

耳聋相关基因 *COCH* 的非同义单核苷酸多态性致聋表型预测

钱旭丽, 曹新

南京医科大学基础医学院, 南京 210029

摘要: 群体凝血因子 C 同源物基因 (Coagulation factor C homology, *COCH*) 是人类发现的第一个伴前庭功能障碍的耳聋基因, 位于人类染色体 14q12-q13 上。迄今, 在 *COCH* 基因上发现 16 个位点突变导致常染色体显性遗传非综合征型耳聋 DFNA9 的发生, 其中包括 13 个非同义单核苷酸多态性 (Non-synonymous single nucleotide polymorphisms, nsSNPs) 位点。由于该基因其他 nsSNPs 的基因型与表型关系尚不清楚, 因此文章采用生物信息学方法, 从 *COCH* 基因全部的 SNPs 中分级筛选, 结合已知的致病 nsSNPs 信息及蛋白三维结构验证, 首次预测出由 *COCH* 基因编码的 cochlin 蛋白的 vWFA (Von Willebrand factor type A domain) 区的 8 个高风险致病性 nsSNPs (I176T、R180Q、G265E、V269L、I368N、I372T、R416C 和 Y424D)。同时, 对位于 LCCL (Limulus factor C, cochlin, and late gestation lung protein Lgl1) 区域的 6 个已知致病突变的 nsSNPs (P51S、G87W、I109N、I109T、W117R 和 F121S) 进行了三维结构模拟, 发现突变体均发生了环状结构或链状结构的改变。本研究对 *COCH* 基因的基因型与表型的相关性研究为遗传性耳聋筛查提供了相应的理论依据, 也对该基因所编码的 cochlin 蛋白的功能研究具有一定的指导意义。

关键词: *COCH*; nsSNPs; 耳聋; 基因型; 表型

Phenotype predictions of the pathogenic nonsynonymous single nucleotide polymorphisms in deafness-causing gene *COCH*

Xuli Qian, Xin Cao

School of Basic Medicine, Nanjing Medical University, Nanjing 210029, China

Abstract: The *COCH* (Coagulation factor C homology) gene, located in human chromosome 14q12-q13, is the first gene identified to cause vestibular dysfunction. *COCH* encodes cochlin, which contains an N-terminal LCCL (Limulus factor C, cochlin, and late gestation lung protein Lgl1) domain and a C-terminal vWFA (Von Willebrand factor type A) domain. Recently, functional research of *COCH* mutations and cochlin have come under the spotlight in the field of hereditary deafness. Approximately 16 mutations in *COCH* have been confirmed to date, among which 13 non-synonymous single nucleotide polymorphisms (nsSNPs) are the most common form of genetic variations. None-

收稿日期: 2014-12-31; 修回日期: 2015-01-21

基金项目: 国家自然科学基金项目 (编号: 31171217) 资助

作者简介: 钱旭丽, 硕士研究生, 专业方向: 遗传学。E-mail: qian_0119@163.com

通讯作者: 曹新, 教授, 研究方向: 分子遗传学。E-mail: caoxin@njmu.edu.cn

DOI: 10.16288/j.ycz.14-465

网络出版时间: 2015-4-21 11:35:05

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150421.1135.001.html>

theless, there is poor knowledge on the relationship between the genotype and the phenotype of the other nsSNPs in *COCH*. Here we analyzed deleterious nsSNPs from all SNPs in the *COCH* gene in the vWFA domain based on different computational methods and identified eight potential pathogenic nsSNPs (I176T, R180Q, G265E, V269L, I368N, I372T, R416C and Y424D) after combining literatures with 3D structures. Meanwhile, the protein structures of six reported pathogenic nsSNPs (P51S, G87W, I109N, I109T, W117R and F121S) in the LCCL domain have been constructed, and we identified aberrant structural changes in loops and chains. The prediction of pathogenic mutations for *COCH* nsSNPs will provide a blueprint for screening pathogenic mutations, and it will be beneficial to the functional research of *COCH* and cochlin in this field.

