



## 研究报告

### Research Report

# 人 *GJB2* 分子进化与耳聋遗传效应的关联性分析

范焱<sup>1</sup>, 喻岸竹<sup>1</sup>, 郁芸<sup>2</sup>, 曹新<sup>2</sup>, 韩新焕<sup>2</sup>

1 南京医科大学临床医学院, 南京, 210029;

2 南京医科大学基础医学院, 南京, 210029

✉ 通讯作者: hxinhuan@163.com; ✉ 作者

医学遗传学与疾病研究, 2012 年, 第 1 卷, 第 2 篇 doi: 10.5376/mgdr.cn.2012.01.0002

收稿日期: 2012 年 04 月 02 日

接受日期: 2012 年 07 月 23 日

发表日期: 2012 年 07 月 25 日

本文首次发表在《基因组学与医学生物学》(2012 年第 31 卷第 1 期 26-34 页)上。现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

建议最佳引用格式:

引用格式(中文):

范焱等, 2012, 人 *GJB2* 分子进化与耳聋遗传效应的关联性分析, 医学遗传学与疾病研究(online) Vol.1 No.2 pp.6-14 (doi: 10.5376/mgdr.cn.2012.1.0002)

引用格式(英文):

Fan et al., 2012, Analysis on Association between Molecular Evolution of Human *GJB2* Gene and Hereditary Effect with Deafness, Yixue Yichuanxue Yu Jibing Yanjiu (online) Vol.1 No.2 pp.6-14 (doi: 10.5376/mgdr.cn.2012.2.0002)

**摘要** 间隙连接蛋白  $\beta 2$  (*GJB2*) 基因突变与遗传性非综合征性耳聋密切相关, 其广泛的突变类型及特异性的热点突变被认为是一种独特的致聋基因。本研究应用生物信息学方法对 17 个物种的 *GJB2* 蛋白进行了系统发育、保守性、跨膜区结构、三维结构和错义突变的分析, 并结合已有报道的实验结果进行关联性分析。分析预测获得了 166 个固定的氨基酸位点, 2 个非保守区以及 2 个空间结构保守位点; 关联性分析证实发生在保守位点的突变致病性高, 非保守区突变的概率致病性小, 跨膜区且改变氨基酸性质的突变, 可能影响蛋白的空间结构而改变膜通道的通透性。本文为进一步研究 *GJB2* 基因突变与聋病的关联性及其发病机制提供了理论依据, 同时, 这种研究思路对其它疾病的相关研究具有一定的借鉴价值。

**关键词** *GJB2* 基因; 非综合征型耳聋; 分子进化; 遗传效应, 关联分析

## Analysis on Association between Molecular Evolution of Human *GJB2* Gene and Hereditary Effect with Deafness

Fan Yi<sup>1</sup>, Yu Anzhu<sup>1</sup>, Yu Yun<sup>2</sup>, Cao Xin<sup>2</sup>, Han Xinhuan<sup>2</sup>

1 Clinical Medical College, Nanjing Medical University, Nanjing, 210029;

2 School of Basic Medical Sciences, Nanjing Medical University, Nanjing, 210029

✉ Corresponding author: hxinhuan@163.com; ✉ Authors

**Abstract** Gap junction protein  $\beta 2$  (*GJB2*) gene mutations are close associated with the hereditary non-syndromic hearing loss. Because of their wide mutation types and specific high frequency mutations, it was considered to be an unique deafness gene. In this research, *GJB2* proteins were systematically studied with the means of bioinformatics including molecular phylogeny, conservation, transmembrane region, three-dimensional structure and missense mutation analysis, and also correlation analysis was carried out by combining some previous reported experimental results. 166 amino acid fixed sites, two non-conserved regions and 2 sites of conserved spatial structure were predicted. Association analysis confirmed that high pathogenicity was the occurrence of mutations in the conserved sites, low probability of pathogenicity happened in non-conservative mutations as well. The mutations in transmembrane regions with the property of amino acid changing may affect the spatial structure of proteins leading to change the permeability of membrane channels. This study would provide a theoretical basis for further study of the deafness associating with *GJB2* gene mutations and molecular mechanics of pathogenesis, and also the research ideas have a certain reference value for other diseases.

**Keywords** *GJB2* gene, Non-syndromic hearing loss, Molecular evolution, Hereditary effect, Correlation analysis

*GJB2* 基因编码的间隙连接蛋白 26 (Connexin26, Cx26) 是间隙连接蛋白家族中重要的成员, 在内耳中是  $K^+$  和  $IP_3$  等第二信使分子的通道, 对听觉生理和病理具有重要作用。*GJB2* 基因突变与遗传性非

综合征型耳聋(nonsyndromic hearing loss, NSHL)有关, 是最常见的耳聋基因(Mahdieh et al., 2010; Wu et al., 2011)。在 *GJB2* 基因中, 已发现 100 多种突变与耳聋有关(<http://davinci.crg.es/deafness/>), 该基



因的突变具有种族和地域特征, 如 35delG、167delT 和 235delC 分别在高加索人、犹太人和亚洲人群中较为常见。在我国 *GJB2* 基因 235delC 的突变致聋的频率也高于其它亚洲国家(Dai et al., 2007)。由于 *GJB2* 基因突变种类多, 且仍有新的突变被不断发现, 如何科学地预测新生致病突变以及与疾病表型的关联性, 是一个值得探索的课题。现今利用生物信息学方法从已有的生物数据库中挖掘有意义的信息, 对新发现的突变进行致病性预测和可能的功能分析, 此为疾病基因组学研究提供了一个有用的平台。基于此, 本文利用生物信息学的方法分析 *GJB2* 基因的序列进化和分子特征, 并结合突变数据库与相关的研究报道, 探讨 *GJB2* 基因变异与耳聋遗传效应之间的关系。

