 个同义 SNPs (synonymous SNPs, sSNPs) 和 31 个非同义 SNPs (non-synonymous single nucleotide polymorphisms, nsSNPs)。分别采用 SIFT、SNPs3D 和 PolyPhen-2 方法分析 nsSNPs 引起的氨基酸替换是否可能影响 hRFT2 的功能。结果表明, 5 个 nsSNPs (rs11477762、rs146302587、rs146492942、rs76947760 和 rs145431028) 可能严重影响 hRFT2 蛋白的功能, 其中 rs76947760 和 rs145431028 的影响已得到临床证明, 另外 3 个 nsSNPs (rs148387972、rs140391358 和 rs3746802) 也可能对 hRFT2 有较大的影响。

**关键词 :** 人类核黄素转运蛋白 2 单核苷酸多态性 氨基酸替换

## Bioinformatics Analysis of Single Nucleotide Polymorphisms in hRFT2 Gene

Xie Yunfei<sup>1</sup> Xie Bohong<sup>2</sup> Yang Zishan<sup>1</sup>

(<sup>1</sup>Department of Life Sciences and Technology, Henan Key Laboratory of Hereditary Diseases and Molecular Targeted Drugs (Cultivating Base), Xinxiang Medical University, Xinxiang 453003 ;<sup>2</sup>School of Basic Medical Sciences, Xinxiang Medical University, Xinxiang 453003 )

**Abstract:** It was to screen disease-associated single nucleotide polymorphisms (SNPs) of hRFT2 gene, 395 useful SNPs were retrieved and screened from SNP databases, including 30 synonymous SNPs (sSNPs) and 31 non-synonymous SNPs (nsSNPs). The possible effects of nsSNPs-caused amino acid substitutions (AASs) were predicted using SIFT, SNPs3D and PolyPhen-2 methods. 5 nsSNPs (rs11477762, rs146302587, rs146492942, rs76947760 and rs145431028) were extremely damaging to hRFT2, in which rs76947760 and rs145431028 were proved by clinical studies. Another 3 nsSNPs (rs148387972, rs140391358 and rs3746802) might also be intolerable to hRFT2.

**Key words:** Human riboflavin transporter 2 Single nucleotide polymorphism Amino acid substitution

单核苷酸多态性 (single nucleotide polymorphism, SNP) 是基因组中广泛存在的可遗传性变异。人类 SNPs 的总量超过  $3 \times 10^6$  个, 其中仅有一小部分位于基因的编码区或表达调控区, 而能够引起细胞功能性改变的 SNPs 更是少之又少, 这部分 SNPs 与人类表型的差异、对疾病和药物的敏感性关系较为密切, 也是长期以来 SNP 研究的热点。近年来随着 SNP 检测技术的改进, 特别是人类基因组变异组计划 (Human Variome Project, HVP) 和全基因组关联

研究 (Genome-Wide Association Studies, GWAS) 的开展, SNP 与疾病关系的研究取得重大突破, 发现了多种重大疾病的易感基因和突变位点。

人类核黄素转运蛋白 2 (human riboflavin transporter 2, hRFT2) 基因位于人类 20 号染色体的短臂 (20p13), 故又被称为第 20 号染色体开放读码框 54 (chromosome 20 open reading frame 54, C20orf54)。hRFT2 基因含有 5 个外显子, 其中编码区位于第 2 至第 5 外显子, 编码的前体蛋白由 469 个氨基酸

收稿日期: 2011-11-03

基金项目: 河南省自然科学基金项目 (102300410139)

作者简介: 谢云飞, 男, 博士, 讲师, 研究方向: 肿瘤易感基因表达调控; E-mail: xieyunfei@xxmu.edu.cn

(amino acids, aa) 组成, N-末端的 22 aa 为信号肽。hRFT2 蛋白是一种跨膜蛋白, 含有 11 个跨膜区, 具有转运核黄素 (维生素 B<sub>2</sub>) 的功能<sup>[1,2]</sup>。hRFT2 蛋白在进化上高度保守, 与斑马鱼、鸡和大鼠来源的对应蛋白的同源性分别达到 36%、59% 和 83%<sup>[3]</sup>。