Keywords: *COCH*; nsSNPs; deafness; genotype; phenotype

耳聋是一种人类常见、高发的致残性疾病,据2013年世界卫生组织公布,全球约有3.6亿的人口患有不同程度的耳聋,新生儿中大约有1/1000患有某种先天性耳聋^[1]。耳聋直接影响患者的认知、思维和记忆能力,造成生活质量的极度下降,给社会及家庭造成重负。引起耳聋的病因主要包括环境因素和遗传因素,其中遗传因素约占60%左右^[2,3]。遗传性耳聋根据是否伴有耳外组织的异常或病变分为综合征型耳聋和非综合征型耳聋,其中非综合征型耳聋占70%左右。目前已发现145个染色体位点与非综合征型耳聋有关,包括55个常染色体显性遗传非综合征型耳聋基因位点,其中30个常染色体显性遗传非综合征型耳聋基因已经被定位克隆^[4,5]。群体凝血因子C同源物(Coagulation factor C homology, *COCH*)(OMIM 603196)基因是人类发现的第一个、且是目前唯一的伴前庭功能障碍的常染色体显性遗传非综合征型耳聋基因^[6],该基因定位于14号染色体长臂14q12-q13,广泛表达于耳蜗及前庭组织^[7]。由*COCH*基因编码的cochlin蛋白是一种分泌蛋白,包括两个主要结构—LCCL (Limulus factor C, cochlin, and late gestation lung protein Lgl1)和vWFA (Von Willebrand factor type A domain)结构域。目前,研究已表明*COCH*基因突变可导致渐进性感音神经性耳聋、前庭功能障碍症状,临床诊断为DFNA9(Autosomal dominant nonsyndromic sensorineural deafness 9)。遗传学分析发现,与DFNA9相关的基因点突变已有16种,其中13种为非同义单核苷酸多态性(Non-synonymous single nucleotide polymorphisms, nsSNPs)^[5]。因此,推测*COCH*基因尚存在其他与耳聋表型相关的nsSNPs。由于基因筛查及分子生物学实验方法需要大量的时间和费用,通过生物信息学方法对未知

表型的nsSNPs进行功能性预测是一个理想的策略,并且已有较多的文献报道验证了其预测的可行性及准确性^[8-15]。因此,本研究整合了dbSNP、HGMD和Deafness Variation Database数据库中大约2037个*COCH*基因的SNP,经过初步筛选后应用专业软件及程序(包括SIFT、PolyPhen-2、MutPred、PhD-SNP、Swiss Model等)对*COCH*基因的nsSNPs进行分级筛选预测,最终发现8个高风险致病性nsSNPs。该研究为*COCH*基因突变的筛选提供了理论依据,也为其他遗传病的致病基因的突变筛选提供了可借鉴的分析方法。

1 材料和方法

1.1 数据收集

通过dbSNP数据库(<http://www.ncbi.nlm.nih.gov/projects/SNP/>)^[16]、HGMD数据库(<http://www.hgmd.cf.ac.uk/ac/index.php>)^[17]以及Deafness Variation Database数据库(<http://deafnessvariationdatabase.org>)获得*COCH*基因的SNP。Cochlin蛋白的氨基酸序列通过UniProt数据库获得。*COCH*基因的相关信息通过美国国立生物信息中心NCBI数据库(<http://www.ncbi.nlm.nih.gov/>)获得。

1.2 nsSNPs数据集筛选

从所获得的*COCH*基因的SNP中筛选出nsSNPs,在此过程中去除3'或5'UTR,以及同义SNP。

1.3 nsSNPs相关信息整理

为了进一步对获得的*COCH*基因的nsSNPs进行预测及分析,所有*COCH*基因的nsSNPs相关信息(包括SNP ID、染色体位置、氨基酸变化、有无文献报道等)均从dbSNP数据库、HGMD数据库、Deafness Variation Database数据库及美国国立生物