## 1 结果与分析

### 1.1 17 个物种及其 *GJB2* 蛋白序列信息

本研究选取的 17 个物种的 *GJB2* 蛋白序列信息(表 1), 包括了 15 个哺乳纲和 2 个两栖纲物种, 在哺乳纲中包括了灵长目、食肉目、奇蹄目、偶蹄目及啮齿目。选取的 17 个 *GJB2* 蛋白序列中, 除了西

部爪蛙和非洲爪蛙的 *GJB2* 蛋白序列长度为 248 以外, 均为 226。

### 1.2 *GJB2* 系统发育分析

从 *GJB2* 的系统发育树中可见(图 1), 哺乳类动物的 *GJB2* 与两栖类动物的 *GJB2* 亲缘关系较远。此外, 在 15 种哺乳类动物的 *GJB2*、2 种两栖类动物的 *GJB2* 和哺乳类与两栖类 *GJB2* 祖先序列, 分别进行的多序列比对结果中, 相似度(固定氨基酸残基的频率)分别为 83.6% (189/226)、94.0% (233/248) 和 70.7% (159/225)。

可见, 哺乳类动物的祖先序列与两栖类动物的祖先序列相比, 相似度较低(图 2), 与 *GJB2* 系统发育树中的结果一致。

### 1.3 *GJB2* 保守区分析

在 17 个物种的 *GJB2* 多序列比对中(图 2), 人类 *GJB2* 有 65.93% (149/226)的氨基酸残基是固定的, 23.89% (54/226)的氨基酸残基是保守的。比较人类的 *GJB2* 序列与进化树中各结点处的祖先序列, 结果显示, 人类 *GJB2* 中 70.35% (159/226)的氨

表 1 17 个物种 *GJB2* 蛋白序列信息

Table 1 *GJB2* protein sequence information from 17 species

编号 Number	物种名称 Name of species used in this research	长度 Length	纲 Class	目 Order	GenBank 登录号 Accession number
1	人类 human ( <i>homo sapiens</i> )	226	M	PR	AAL87696.1
2	黑猩猩 chimpanzee ( <i>pan troglodytes</i> )	226	M	PR	BAK62975.1
3	苏门答腊猩猩 Sumatran orangutan ( <i>pongo abelii</i> )	226	M	PR	XP_002824091.1
4	家犬 dog ( <i>canis lupus familiaris</i> )	226	M	CA	NP_001184077.1
5	猕猴 rhesus macaque ( <i>macaca mulatta</i> )	226	M	PR	NP_001038200.1
6	马 horse ( <i>equus caballus</i> )	226	M	PE	XP_001488623.1
7	绵羊 sheep ( <i>Ovis aries</i> )	226	M	AR	NP_001009780.1
8	牛 cattle ( <i>Bos taurus</i> )	226	M	AR	NP_001077106.1
9	中华仓鼠 Chinese hamster ( <i>Cricetulus griseus</i> )	226	M	RO	EGV92232.1
10	褐家鼠 Norway rat ( <i>Rattus norvegicus</i> )	226	M	RO	NP_001004099.1
11	家鼠 house mouse ( <i>Mus musculus</i> )	226	M	RO	NP_032151.1
12	南美洲豚鼠 Domestic guinea pig ( <i>Cavia porcellus</i> )	226	M	RO	NP_001166294.1
13	西部爪蛙 Western clawed frog ( <i>Xenopus (Silurana) tropicalis</i> )	248	A	AN	NP_001016984.1
14	非洲爪蛙 African clawed frog ( <i>Xenopus laevis</i> )	248	A	AN	NP_001080478.1
15	白掌长臂猿 common gibbon ( <i>Hylobates lar</i> )	226	M	PR	AAL03975.1
16	婆罗洲猩猩 Bornean orangutan ( <i>Pongo pygmaeus</i> )	226	M	PR	AAL03974.1
17	西部猩猩 Western gorilla ( <i>Gorilla Gorilla</i> )	226	M	PR	AAL03973.1

注: M: 哺乳纲; A: 两栖纲; PR: 灵长目; CA: 食肉目; PE: 奇蹄目; AR: 偶蹄目; RO: 啮齿目; AN: 无尾目

Note: M: Mammalian; A: Amphibian; PR: Primates; CA: Carnivore; PE: Perissodactyla; AR: Artiodactyla; RO: Rodentia; AN: Anura