尽管对 hRFT2 基因的研究刚刚开始, 其功能也可能仅仅揭开冰山一角, 但由于发现 hRFT2 基因突变是引起进行性桥延麻痹伴耳聋 (Brown-Vialetto-Van Laere, BVVL) 综合征和法齐奥 - 隆德 (Fazio-Londe) 综合征的直接病因<sup>[3-5]</sup>, 同时有报道称其多态性与食管鳞状细胞癌 (esophageal squamous cell carcinoma, ESCC) 和贲门癌 (gastric cardia adenocarcinoma, GCA) 存在密切的联系<sup>[6]</sup>。hRFT2 已经引起科研人员的极大关注。已经发现的 hRFT2 SNPs 还有很多, 这些 SNPs 位点中, 必然有少数会影响 hRFT2 基因的表达或 hRFT2 蛋白的功能, 甚至和某些重大疾病存在密切联系。

非同义 SNPs (non-synonymous SNPs, nsSNPs) 可引起蛋白质氨基酸替换 (amino acid substitutions, AASs), 大部分 AASs 是蛋白质可以耐受的, 不会对蛋白质功能造成影响。少数 AASs 则可能改变蛋白质的表达水平、稳定性、定位或生物活性, 使其功能受损, 这些 AASs 往往位于蛋白的保守区。在已经揭秘的人类遗传性疾病中, 50% 以上与 AASs 有关。因此, 研究 nsSNPs 导致的 AASs 是否会影响蛋白质的功能十分重要。但 nsSNPs 数量众多, 通过逐一研究其对基因的影响显然是不现实的, 理想的途径是通过生物信息学方法从大量候选者中筛选出具有生物学意义的 nsSNPs, 再通过试验进一步研究。近十年来, 国内外已经建立了多种分析 nsSNPs 影响的生物信息学方法, 其原理不尽相同, 有些方法侧重于分析蛋白质序列的同源性<sup>[7,8]</sup>, 另一些侧重于分析蛋白质的结构<sup>[9,10]</sup>, 还有一些则二者并重<sup>[11,12]</sup>。SIFT、PolyPhen-2 和 SNPs3D 是其中 3 种具有代表性并提供网络服务平台的分析方法。SIFT 基于序列同源性比较, 通过对一些已知可导致疾病或蛋白功能受损的 nsSNPs 的分析表明, 其假阴性率和假阳性率分别为 31% 和 20%<sup>[7,13]</sup>。SNPs3D 是一种完全基于蛋白质结构分析的预测方法, 其假阴性率和假阳性率分别为 26% 和 15%<sup>[14,15]</sup>。PolyPhen-2 则同时分析

蛋白质的序列同源性和结构, 是 PolyPhen 方法的升级版, 根据分析对象和目的的不同, 又分为 HumDiv 和 HumVar 两种方法, 前者适用于复杂表型相关的稀有等位基因位点的评价和自然选择序列数据的分析等, 后者适用于将少数有极大影响的突变同众多轻度有害的突变区分开来, 以利于孟德尔疾病 (单基因疾病) 的诊断, 两种方法的假阳性率为 20%, 甄别已知功能性突变位点的阳性率分别为 92% 和 73%<sup>[16]</sup>。本研究从几个主要的人类 SNP 数据库中收集了 hRFT2 基因的 SNPs, 确定这些 SNPs 位点在 hRFT2 基因中的位置, 从中剔除假 hRFT2 SNPs, 进而采用 3 种 AASs 预测方法分析 nsSNPs 对 hRFT2 蛋白功能的影响, 以期筛选出可能和疾病有关的 nsSNPs, 为后续研究提供线索。

## 1 材料与方法

### 1.1 hRFT2基因SNPs的搜集

以 hRFT2 基因序列号 (NG\_027687.1) \ hRFT2 或 C20orf54 为关键词从 3 个 SNP 数据库中检索 SNPs, 包括美国 NCBI SNP 数据库 (<http://www.ncbi.nlm.nih.gov/snp>), 欧洲 SNP 数据库 (<http://www.gwascentral.org/index>), 以及日本 SNP 数据库 (<http://snp.ims.u-tokyo.ac.jp/>)。