信息中心 NCBI 数据库获得, 其中以 dbSNP 数据库为主, 其他数据库作为补充。

1.4 nsSNPs 表型分析

利用 SIFT 及 PolyPhen-2 软件对初步筛选获得的 nsSNPs 进行突变表型有害性分析。其中 SIFT 软件平台(<http://SIFT-dna.org>)主要利用 SIFT 算法, 基于序列保守性, 并结合多序列比对信息对 nsSNPs 进行表型预测。该软件所预测出的结果中分值 <0.05 被认为该突变对表型有害(deleterious), 而 0.05 则被认为无害(tolerated)^[18,19]。PolyPhen-2 软件(<http://genetics.bwh.harvard.edu/pph2/>)则是基于序列、系统发育和结构信息对蛋白氨基酸替换产生的结构、功能影响进行预测的软件。该软件所预测的分值越趋近 1, 该突变导致表型改变的可能性越大, 其结果可以定性描述为 benign、possibly damaging 及 probably damaging^[20]。

1.5 nsSNPs 疾病相关性分析

基于 SVM(Support vector machine)算法及 SIFT 算法, 利用 PhD-SNP 及 MutPred 软件进行 nsSNPs 疾病相关性分析。PhD-SNP 平台主要对目标 nsSNPs 数据集进行致病性分析, 预测结果可描述为致病(disease)及中性(neutral)^[21,22]。MutPred 软件可以预测突变与疾病的相关性, 还可以描述该突变致病的分子病因, 在结果中包含两个重要分值 g 和 p , 其中 g 主要表示该突变是否与疾病相关, p 则表示其某一结构或功能属性是否受影响。结合 g 和 p 分值, 该平台将预测结果评定为假设可行(Actionable hypotheses)、假设可信(Confident hypotheses)及假设非常可信(Very confident hypotheses)^[23]。

1.6 突变前后三维结构模型构建及突变分析

为了验证 *COCH* 基因的 nsSNPs 所造成的氨基酸改变对 cochlin 三维结构的影响, 利用 Swiss Model 软件(<http://swissmodel.expasy.org/interactive>)对致病 nsSNPs 进行突变前后 LCCL 区三维结构模型构建, 并利用 VMD 软件对突变前后三维结构模型进行分析^[24]。

2 结果与分析

2.1 nsSNPs 数据集收集及整理

从 dbSNP 数据库、HGMD 数据库以及 Deafness

Variation Database 数据库获得 *COCH* 基因 SNP 共 2037 个, 其中人类 *COCH* SNP 为 1132 个, 经过人工筛选共获得 73 个 nsSNPs(表 1), 其中 4 个来自 HGMD 数据库, 8 个来自 Deafness Variation Database 数据库, 其余 61 个来自 dbSNP 数据库。

2.2 nsSNPs 有害性预测结果

本研究选用 SIFT 及 PolyPhen-2 软件对数据库中所整合的 73 个 nsSNPs 进行表型预测。在 73 个 nsSNPs 中, 35 个 nsSNPs 被 SIFT 软件预测为 deleterious, 且预测分值为 0 的有 9 个。45 个 nsSNPs 被 PolyPhen-2 软件预测为 possibly damaging 及 probably damaging, 其中预测分值为 1 的有 6 个。在所筛选出的 73 个 nsSNPs 中, 3 个(K151E、S365L 和 R180X)无法用 SIFT 软件和 PolyPhen-2 软件共同预测, 因而最终 SIFT 软件和 PolyPhen-2 软件预测为有害的 nsSNPs 为 31 个(E25K、P51S、V66A、G75E、G87W、G88E、P89H、I109N、I109T、W117R、F121S、G135R、I176T、R180Q、V191M、A231V、G265E、V269L、I372T、G379R、I368N、R384H、R416C、Y424D、T444A、W503L、M512T、R525I、F527C、C542Y 和 C542F), 其中 G379R、G87W 和 F527C 预测结果中 SIFT 及 PolyPhen-2 软件预测分值分别为 0 和 1, 为高度有害(表 2)。

2.3 nsSNPs 疾病相关性预测结果

为了对所预测的 31 个 nsSNPs 进行疾病相关性分析, 本文通过 PhD-SNP 及 MutPred 软件分析, 分别获得了 26 和 15 个 nsSNPs 与疾病相关, 且所有被 MutPred 软件预测为假设可行、假设可信及假设非常可信的 nsSNPs 在 PhD-SNP 软件预测结果中均与疾病相关。即 PhD-SNP 及 MutPred 软件共同预测出 15 个与疾病相关联的 nsSNPs (P51S、G87W、I109N、I109T、W117R、F121S、I176T、R180Q、G265E、V269L、I368N、I372T、R416C、Y424D 和 F527C)(表 3)。