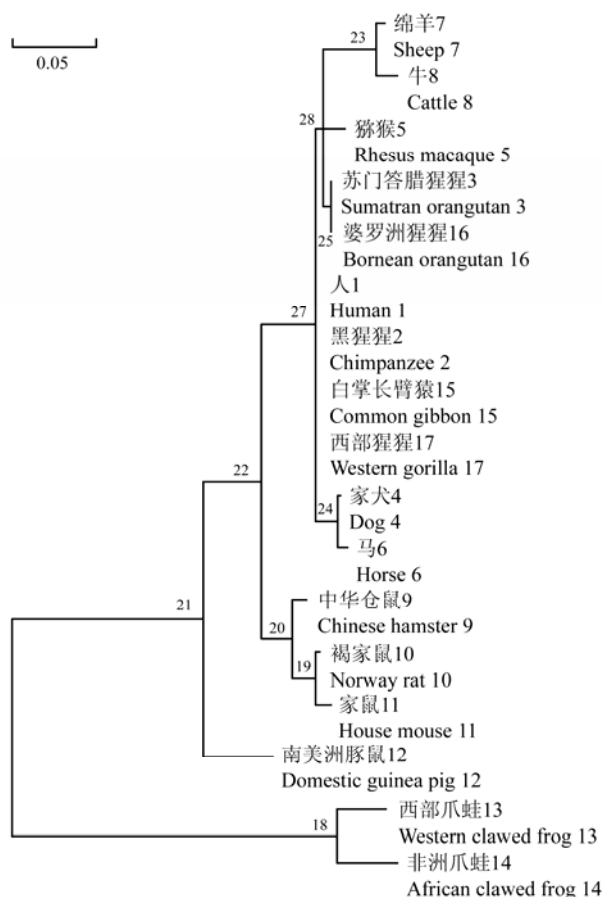


图 1 GJB2 蛋白系统发育树

注: 物种名称后面的数字代表该物种序列编号, 系统发育树各结点处的数字代表该处祖先序列的编号; 此外, 编号 21 的祖先序列为树中哺乳动物的祖先序列, 编号 18 的祖先序列为树中两栖类动物的祖先序列

Figure 1 The phylogenetic tree of GJB2 proteins

Note: The number behind the species name represents the species' sequence number. The number in the node represents the ancestral sequence number of this node; No.21 represents the sequence number of the tree's mammalian ancestor and No.18 represents the sequence number of this tree's amphibian ancestor

氨基酸残基是固定的, 20.80% (47/226)的氨基酸残基是保守的。此外, 在 17 个物种序列比对中, 其中固定的氨基酸残基在祖先序列中也是一致的。将人类 GJB2 序列、哺乳动物的祖先序列和两栖类动物的祖先序列进行多序列比对, 有 73.45% (166/226)的位点显示是固定的。由于基因进化作用, GJB2 出现多态性位点从而形成了不同物种, 其中部分多态性位点在人类与祖先序列比对中是固定的, 因此, 将人类 GJB2 序列中的这些位点视为固定位点可提高分析的精确度。所以, 最终得到了 166 个固定的氨基酸残基位点。

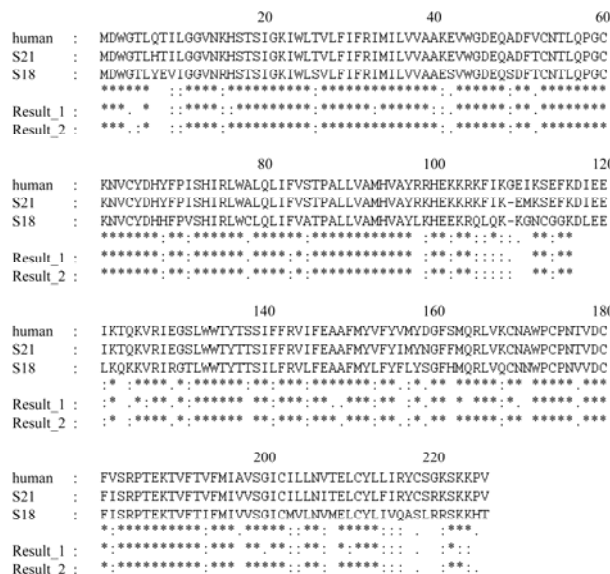


图 2 人类、哺乳动物和两栖类动物祖先序列比对分析结果  
 注: S21 为进化树中哺乳类动物祖先序列, S18 为进化树中两栖类动物祖先序列; 序列下方的结果为 human, S21 和 S18 比对结果; Result\_1 为 17 个物种多序列比对结果, Result\_2 为 human 和进化树中各结点处祖先序列多序列比对结果

Figure 2 The results of multiple sequence alignment among human, mammalian and amphibian ancestor

Note: S21 represents the sequence number of the tree's mammalian ancestor and S18 represents the sequence number of this tree's amphibian ancestor; The alignment result among human, S21 and S18 is below the sequences; Result\_1 is the multiple sequence alignment result of 17 species, and Result\_2 is the result of ancestors of the tree's nodes and human

人类 GJB2 序列的 SMART 分析(Kelley et al., 1998)显示, GJB2 是由 5 个非跨膜区(2 个细胞外环, 1 个 N 末端, 1 个 C 末端和 1 个细胞内环)和 4 个跨膜区组成(图 3), 53↔180、60↔174 及 64↔169 为二硫键结合位点。研究表明, 间隙连接蛋白家族成员之间的同源性主要在细胞外的两个结构域, 其次是跨膜结构域, 细胞质内的结构域差别最大。经比对分析各区域的保守度, 结果显示细胞外环区、跨膜区与胞质区的平均保守度分别为 94.74%、93.38% 和 72.78%(表 2)。GJB2 蛋白细胞外环的高度保守性, 推测其可能与间隙连接蛋白其他家族成员在细胞外侧存在相互作用。此外, 由于胞质区 CT 结构域的保守度极低, 在本文定义为“1 号非保守区”。

#### 1.4 人类 GJB2 蛋白三维结构比较

人类 GJB2 的三维结构包括了 6 个  $\alpha$  螺旋、5 个  $\beta$  折叠、10 个转角以及 7 个卷曲(图 3)。将人类 GJB2 三维结构与其余 16 个物种进行两两比较, 得分  $\geq 98$  的序列三维结构与人类的几乎完全重叠, 分别是编