### 1.2 SNP的确认和定位

以 hRFT2 基因 (NG\_027687.1) 和 mRNA (NM\_033409) 序列为标准, 根据 NCBI SNP 数据库提供的 SNPs 信息确定各 SNP 在 hRFT2 基因或 mRNA 中的位置。对于 hRFT2 基因中不存在的 SNPs 位点, 通过 NCBI BLAST 检索其在人类基因组中的位置。应用蛋白质跨膜区预测软件 TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>) 分析 hRFT2 蛋白的跨膜区。

### 1.3 分析方法

应用 3 种 AASs 效应预测方法 SIFT (<http://sift.bii.a-star.edu.sg/>) \ PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2>) 和 SNPs3D (<http://www.snps3d.org>) 分析各 nsSNPs 对 hRFT2 蛋白功能的影响。

## 2 结果

### 2.1 hRFT2基因SNPs分类

从 3 个 SNP 数据库中检索到 hRFT2 基因 SNPs 566 个。NCBI SNP 数据库是目前最权威的 SNP

数据库, 根据 hRFT2 基因序列号共检索到 507 个 SNPs, 其中有 112 个属重复提交。欧洲 SNP 数据库收录 hRFT2 基因 SNPs 37 个, 其中 21 个 SNPs 位点与 NCBI 数据库重复。日本 SNP 数据库收录的 22 个 SNPs 已全部收录入 NCBI SNP 数据库。剔除全部 155 个重复提交的 SNPs, 共得到 411 个唯一性 SNPs 位点。经过序列比较发现, 来自欧洲数据库的 16 个唯一性 SNPs 并非定位于 hRFT2 基因中, 而是介于 C20orf55 和 hRFT2 基因之间, 属于基因间 SNPs (intergenic SNPs, iSNP)。剩余 395 个唯一 SNPs 则全部定位于 hRFT2 基因中, 这些位点可视为有效 hRFT2 SNPs (表 1)。

表 1 不同数据库中 hRFT2 基因 SNPs 确认

SNP 数据库	唯一性 SNP		重复提交 SNP	合计
	有效 SNP	无效 SNP		
NCBI	395	0	112	507
欧洲	0	16	21	37
日本	0	0	22	22
合计	411		155	566

以 NCBI 数据库为参照

有效 SNPs 散布在 hRFT2 基因的不同区域, 其中非转录区分布最多, 共有 172 个 SNPs, 占总数的 43.5%。内含子区 SNPs 共 127 个, 占 32.2%。其余 96 个 SNPs 则位于外显子区, 包括 5' 非翻译区 (untranslated region, UTR)、编码区、3'UTR, 数目仅占总数的 24.3% (表 2)。

表 2 有效 SNPs 在 hRFT2 基因中定位

定位	数目
5' 非转录区	111
5'UTR	19
编码区	61
内含子区	127
3'UTR	16
3' 非转录区	61
合计	395

hRFT2 基因中对应于 mRNA 和内含子以外的 DNA 序列全部定义为非转录区

2.2 nsSNPs对hRFT2基因功能的影响

在 395 个有效 SNPs 位点中, 有 61 个位于基因编码区, 包括 30 个 sSNP 和 31 个 nsSNPs, 后者包含 1 个无义 SNP (nonsense SNP)。无义 SNP

(rs11477762) 导致翻译提前终止, 所编码的多肽链由野生型 469 aa 缩短为 123 aa, 这对 hRFT2 功能的影响必然是巨大的, 甚至可能是致死的。

采用 SIFT、PolyPhen-2 和 SNPs3D 三种方法分别预测 30 个 nsSNPs 导致的 AAS 对 hRFT2 蛋白有无影响, 仅有 12 个 nsSNPs 的分析结果完全一致, 16 个 nsSNPs 的预测结果不完全相同, 还有 2 个 nsSNPs 采用 SIFT 和 PolyPhen-2 预测的结果一致, 但不能使用 SNPs3D 方法进行分析。分析结果 (表 3) 表明, rs146302587、rs146492942 和 rs145431028 对 hRFT2 的功能影响很大, 这 3 个 nsSNPs 可分别引起 Gly13Arg、Leu150Phe 和 Phe457Leu 突变, 3 个氨基酸残基分别位于 hRFT2 前体蛋白的第 1、第 5 跨膜区和羧基端膜外区 (图 1)。16 个 nsSNPs 可能对 hRFT2 蛋白的功能有一定影响, 其中 rs148387972、rs140391358、rs3746802 和 rs76947760 的可能性更大, 因为每个位点都有两种方法给出阳性结果, 而且唯一的阴性结果也十分接近相应方法的临界点。剩余 11 个 nsSNPs 则是 hRFT2 可以耐受的, 因为 3 种分析方法均做出了阴性判断。