2.4 致病 nsSNPs 预测结果验证

为了验证筛选方法的可行性, 本文通过 PubMed、ScienceDirect、Web of Science 等数据库查阅, 发现被预测的 15 个与疾病相关的 nsSNPs 中共有 7 个 nsSNPs(P51S、G87W、I109N、I109T、W117R、F121S 和 F527C)已被报道^[25-31], 且均导致了 DFNA9 的发

生。对这 15 个 nsSNPs 改变后的 cochlin 蛋白结构域进行分析,发现 6 个 nsSNPs(P51S、G87W、I109N、I109T、W117R 和 F121S)位于该蛋白的 LCCL 结构域,而其他 9 个均位于 vWFA 结构域^[32](图 1)。研究

表 1 从数据库初步筛选得到的 *COCH* 基因的 nsSNPs 数据集及相关信息

SNP 序列号	染色体位置 (GRCh37)	突变前后碱基	突变前后 氨基酸	SNP 序列号	染色体位置 (GRCh37)	突变前后碱基	突变前后 氨基酸
rs28938175	31346846	151C>T	P51S	rs201578645	31358857	1513C>A	P505T
rs121908927	31346892	197T>C	V66A	rs367884240	31355287	1246C>T	R416C
rs121908928	31348040	263G>A	G88E	rs369882771	31349850	539G>A	R180Q
rs121908929	31348126	349T>C	W117R	rs370382942	31355174	1133T>C	V378A
rs121908930	31348103	326T>A	I109N	rs370525728	31355050	1009T>C	F337L
rs121908931	31348132	355G>A	A119T	rs371099832	31355417	1376G>A	G459D
rs121908932	31358969	1625G>A	C542Y	rs371248806	31355371	1330A>G	T444A
rs121908934	31358879	1535T>C	M512T	rs371802597	31344300	73G>A	E25K
rs267603975	31355176	1135G>A	G379R	rs375542487	31355144	1103T>A	I368N
rs1045644	31355096	1055C>G	T352S	rs376820760	31355036	995A>G	H332R
rs1801963	31358938	1594C>T	P532S	rs531508748	31354671	805G>T	V269L
rs17097468	31358897	1553A>G	E518G	rs536511561	31353821	692C>T	A231V
rs28362775	31354707	841G>A	D281N	rs540895576	31348048	271C>G	R91G
rs28362778	31355245	1204A>G	I402V	rs541167867	31349838	527T>C	I176T
rs28400035	31348658	403G>A	G135R	rs543399762	31349664	443G>A	R148Q
rs34622961	31349671:31349672	450_451insG	K151E	rs544358450	31355068	1027G>A	V343I
rs36109457	30885927:30885928	1092_1093insC	S365L	rs544874880	31344151	7G>C	A3P
rs61759484	31355311	1270T>G	Y424D	rs545418366	31355192	1151G>A	R384H
rs77074957	31354678	812T>G	V271G	rs550268175	31354758	892A>G	I298V
rs113782841	31344307	80C>A	A27D	rs559481815	31349849	538C>T	R180X
rs139503327	31355389	1348A>G	I450V	rs564563172	31358865	1521T>A	D507E
rs142073417	31348679	424C>T	H142Y	rs571941613	31355201	1160T>G	L387R
rs143098658	31346813	118G>A	D40N	rs576712224	31344295	68G>A	G23D
rs146115619	31346821	126G>C	R42S	rs577912568	31355181	1140T>A	D380E
rs147643069	31358918	1574G>T	R525I	CM062521	31348036	259G>T	G87W
rs148102471	31346888	193A>G	I65V	CM076118	31348103	326T>C	I109T
rs149072811	31358935	1591G>C	E531Q	CM108268	31348139	362T>C	F121S
rs149903169	31348043	266C>A	P89H	CM053812	31358969	1625G>T	C542F
rs186875564	31354680	814G>A	V272I		31346919	224G>A	G75E
rs188283330	31348049	272G>A	R91Q		31348070	293G>A	R98Q
rs200252533	31358852	1508G>T	W503L		31349882	571G>A	V191M
rs200522736	31344283	56C>T	P19L		31354660	794G>A	G265E
rs200549732	31349921	610G>A	V204M		31355156	1115T>C	I372T
rs200904881	31355225	1184A>G	K395R		31358885	1541C>T	S514F
rs200935305	31344166	22G>A	A8T		31358924	1580T>G	F527C
rs201294938	31353823	694A>G	I232V		31358972	1628G>A	R543K
rs201427920	31344164	20C>T	P7L				