表 2 人类 GJB2 结构域保守度与突变位点分布频率

Table 2 The degrees of conservation of human GJB2 protein domains and the distribution frequencies of GJB2 mutation sites

区号 Number	范围 Range	结构域 Domain	类型 Type	保守度(%) Degrees of Conservation (%)	突变位点分布频率(%) Frequency (%)
1	1~20	NT	ND	85.00	25.0
2	21~40	M1	TD	90.91	22.7
3	41~75	E1	ND	100.00	22.9
4	76~98	M2	TD	95.65	47.8
5	99~131	CL	ND	78.79	27.3
6	132~154	M3	TD	95.65	21.7
7	155~192	E2	ND	89.47	28.9
8	193~215	M4	TD	91.30	26.1
9	216~226	CT	ND	54.55	0.0

注: ND 表示非跨膜区; TD 表示跨膜区

Note: ND: The non-transmembrane domain; TD: The transmembrane domain

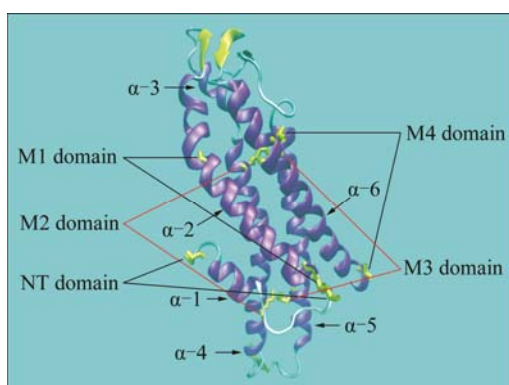


图 3 人类 GJB2 蛋白三维结构

Figure 3 The three-dimensional structure of human GJB2 protein

号为 2、3、4、5、6、15、16 和 17 的 8 个物种(图 4)。将比对得分较低的西部爪哇(score=73)的 GJB2 三维结构与人类的进行空间比对, 发现两者在第 109 至 125 氨基酸残基段存在空间结构的差异(图 5)。进而, 将人类与其余物种的 109~125 区段分别进行了结构比对, 发现物种编号为 2、3、4、5、6、9、10、11、15、16 及 17 的三维结构与人类的几乎完全重叠, 而编号为 7、8、12、13 和 14 的三维结构在 109~125 残基段与人类存在差异(图 6)。分析发现, 编号为 4、5、6、9、10、11 仅在第 113 位或第 123 位存在差异, 编号 7、8、13、14 在 113 和 123 位同时存在差异, 编号 12 仅在 123 位有差异, 但其与人类亲缘关系较远(图 7)。所以, 推测当人类 GJB2 序列的 113 位和 123 位同时发生变化时, 其序列 109-125 段对应的空间结构将有可能发生改变。因此, 人类 GJB2 序列第 113 和 123 位点可能是“空间结构保守位点”。

由于 109~125 区段位于 99~131 非跨膜区域内, 其保守度(78.79%)相比其他区域较低(除 1 号非保守区), 在本文将该区域定义为“2 号非保守区”。

### 1.5 人类 GJB2 基因错义突变分析

从 CRG (Center for Genomic Regulation) 数据库中获得 GJB2 基因的 77 种与 NSHL 相关的致病性错义突变位点信息, 其中包括 60 种常染色体隐性遗传错义突变, 8 种常染色体显性遗传错义突变和 9 种功能尚未明确的致病性错义突变(表 3)。这些突变共涉及了 60 个氨基酸位点, 其中 90% (54/60) 为保守性位点。从各区域的分布频率可知(表 2), 突变位点多数集中在区 4 (47.8%), 1 号非保守区的突变位点分布频率极低(0%), 2 号非保守区较低(29.4%)。因此, 可以推测新的致病性错义突变位点出现在 M2 结构域的概率较高, 出现在 CT 结构域的概率较低。此外, 在上述两个空间结构保守位点均出现了错义突变(S113R 和 T123N), 有研究表明, S113R 突变使 GJB2 不能诱导同型连接通道的形成(Martínez et al., 2009), 影响了其膜通道转运功能, 生物信息学预测结果也证实该位点的突变将改变 GJB2 蛋白的三维结构。