2.3 hRFT2基因与ESCC、GCA的关系

通过 BLAST 搜索 rs13042395 在人类基因组中的定位, 并比较 rs13042395 和 hRFT2 基因 (NG\_027687.1) 在 AL118502.38 序列中的位置, 结果 (图 2) 表明, rs13042395 在第 20 号染色体中定位于 hRFT2 基因下游, 介于 hRFT2 和 C20orf55 基因之间, 与两种基因的起始密码子分别间隔 8.1 kb 和 70.9 kb, 属于基因间 SNP。

3 讨论

本研究从 NCBI SNP 数据库、欧洲 SNP 数据库以及日本 SNP 数据库中共检索到 hRFT2 基因 SNPs 566 个, 从中筛选出 395 个有效 SNPs, 并对其进行定位、分类以及功能预测。从数据库来源看, 欧洲 SNP 数据库在 3 个数据库中的可信度最低, 收录的 SNPs 中 43.2% (16/37) 属于基因间 SNPs。NCBI SNP 数据库收录了所有的唯一性 SNPs, 而且数据库的数据更新速度很快, 在笔者跟踪 hRFT2 基因近一年的时间里, NCBI SNP 数据库新增 hRFT2 SNPs 130 余个, 同时删除了曾经收录的 21 个 SNPs。但

表 3 hRFT2 基因 nsSNPs 及其影响预测

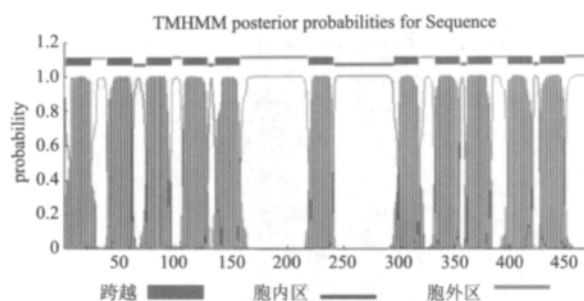
SNP 位点	核苷酸变化	氨基酸变化	外显子序号	分值			可能影响 <sup>d</sup>
				SIFT <sup>a</sup>	SNPs3D <sup>b</sup>	PolyPhen <sup>c</sup>	
rs146302587	c.37G>C	p.G13R	2	0.00	-2.83	1.00	+
rs148387972	c.58A>C	p.I20L	2	0.11	-0.06	0.972	±
rs144337813	c.194G>A	p.R65Q	2	0.37	1.43	0.141	-
rs149622425	c.203G>C	p.C68S	2	0.60	1.85	0.004	-
rs35655964	c.222C>G	p.I74M	2	0.01	1.55	0.427	±
rs142064992	c.250G>A	p.V84I	2	0.13	0.96	0.010	-
rs140391358	c.328G>T	p.V110F	2	0.05	0.04	0.989	±
rs142157418	c.395G>A	p.R132Q	2	0.49	1.13	0.275	-
rs146492942	c.448C>T	p.L150F	2	0.00	-0.33	0.571	+
rs112034541	c.499G>A	p.V167I	2	0.26	1.61	0.816	±
rs139629967	c.511G>A	p.E171K	2	0.00	1.47	0.055	±
rs6054614	c.521A>G	p.D174G	2	0.72	1.06	0.000	-
rs150702278	c.548G>A	p.R183K	2	1.00		0.000	-
rs113864305	c.551A>G	p.E184G	2	0.45		0.000	-
rs143511669	c.676C>T	p.L226F	3	0.16	2.40	0.159	-
rs146925273	c.706G>A	p.V236M	3	0.06	0.39	0.001	-
rs148832841	c.733C>A	p.P245T	3	0.14	0.92	0.880	±
rs150159842	c.787C>T	p.H263Y	3	0.28	0.65	0.128	±
rs3746804	c.800C>T	p.P267L	3	0.02	2.01	0.000	±
rs111912321	c.803G>A	p.R268Q	3	0.43	2.31	0.011	-
rs3746803	c.833C>T	p.T278M	3	0.09	2.78	0.551	±
rs142265627	c.865G>A	p.E289K	3	0.06	0.33	0.701	±
rs3746802	c.907A>G	p.I303V	3	0.17	-0.55	0.999	±
rs145194879	c.928G>A	p.V310I	3	0.29	1.31	1.000	±
rs139542858	c.1009G>A	p.A337T	3	0.04	0.26	0.563	±
rs76947760	c.1048T>A	p.L350M	3	0.06	-0.74	0.794	±
rs148181353	c.1102G>A	p.V368M	4	0.08	1.78	0.009	-
rs143033168	c.1175A>C	p.H392P	4	0.21	-0.17	0.666	±
rs145431028	c.1371C>G	p.F457L	5	0.00	-1.90	1.000	+
rs140360713	c.1381G>T	p.D461Y	5	0.07	1.37	1.000	±