证实, *COCH* 突变所导致的氨基酸改变几乎均位于 LCCL 和 vWFA 结构域, 本研究预测出的 6 个位于 LCCL 区域的 nsSNPs 均已被报道, 预测出的位于 vWFA 结构域的 9 个 nsSNPs 中, 除已被报道的 F527C 外, 其余 8 个 nsSNPs(I176T、R180Q、G265E、V269L、I368N、I372T、R416C 和 Y424D)可认为是 *COCH* 中具有潜在高风险致病的 nsSNPs。为了进一步说明这些 nsSNPs 造成疾病表型可能的原因, 本文借助

Swiss Model 平台对 cochlin 蛋白进行三维模型构建。由于该蛋白的 LCCL 结构域已经被精确测定(PDB ID: 1JBI), 因而本研究仅对该结构域内的 6 个 nsSNPs(P51S、G87W、I109N、I109T、W117R 和 F121S)进行突变前后氨基酸结构改变及结构域的变化分析, 结果表明 I109N 和 I109T 发生侧链结构改变, 而 P51S、G87W、W117R 和 F121S 突变均导致环状结构与链状结构互换(图 2)。

表 2 通过 SIFT 和 PolyPhen-2 软件共同预测为有害的 nsSNPs

SNP ID	染色体位置	碱基改变	氨基酸改变	正常人中突变频率(MAF)	SIFT 分值	SIFT 预测结果	PolyPhen-2 分值	PolyPhen-2 预测结果
rs28938175	31346846	151C>T	P51S	NA	0.000	Deleterious	0.973	Probably damaging
rs121908927	31346892	197T>C	V66A	0.0002/1	0.001	Deleterious	0.58	Possibly damaging
rs121908928	31348040	263G>A	G88E	NA	0.02	Deleterious	1	Probably damaging
rs121908929	31348126	349T>C	W117R	NA	0.003	Deleterious	0.999	Probably damaging
rs121908930	31348103	326T>A	I109N	NA	0.02	Deleterious	0.956	Possibly damaging
rs121908932	31358969	1625G>A	C542Y	NA	0.000	Deleterious	0.799	Possibly damaging
rs121908934	31358879	1535T>C	M512T	NA	0.003	Deleterious	0.818	Possibly damaging
rs267603975	31355176	1135G>A	G379R	NA	0.000	Deleterious	1	Probably damaging
rs28400035	31348658	403G>A	G135R	0.0004/2	0.028	Deleterious	0.596	Possibly damaging
rs61759484	31355311	1270T>G	Y424D	0.0002/1	0.002	Deleterious	0.959	Probably damaging
rs147643069	31358918	1574G>T	R525I	NA	0.023	Deleterious	0.999	Probably damaging
rs149903169	31348043	266C>A	P89H	NA	0.007	Deleterious	0.983	Probably damaging
rs200252533	31358852	1508G>T	W503L	NA	0.031	Deleterious	0.999	Probably damaging
rs367884240	31355287	1246C>T	R416C	NA	0.000	Deleterious	0.996	Probably damaging
rs369882771	31349850	539G>A	R180Q	NA	0.008	Deleterious	1	Probably damaging
rs371248806	31355371	1330A>G	T444A	NA	0.002	Deleterious	0.979	Probably damaging
rs371802597	31344300	73G>A	E25K	0.0004/2	0.023	Deleterious	0.476	Possibly damaging
rs375542487	31355144	1103T>A	I368N	NA	0.000	Deleterious	0.934	Possibly damaging
rs531508748	31354671	805G>T	V269L	0.0002/1	0.003	Deleterious	0.773	Possibly damaging
rs536511561	31353821	692C>T	A231V	0.0002/1	0.000	Deleterious	0.999	Probably damaging
rs541167867	31349838	527T>C	I176T	0.0002/1	0.001	Deleterious	0.999	Probably damaging
rs545418366	31355192	1151G>A	R384H	0.0002/1	0.013	Deleterious	0.019	Benign
CM062521	31348036	259G>T	G87W	NA	0.000	Deleterious	1	Probably damaging
CM076118	31348103	326T>C	I109T	NA	0.018	Deleterious	0.503	Possibly damaging
CM108268	31348139	362T>C	F121S	NA	0.001	Deleterious	0.999	Probably damaging
CM053812	31358969	1625G>T	C542F	NA	0.000	Deleterious	0.567	Possibly damaging
	31346919	224G>A	G75E	NA	0.002	Deleterious	1	Probably damaging
	31349882	571G>A	V191M	NA	0.038	Deleterious	0.996	Probably damaging
	31354660	794G>A	G265E	NA	0.002	Deleterious	0.996	Probably damaging
	31355156	1115T>C	I372T	NA	0.001	Deleterious	0.978	Probably damaging
	31358924	1580T>G	F527C	NA	0.000	Deleterious	1	Probably damaging