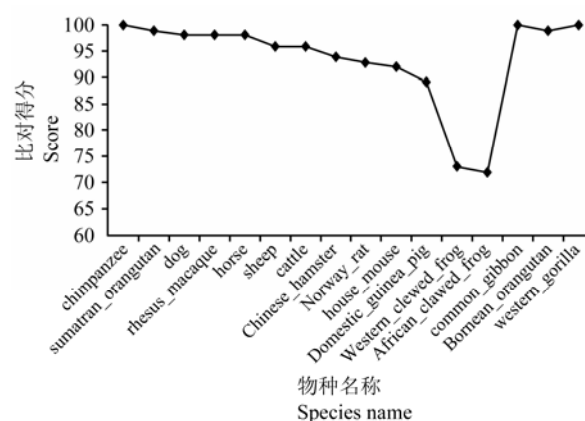


图 4 人类与其余物种两两比对结果

Figure 4 The alignment scores between human and any other species



表 3 非综合征耳聋 GJB2 致病性错义突变

Table 3 The pathogenic missense mutations of GJB2 in nonsyndromic hearing loss

隐性遗传突变 Recessive mutation			显性遗传突变 Dominant mutation	功能未明的突变 Unknown mutation
A	B	C		
T8M	Y65X	K122I	W44S	N54I
G12V	W77R	Q124X	W44C	V84A
K15T	W77X	R127H	R75Q	T123N
S19T	L79P	W133X	R143Q	E129K
I20T	Q80X	Y136X	M163L	Y155X
W24X	Q80P	S139N	D179N	M163V
V27I+E114G	Q80R	R143W	R184Q	A171T
R32C	I82M	E147K	C202F	F191L
R32L	V84L	E147X		A197S
R32H	S85P	R165W		
V37I	A88S	W172X		
A40E	L90V	C174R		
A40G	L90P	P175T		
W44X	M93I	V178A		
G45E	V95M	R184W		
E47X	Y97X	R184P		
E47K	H100Y	S199F		
Q57X	H100L	I203K		
C64X	E101G	N206S		
Y65X	S113R	L214P		

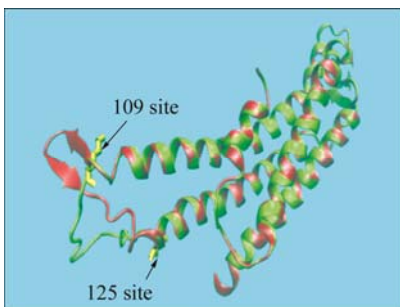


图 5 人类与西部爪蛙 GJB2 蛋白三维空间结构对比注: 图中柱状体标出的分别为位点 109 和位点 125

Figure 5 The molecular overlap map of the three-dimensional structure alignment of human and western clawed frog GJB2 protein

Note: The pillars represent the sites 109 and 125, respectively

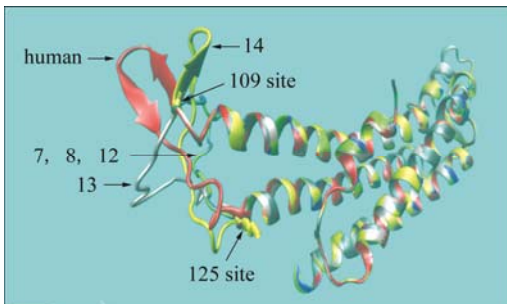


图 6 人类与绵羊, 牛, 南美豚鼠, 西部爪蛙及非洲爪蛙的 GJB2 蛋白三维结构重叠图

Figure 6 The molecular overlap map of the three-dimensional structure alignment of human, sheep, cattle, Domestic guinea pig, western clawed frog and African clawed frog GJB2 protein

Species	Sequence
Norway_rat	LIMVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
House_mouse	LIMVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Chinese_hamster	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Sheep	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Cattle	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
dog	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
horse	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Sumatran_orangutan	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Bornean_orangutan	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Rhesus_macaque	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Human	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Chimpanzee	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Common_gibbon	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Western_gorilla	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Domestic_guinea_pig	LIFVSTPALLVAMHVAYRRHEKRRKFKMGEIKRNEFKDIEEIKTKQVRIEG
Western_clawed_frog	LIFVATPALLVAMHVAYLRHEEKRRKQKQK-RGNTGGKDLLEELKRRKVRIRG
African_clawed_frog	LIFVATPALLVAMHVAYLRHEEKRRKQKQK-RGNTGGKDLLEELKRRKVRIRG

图 7 17 个物种 GJB2 序列部分比对结果

Figure 7 Multi-alignment of GJB2 protein sequences from 17 species

### 1.6 人类 GJB2 蛋白 SOSUI 分析

对人类 GJB2 蛋白进行 SOSUI 分析, 结果显示的跨膜区与 SMART 分析结果趋于一致。60 个突变位点近半数集中在跨膜区(图 8), 其中 23 个位于疏水性残基, 37 个位于亲水性残基。在组成的跨膜区的残基中, 28.3% (26/92)为亲水性残基, 而分布在跨膜区的突变位点中有 46.2% (12/26)位于亲水性残基, 由此推测, 跨膜区的突变发生在亲水性残基的概率较大。杨中纯等(2010)人通过实验对 GJB2 蛋白跨膜区的突变 R32H、V95L、R143W、S199F 分别进行研究, 结果表明 R32H、R143W、S199F 突

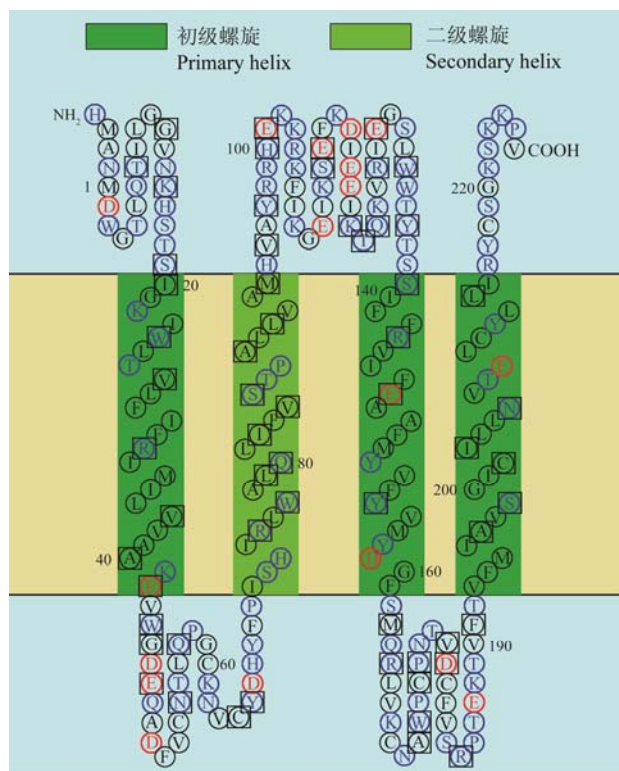


图 8 人类 GJB2 蛋白 SOSUI 分析结果  
 注: 蛋白跨膜模拟图中, 方框框出的为致病性突变位点; 4 个跨膜螺旋图中, 黑色的为疏水性残基, 蓝色的为亲水性残基, 蓝色加粗的为亲水碱性残基, 红色加粗的为亲水酸性残基