a, 分值 > 0.05, 可以耐受 ;分值 0.05, 不可耐受, 分值越低, 破坏性越大。b, 分值 > 0.00, 可以耐受 ;分值 0.00, 不可耐受, 分值越低, 破坏性越大。  
c, 分值 < 0.50, 可以耐受 ;分值 0.50, 不可耐受, 分值越高, 破坏性越大。d, +, 有影响 ;-, 没有影响 ;±, 可能有影响。 , 无法分析

是数据库的检索功能存在明显不足，如果直接使用基因名称（hRFT2 或 C20orf54）而不是基因序列号进行检索，159 个已被 NCBI SNP 数据库收录的有效 SNPs 将被遗漏。另外，NCBI SNP 数据库收录的条目并非完全是严格意义上的 SNP，而是包含了 SNP 和突变，由于资料的限制，本研究并未加以区分。

nsSNP 引起的 AAS 可能对蛋白质的结构和稳

定性产生影响，进而影响蛋白质的功能和活性，因而重点分析了位于基因编码区的 nsSNPs。从这些 nsSNPs 在 cDNA 中的分布看，第二外显子包含了近半数的 nsSNPs，出现 nsSNPs 的频率最高，平均每 40.5 个核苷酸出现 1 个。同时，第二外显子还包含唯一已发现的无义 SNP。第五外显子出现 nsSNPs 的频率最低，每 105 个核苷酸才出现 1 个，提示第



胞内区: 1, 62-73, 129-134, 242-295, 356-361, 422-427 aa; 胞外区: 25-38, 97-105, 158-218, 319-332, 385-398, 451-469 aa; 跨膜区: 2-24, 39-61, 74-96, 106-128, 135-157, 219-241, 296-318, 333-355, 362-384, 399-421, 428-450 aa

图1 hRFT2 前体蛋白跨膜区预测

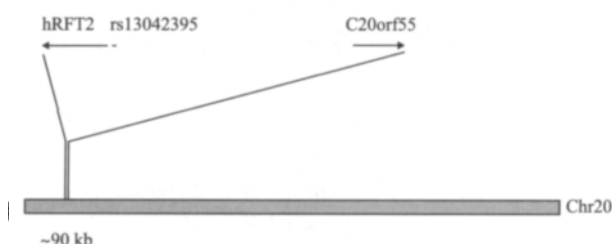


图2 hRFT2、rs13042395 和 C20orf55 基因在人类第20号染色体中的相对位置

五外显子所编码的羧基端 70 个氨基酸序列可能对 hRFT2 蛋白具有重要意义。尽管对 hRFT2 蛋白功能区进行预测结果表明, 羧基端 70 个氨基酸序列不构成任何有重要功能的结构域或模体, 但对 hRFT2 定位于细胞顶端膜及转运核黄素的功能至关重要<sup>[17]</sup>。