注: 黑色加粗字体表示 SIFT 或 PolyPhen-2 预测结果为高度有害的 nsSNPs, 即 SIFT 预测结果分值为 0.000, PolyPhen-2 预测结果分值为 1。NA 表示暂无相关数据。

表 3 PhD-SNP 和 MutPred 软件共同预测为疾病相关的 nsSNPs

SNP 序列号	氨基酸改变	PhD-SNP 预测结果	Mutpred 分值(g)	Mutpred 分值(p)	Mutpred 分子改变预测	可信度
rs28938175	P51S	Disease	0.953	0.0228	Loss of sheet	Confident hypotheses
rs121908929	W117R	Disease	0.825	0.0486	Gain of catalytic residue at W117	Confident hypotheses
rs121908930	I109N	Disease	0.839	0.0126	Gain of disorder	Confident hypotheses
rs61759484	Y424D	Disease	0.824	0.014	Gain of disorder	Confident hypotheses
rs367884240	R416C	Disease	0.748	0.0418	Loss of disorder	Actionable hypotheses
rs369882771	R180Q	Disease	0.605	0.0233	Loss of catalytic residue at R180	Actionable hypotheses
rs375542487	I368N	Disease	0.884	0.0466	Loss of stability	Confident hypotheses
rs531508748	V269L	Disease	0.683	0.0405	Loss of methylation at K268	Actionable hypotheses
rs541167867	I176T	Disease	0.819	0.0065	Loss of stability	Very confident hypotheses
CM062521	G87W	Disease	0.944	0.0127	Loss of glycosylation at S86	Confident hypotheses
CM076118	I109T	Disease	0.867	0.0184	Gain of disorder	Confident hypotheses
CM108268	F121S	Disease	0.762	0.0079	Gain of disorder	Actionable hypotheses
	G265E	Disease	0.506	0.0456	Gain of solvent accessibility	Actionable hypotheses
	I372T	Disease	0.871	0.0031	Loss of stability	Very confident hypotheses
	F527C	Disease	0.899	0.0331	Loss of catalytic residue at F527	Confident hypotheses