Figure 8 The result of SOSUI analysis in human GJB2 protein  
 Note: In the transmembrane protein figure, boxes highlight residues that are pathogenic mutation sites; In the transmembrane helix figure, hydrophobic residues are black; Neutral hydrophilic residues are blue; Hydrophilic basic residues are blue bold; Hydrophilic acidic residues are red bold

变 GJB2 蛋白的表达缺失, V95M 突变蛋白虽然在细胞膜上形成了间隙连接蛋白孔道却无生化偶联通透性的功能。Choi 等(2009)人也通过实验证明跨膜区 T86R 突变使蛋白无法形成间隙连接功能; Ambrosi 等(2010)人对跨膜区 4 种错义突变进行研究, 发现其中 T135A 突变导致膜通道关闭, M34A 突变导致膜通道的活性严重降低。此外, Oshima 等(2011)人对 M34A 突变蛋白的电子晶体结构进行了研究, 发现 M34A 对 N 端通道口的形成产生影响, 使通道几乎呈关闭状态。由此推测, 当错义突变发生在跨膜区且突变改变了氨基酸的性质时, 该突变将会影响蛋白在细胞膜上的表达或膜通道的通透性。

## 2 讨论

间隙连接是细胞间直接进行物质交流的唯一通道, 无论在脊椎还是无脊椎动物都发挥着重要的作用。在脊椎动物间隙连接的通道由 connexin (Cx) 蛋白家族构成, 在老鼠体内约有 20 种间隙连接蛋白基因, 在人类已知有 21 种间隙连接蛋白基因。所有的 Cx 都是具有 4 次横穿细胞膜的肽, 形成了细胞外环、内环和跨膜区的三部分结构。跨膜区和细胞外区高度保守, 然而不同 Cx 的功能差异是由这些区域的氨基酸序列的不同造成的。间隙连接介导的细胞间信号的异常可以引起全身多系统的疾病, 由于不同的 Cx 在各器官、组织的分布不同, 因此引起与之相关的疾病也不同。Cx26 在全身大部分组织器官均有表达, 并以皮肤和内耳表达水平最高。在内耳中, Cx26 蛋白在人耳蜗的血管纹、基底膜、螺旋器均呈现较高水平的表达, GJB2 基因的突变可致 Cx26 蛋白结构或功能发生障碍, 从而导致遗传性 NSHL。然而不同耳聋人群具有不同 GJB2 基因的突变谱(种族特异性), 以及 GJB2 基因突变所致临床表型多样性, 一直是该基因突变致聋研究的热点之一。

本文利用生物信息学方法对 GJB2 分子进化特征与 NSHL 的遗传效应进行了关联性分析, 结果表明, 人类 GJB2 氨基酸序列具有高度保守性, 分析推测约有 166 个固定的氨基酸位点。GJB2 序列的膜外区保守度高于跨膜区, 跨膜区保守度高于胞质区, 其中 CT 结构域保守度极低。序列 109~125 区域空间结构保守性低, 两个非保守区位点发生致病性错义突变的概率较低。序列第 113 位和第 123 位推测为“空间结构保守位点”, 预测发生在这两位上的错义突变会影响蛋白质的功能。发生在跨膜区



错义突变如果改变了氨基酸的性质, 则该突变将会影响蛋白在细胞膜上的表达或改变膜通道的通透性。参照 Dai 等(2009)对 2 063 例 NSHL 患者中发生了 *GJB2* 基因致病性错义突变的研究, 从 58 例发生致病性错义突变的患者中得到 10 种错义突变, 其中有 1 例患者同时发生了 W3X 和 R143W 突变, 另 1 例患者同时发生 Y152X 和 R127H 突变。经比对, 所检出的 10 种错义突变全部位于保守区, 且发生在跨膜区的突变(R32C, S85P, T86R, R143W)均改变了氨基酸的性质, 而在 1 号和 2 号非保守区均未发现致病性错义突变, 这一研究结果也验证了保守性位点氨基酸的改变比非保守性位点更易致病, 在非保守区出现致病性突变的可能性就较小。

通过对 *GJB2* 系统发育、三维结构及 SOSUI 分析, 对 *GJB2* 分子进化特征有了进一步的理解, 如我们首次推测并提出了人类 *GJB2* 序列第 113 位和第 123 位为“空间结构保守位点”的概念, 并将继续予以验证。在对多物种序列比对中, 我们发现绵羊与人类的序列比对得分较高(score=96), 但两者空间结构存在一定的差异, 这一结果与进化树中绵羊与人类的亲缘关系较远相一致。同样, 褐家鼠与人类序列比对得分不如绵羊的高(score=93), 但其三维结构却与人类几乎完全重叠, 符合进化树中的亲缘关系。这些结果也进一步丰富了 *GJB2* 的生物信息学的资料。