因为已发现与疾病相关联的 SNPs 多数为 nsSNPs, 所以 nsSNP 一直是 SNP 研究的中心。在本研究中, 我们着重分析了 hRFT2 基因的 nsSNPs。在 31 个 nsSNPs 中, 无义 SNP rs11477762 将使 hRFT2 蛋白缺失羧基端 346 aa, 占成熟蛋白全长的 77.4%, 这无疑会导致 hRFT2 缺失大部分功能。如果此 SNP 确实存在, 必然会引起极其严重的疾病。对于另外 30 个 nsSNPs, 我们通过 SIFT、SNPs3D 和 PolyPhen-2 等 AAS 预测方法分析了 AASs 对 hRFT2 的影响。从分析结果看, SIFT 和 SNPs3D 的阳性率比较接近, 分别为 26.7% 和 23.3%, PolyPhen-2 的阳性率则达到 50%, 根据已有的 nsSNPs 研究结果判断, 这个比例明显偏高, 因为大部分自然选择的 AASs 对蛋白的影响很小甚至可以忽略。SNPs3D 对 rs150702278

和 rs113864305 引起的 Arg183Lys 和 Glu184Gly 无法进行分析, 在分析 Thr278Met (rs3746803) 和 Pro267Leu (rs3746804) 时则误认为所提交的野生型错误, 表明该方法在程序设计上仍有待改进。3 种方法共同鉴定出 3 个可能对 hRFT2 极为有害的位点, 其中 rs146302587 导致第 13 位氨基酸由 Gly 突变为 Arg, 该位点位于 hRFT2 蛋白的信号肽中, 有可能影响 hRFT2 在细胞中的靶向运输。rs146492942 引起 Leu150Phe 替换, 该位点位于成熟蛋白的第 4 跨膜区, 与此区相邻的是 hRFT2 蛋白最长的膜外区, Leu 和 Phe 均为疏水性很强的氨基酸, 二者替换似乎不会影响蛋白的定位和稳定性。rs145431028 和 rs146492942 刚好相反, 引起 Phe457Leu 替换, 该突变已被证明是引起 BVVL 综合征的主要病因<sup>[3]</sup>。rs76947760 导致的 Leu350Met 替换也已被证明可引起 BVVL 综合征<sup>[3]</sup>, 3 种分析方法中, 只有 SIFT 认为 Leu350Met 可能对 hRFT2 的功能无影响, 但综合分值 (0.06) 已经极其接近临界点 (0.05), 而且在 SIFT 的上一个版本中 (JCVI 版, <http://sift.jcvi.org>), Leu350Met 替换曾被认为是非常有害的 (分值 0.02)。rs3746802 引起的 I303V 突变也只有 SIFT 判定为阴性, 但此前两个阶段的 JCVI 版均将其判为阳性 (分值 0.02/0.04)。rs148387972 和 rs140391358 引起的 AASs 也可能对 hRFT2 的功能有一定影响, 因为每个位点都有 2 个阳性结果, 并且阴性结果也十分接近临界点。因此, 这 3 个位点也值得研究者关注。

本研究还对近来广受关注的 SNP 位点 rs13042395 进行了分析。根据 hRFT2 基因、C20orf55 基因以及 rs13042395 在人类第 20 号染色体片段 AL118502.38 中的定位判断, rs13042395 应属基因间 SNP, 因而 NCBI SNP 数据库并没有将该位点定义为 hRFT2 基因 SNP。虽然 rs13042395 距离 hRFT2 基因的 5' 端仅有 281 bp, 但基因的范围因为非转录区的存在而很难准确界定, 而且目前尚没有任何证据表明 rs13042395 所在的区域是 hRFT2 基因的远距离调控元件 (如增强子或沉默子), 在缺乏试验数据的情况下, 很难判断 rs13042395 是否会影响 hRFT2 基因的表达。因此, 即使 rs1304239 确实与 ESCC 和 GCA 有关, 推测 hRFT2 是 ESCC 和 GCA 的易感基因仍显牵强。最近 Zhang 等<sup>[18]</sup>的研究结果表明,

rs13042395 与 GCA 并没有明确的关联,也从另一个方面调低了 hRFT2 基因和 GCA 有关的可能性。

通过生物信息学的方法,从现有 hRFT2 基因 SNPs 中筛选出了可能会影响 hRFT2 功能的 nsSNPs。尽管 3 种分析方法得出的结论不完全一致,但有关 BVVL 综合征的研究已经部分证实了数据的准确性,为今后的研究提供了有价值的线索。