注：黑色加粗字体表示经 PhD-SNP 和 MutPred 软件预测后为疾病相关且已有文献报道的 nsSNPs，其余为本研究所预测出的 nsSNPs。

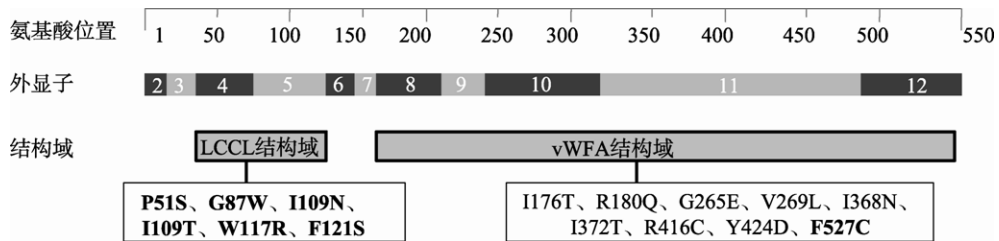


图 1 15 个 nsSNPs 在 cochlin 蛋白结构域中分布情况
黑色加粗部分表示该突变已经被报道。

3 讨 论

自 1994 年 Robertson 等^[7]发现了新的耳蜗基因—*COCH* 基因以来,在已建立的人类 *COCH* 基因 cDNA 文库基础上,通过表达序列与遗传图谱比较,证实 *COCH* 基因位于常染色体 14q12-q13 上,且定位在 DFNA9 候选区内。迄今已有 1132 个人类 *COCH* 基因 SNPs 被发现,其中仅有 16 个 SNPs 的变异与耳聋相关,大量 SNP 对疾病表型的影响有待进一步研究。

人类基因组图谱的完成和 HapMap 计划的实施,使通过 SNP 位点追踪复杂疾病相关基因突变成为可能。研究表明,占人类基因组变异的 90% 的 SNP 可以在 DNA、RNA 和蛋白质水平影响基因功能,其中最显著的作用是使变异后的基因产物失去部分、全

部功能或产生新的功能。关于 nsSNPs 与疾病相关性的研究可以追溯到 2001 年,其研究方法是基于致病 nsSNP 位点对蛋白质结构影响的共性的基础上,制定一系列的规则,用来预测未知表型的 nsSNPs^[33]。自此以后,通过多途径研究,在已知的人类遗传病中,发现涉及 nsSNPs 位点的遗传病已超半数^[34]。

生物信息学的研究与应用已广泛渗透到生物学、医学领域的实践中,对未知表型的 nsSNPs 的分析不仅仅局限于简单的预测,而是采用一系列离散或连续的特征,如 nsSNPs 位点的氨基酸物理和化学特征、nsSNP 位点和附近区域的序列保守性、蛋白质结构特征和进化特征等来构造预测模型,如概率模型、决策树、WEKA 软件包、支持向量机和随机森林等。本研究结合相关文献,并通过上述多种方法来预测