目前, 对 *GJB2* 致病突变性的研究大多采用基因序列分析、电子晶体结构分析、细胞与分子生物学等方法(Choi et al., 2009; Ambrosi et al., 2010; 杨中纯等, 2010; Oshima et al., 2011), 而结合系统生物信息学方法的研究结果未见报道。本文利用生物信息学, 对 *GJB2* 基因进行序列同源搜索, 建立了该基因的物种序列库, 对各物种基因进行多序列比对、系统发育与蛋白质三维结构空间比对分析, 鉴定出序列保守位点和区域。并对保守位点(区域)、功能性位点(区域)和致病突变分布规律进行关联性分析, 此为 *GJB2* 新致病性突变的预测以及突变致病机制研究提供了有用的前瞻性信息, 同时也为其他疾病的相关研究提供了新的思路。

## 3 材料和方法

### 3.1 数据资料

从美国国立生物技术信息中心(NCBI)的数据库中, 通过 <http://blast.ncbi.nlm.nih.gov/Blast.cgi> 查询获得 17 个物种的 *GJB2* 蛋白完整的氨基酸序列。

## 3.2 *GJB2* 蛋白序列的生物信息学分析

### 3.2.1 ClustalW 多重序列比对分析

用 ClustalW2 软件([www.ebi.ac.uk/Tools/mas/clustalw2](http://www.ebi.ac.uk/Tools/mas/clustalw2))对纳入研究的 17 个物种的 *GJB2* 蛋白序列进行多重序列比对分析(Thompson et al., 1994)。ClustalW 是目前使用最广泛的多序列比对软件, 它基于累进方法, 采用两两比对、构建向导树、依次加入序列比对三个步骤进行多序列比对。比对时, 先将多个序列两两比对构建距离矩阵; 然后在根据距离矩阵计算产生系统进化指导树, 按关系密切程度对序列进行加权; 然后从最紧密的两条序列开始, 逐步加入临近的序列并不断重新构建比对, 直到所有序列都被加入为止。

### 3.2.2 *GJB2* 的进化分析

用 MEGA5.0 软件绘制出纳入研究的 17 个物种的 *GJB2* 进化树, 并且得到进化树中各结点处的祖先序列以供进一步研究(Jones et al., 1992; Tamura et al., 2011)。Mega 是目前应用最广泛的进化分析软件之一, 在本研究中采用了 Maximum Likelihood 法。

### 3.2.3 蛋白质跨膜区与结构域分析

用 SMART (<http://smart.embl-heidelberg.de>)对 *GJB2* 蛋白进行结构域功能分析 SMART 是能识别并注释蛋白质结构域的在线软件, 它可对蛋白质的结构域进行研究和相互比较(Schultz et al., 1998; Letunic et al., 2009)。

### 3.2.4 蛋白质跨膜区与结构域分析

用 SOSUI 对 *GJB2* 蛋白的跨膜区进行分析, SOSUI (<http://bp.nuap.nagoya-u.ac.jp/sosui>)是用于对膜蛋白进行识别分析并预测其跨膜螺旋结构的软件, 其识别与预测精确度分别高达 99%和 97%(Hirokawam et al., 1998)。

### 3.2.5 蛋白质三维同源模型构建与三维结构比较分析

用 SWISS-MODEL (<http://swissmodel.expasy.org/>)对纳入研究的 17 条 *GJB2* 蛋白序列分别进行了同源建模(Guex and Peitsch, 1997; Schwede et al., 2003; Arnold et al., 2006)。然后, 用 VMD1.9 软件(Board of Trustees of the University of Illinois and others)对序列与其对应的三维结构进行分析。蛋白质三维同源模型构建法是蛋白质三维结构预测最常见的方法之一, 其原理是在蛋白质结构数据库中寻找未知结构蛋白质的同源构型, 再将同源蛋白质的结构优化构建出预测的蛋白质三维结果。



## 作者贡献

范焱和喻岸竹完成数据处理、分析以及论文写作和修改。郁芸协助数据处理和部分修改工作。曹新在遗传性耳聋的数据分析和遗传效应分析方面给予了技术支持。韩新焕构思本研究、指导实施过程、修改并定稿本论文。全体作者均阅读了本文并同意论文全部内容。