#### 4 结论

在 hRFT2 基因的 31 个 nsSNPs 中,第 1 跨膜区的 rs146302587、第 4 跨膜区的 rs11477762、第 5 跨膜区的 rs146492942、第 8 跨膜区的 rs76947760 和羧基端膜外区的 rs145431028 等 5 个 nsSNPs 可能严重影响 hRFT2 蛋白的功能,从而增加人类患病的风险,其中最后 2 个 nsSNPs 的影响已得到临床证明,第 1 跨膜区的 rs148387972、第 4 跨膜区的 rs140391358 和第 7 跨膜区的 rs3746802 也可能对 hRFT2 的功能或定位有较大影响。

#### 参考文献

- [1] Yamamoto S, Inoue K, Ohta KY, et al. Identification and functional characterization of rat riboflavin transporter 2. *J Biochem*, 2009, 145 (4):437-443.
- [2] Fujimura M, Yamamoto S, Murata T, et al. Functional characteristics of the human ortholog of riboflavin transporter 2 and riboflavin-responsive expression of its rat ortholog in the small intestine indicate its involvement in riboflavin absorption. *J Nutrition*, 2010, 140 (10):1722-1727.
- [3] Green P, Wiseman M, Crow YJ, et al. Brown-Vialetto-Van Laere syndrome, a ponto-bulbar palsy with deafness, is caused by mutations in RFT2. *Am J Human Genet*, 2010, 86 (3):485-489.
- [4] Johnson JO, Gibbs JR, Van Maldergem L, et al. Exome sequencing in Brown-Vialetto-van Laere syndrome. *Am J Human Genet*, 2010, 87 (4):569-570.
- [5] Bosch AM, Abeling NG, Ijlst L, et al. Brown-Vialetto-Van Laere and Fazio Londe syndrome is associated with a riboflavin transporter defect mimicking mild MADD :a new inborn error of metabolism with potential treatment. *J Inherit Metab Dis*, 2011, 34 (1):159-164.
- [6] Wang LD, Zhou FY, Li XM, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies susceptibility loci at PLCE1 and RFT2. *Nat Genet*, 2010, 42 (9):759-763.
- [7] Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res*, 2002, 12 (3):436-446.
- [8] Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res*, 2011, 39 (suppl. 2):W215-222.
- [9] Williams SG, Madan R, Norris MG, et al. Using knowledge of protein structural constraints to predict the evolution of HIV-1. *J Mol Biol*, 2011, 410 (5):1023-1034.
- [10] Stitzel NO, Tseng YY, Pervouchine D, et al. Structural location of disease-associated single-nucleotide polymorphisms. *J Mol Biol*, 2003, 327 (5):1021-1030.
- [11] Ferrer-Costa C, Orozco M, de la Cruz X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *J Mol Biol*, 2002, 315 (4):771-786.
- [12] Bremel RD, Homan EJ. An integrated approach to epitope analysis I : Dimensional reduction, visualization and prediction of MHC binding using amino acid principal components and regression approaches. *Immunome Res*, 2010, 6 :7.
- [13] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 2009, 4 (7):1073-1082.
- [14] Yue P, Melamud E, Moulton J. SNPs3D : candidate gene and SNP selection for association studies. *BMC Bioinformatics*, 2006, 7 :166.
- [15] Yue P, Moulton J. Identification and analysis of deleterious human SNPs . *J Mol Biol*, 2006, 356 (5):1263-1274.
- [16] Adzhubei IA, Schmidt S, et al. A method and server for predicting damaging missense mutations. *Nat Methods*, 2010, 7 (4):248-249.
- [17] Subramanian VS, Rapp L, Marchant JS, et al. Role of cysteine residues in cell surface expression of the human riboflavin transporter-2 (hRFT2) in intestinal epithelial cells. *Am J Physiol Gastrointest Liver Physiol*, 2011, 301 (1):G100-109.
- [18] Zhang H, Jin G, Li H, et al. Genetic variants at 1q22 and 10q23 reproducibly associated with gastric cancer susceptibility in a Chinese population. *Carcinogenesis*, 2011, 32 (6):848-852.

(责任编辑 马鑫)