## 致谢

作者感谢中国药科大学郑珩副教授的热情帮助。

## 参考文献

- Ambrosi C., Boassa D., Pranskevich J., Smock A., Oshima A., Xu J., Nicholson B.J., and Sosinsky G.E., 2010, Analysis of four connexin26 mutant gap junctions and hemichannels reveals variations in hexamer stability, *Biophys J.*, 98(9): 1809-1819 <http://dx.doi.org/10.1016/j.bpj.2010.01.019> PMID: 20441744 PMCID: 2862186
- Arnold K., Bordoli L., Kopp J., and Schwede T., 2006, The SWISS-MODEL workspace: A web-based environment for protein structure homology modeling, *Bioinformatics*, 22(2): 195-201 <http://dx.doi.org/10.1093/bioinformatics/bti770> PMID:16301204
- Choi S.Y., Park H.J., Lee K.Y., Dinh E.H., Chang Q., Ahmad S., Lee S.H., Bok J., Lin X., and Kim U.K., 2009, Different functional consequences of two missense mutations in the *GJB2* gene associated with non-syndromic hearing loss, *Hum. Mutat.*, 30: E716-E727 <http://dx.doi.org/10.1002/humu.21036> PMID:19384972
- Dai P., Yu F., Han B., Yuan Y., Li Q., Wang G., Liu X., He J., Huang D., Kang D., Zhang X., Yuan H., Schmitt E., Han D., and Wong L.J., 2007, The prevalence of the 235delC *GJB2* mutation in a Chinese deaf population, *Genet. Med.*, 9(5): 283-289 <http://dx.doi.org/10.1097/GIM.0b013e31804d2371> PMID:17505205
- Dai P., Yu F., Han B., Liu X.Z., Wang G.J., Li Q., Yuan Y.Y., Liu X., Huang D.L., Kang D.Y., Zhang X., Yuan H.J., Yao K., Hao J.S., He J., He Y., Wang Y.Q., Ye Q., Yu Y.J., Lin H.Y., Liu L.J., Deng W., Zhu X.H., You Y.W., Cui J.H., Hou N.S., Xu X.H., Zhang J., Tang L., Song R.D., Lin Y.J., Sun S.Z., Zhang R.N., Wu H., Ma Y.B., Zhu S.X., Wu B.L., Han D.Y., Wong L.J., 2009, *GJB2* mutation spectrum in 2,063 Chinese patients with non syndromic hearing impairment, *J. Transl. Med.*, 7: 26 <http://dx.doi.org/10.1186/1479-5876-7-26> PMID:19366456 PMCID:2679712
- Guex N., and Peitsch M.C., 1997, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis*, 18(15): 2714-2723 <http://dx.doi.org/10.1002/elps.1150181505> PMID:9504803
- Hirokawa T., Boon-Chieng S., and Mitaku S., 1998, SOSUL: Classification and secondary structure prediction system for membrane proteins, *Bioinformatics*, 14(4): 378-379 <http://dx.doi.org/10.1093/bioinformatics/14.4.378> PMID:9632836
- Jones D.T., Taylor W.R., and Thornton J.M., 1992, The rapid generation of mutation data matrices from protein sequences, *Comput. Appl. Biosci.*, 8(3): 275-282 PMID:1633570
- Kelley P.M., Harris D.J., Comer B.C., Askew J.W., Fowler T., Smith S.D., and Kimberling W.J., 1998, Novel mutations in the connexin 26 gene (*GJB2*) that cause autosomal recessive (DFNB1) hearing loss, *Am. J. Hum. Genet.*, 62(4): 792-799 <http://dx.doi.org/10.1086/301807> PMID:9529365
- Letunic I., Doerks T., and Bork P., 2009, SMART 6: Recent updates and new developments, *Nucleic Acids Research*, 37(1): D229-D232 <http://dx.doi.org/10.1093/nar/gkn808> PMID:18978020 PMCID:2686533
- Mahdih N., Rabbani B., Wiley S., Akbari M.T., and Zeinali S., 2010, Genetic causes of non syndromic hearing loss in Iran in comparison with other populations, *J. Hum. Genet.*, 55(10): 639-648 <http://dx.doi.org/10.1038/jhg.2010.96> PMID:20739942
- Martínez A.D., Acuña R., Figueroa V., Maripillan J., and Nicholson B., 2009, Gap-junction channels dysfunction in deafness and hearing loss, *Antioxid. Redox Signal.*, 11(2): 309-322 <http://dx.doi.org/10.1089/ars.2008.2138>
- Oshima A., Tani K., Toloue M.M., Hiroaki Y., Smock A., Inukai S., Cone A., Nicholson B.J., Sosinsky G.E., and Fujiyoshi Y., 2011, Asymmetric configurations and N-terminal rearrangements in connexin26 gap junction channels, *J. Mol. Biol.*, 405(3): 724-735 <http://dx.doi.org/10.1016/j.jmb.2010.10.032> PMID:21094651 PMCID:3026138
- Schultz J., Milpetz F., Bork P., and Ponting C.P., 1998, SMART, a simple modular architecture research tool: Identification of signaling domains, *Proc. Natl. Acad. Sci.*, 95(11): 5857-5864 <http://dx.doi.org/10.1073/pnas.95.11.5857>
- Schwede T., Kopp J., Guex N., and Peitsch M.C., 2003, SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Research*, 31(13): 3381-3385 <http://dx.doi.org/10.1093/nar/gkg520> PMID:12824332 PMCID:168927
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S., 2011, MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, 28(10):





- 2731-2739 <http://dx.doi.org/10.1093/molbev/msr121> PMID: 21546353
- Thompson J.D., Higgins D.G., and Gibson T.J., 1994, CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Research*, 22 (22): 4673-4680 <http://dx.doi.org/10.1093/nar/22.22.4673> PMID:7984417 PMCID:308517
- Wu C.C., Hung C.C., Lin S.Y., Hsieh W.S., Tsao P.N., Lee C.N., Su Y.N., and Hsu C.J., 2011, Newborn genetic screening for hearing impairment: A preliminary study at a tertiary center, *PLoS One*, 6(7): e22314 <http://dx.doi.org/10.1371/journal.pone.0022314> PMID:21811586 PMCID:3139636
- Yang Z.C., Xiao Z.A., Xie D.H., and Xia K., 2010, Aberrant expression pattern of a novel mutation in connexin 26 gene resulting in autosomal recessive deafness, *Zhonghua Yixue Yichuanxue Zazhi (Chinese Journal of Medical Genetics)*, 27(3): 241-245 (杨中纯, 肖自安, 谢鼎华, 夏昆, 2010, 连接蛋白基因一个新致聋突变体 p.Y155X 及功能分析, *中华医学遗传学杂志*, 27(3): 241-245)