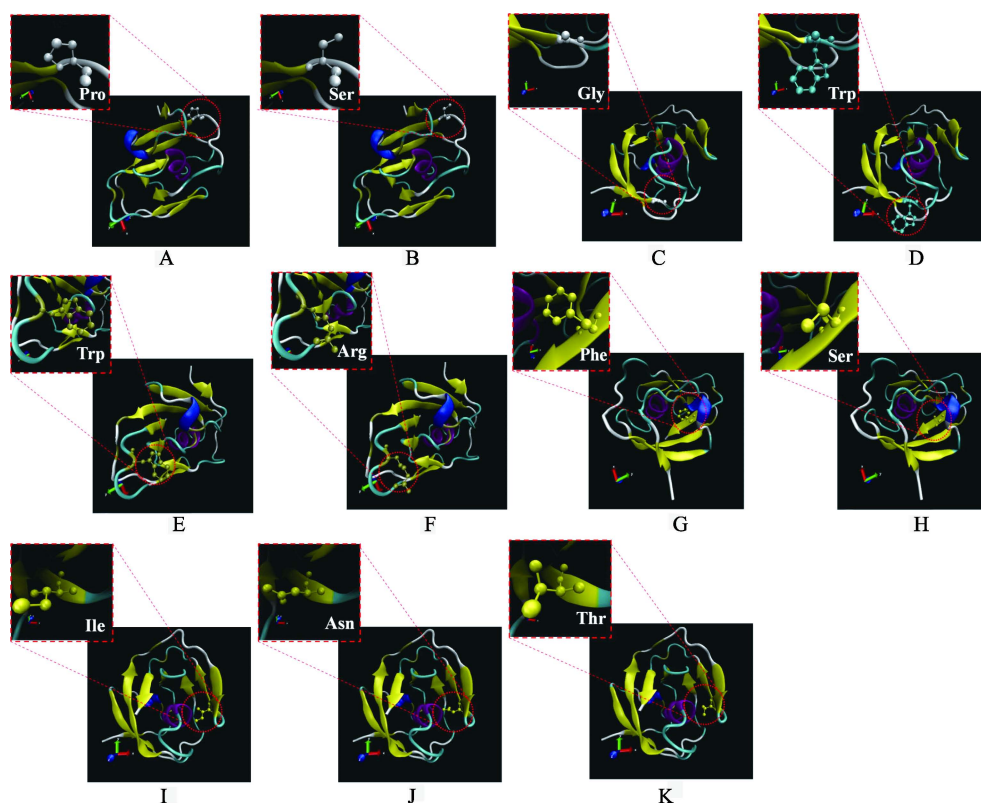


图 2 6 个已报道的位于 LCCL 结构域的致病 nsSNPs 突变前后结构变化

A、B 为 P51S 突变，C、D 为 G87W 突变，E、F 为 W117R 突变，G、H 为 W117R 突变，I、J 为 I109N 突变，I、K 为 I109T 突变。A、C、E、G、I 为突变前结构。

COCH 突变基因型与表型之间的关系，经过分级筛选和准确性验证，从中发掘出 8 个高风险致病的 nsSNPs，为 *COCH* 基因突变致聋的筛查提供了理论依据。在对 P51S、G87W、I109N、I109T、W117R 和 F121S 共 6 个已知的致病突变进行了 LCCL 区域三维结构模拟后，发现绝大多数位点氨基酸均发生了环状结构与链状结构的互换，这一结果提示，突变所导致的氨基酸环状结构与链状结构的改变可能会导致蛋白质功能的改变。在对筛查结果中 15 个 nsSNPs(P51S、G87W、I109N、I109T、W117R、F121S、I176T、R180Q、265E、V269L、I368N、I372T、R416C、Y424D 和 F527C)进行结构域分布分析时，发现 8 个新预测出的 nsSNPs(I176T、R180Q、G265E、V269L、I368N、I372T、R416C 和 Y424D)均位于 vWFA 结构域，这一结果提示在对 *COCH* 的研究中，除了高度保守的 LCCL 结构域易发生致病突变外，vWFA 结构域的突变也可能会导致相应疾病的发生。本研究所采用的筛选方法也对其他基因突变致病性分析具有一定的参考价值。

参考文献

- [1] Looi LM, Ganten D, McGrath PF, Gross M, Griffin GE. Hearing loss: a global health issue. *Lancet*, 2015, 385(9972): 943–944. [DOI]
- [2] Cohen MM, Gorlin RJ. Epidemiology, etiology and genetic patterns. In: Gorlin RJ, Toriello HV, Cohen MM. *Hereditary Hearing Loss and Its Syndromes*. New York: Oxford University Press, 1995: 457. [DOI]
- [3] Tekin M, Arnos KS, Pandya A. Advances in hereditary deafness. *Lancet*, 2001, 358(9287): 1082–1090. [DOI]
- [4] Simon A, Xi L, Xue ZL. Genetics of Hearing and Deafness. *Anat Rec*, 2012, 295(11): 1812–1829. [DOI]
- [5] Van Camp G, Smith R. Hereditary hearing loss homepage. <http://hereditaryhearingloss.org/>. [DOI]
- [6] Fransen E, Van Camp G. The COCH gene: a frequent cause of hearing impairment and vestibular dysfunction? *Br J Audio*, 1999, 33(5): 297–302. [DOI]
- [7] Robertson NG, Khetarpal U, Gutiérrez-Espeleta GA, Bieber FR, Morton CC. Isolation of novel and known genes from a human fetal cochlear cDNA library using subtrac-

- tive hybridization and differential screening. *Genomics*, 1994, 23(1): 42–50. [DOI]
- [8] Kamaraj B, Purohit R. *In silico* screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *BioMed Res Int*, 2013, 2013: 697051. [DOI]
- [9] Kumar A, Rajendran V, Sethumadhavan R, Purohit R. Computational investigation of cancer-associated molecular mechanism in Aurora A (S155R) mutation. *Cell Biochem Biophys*, 2013, 66(3): 787–796. [DOI]
- [10] Wang LL, Li Y, Zhou SF. Prediction of deleterious non-synonymous single nucleotide polymorphisms of genes related to ethanol-induced toxicity. *Toxicol Letter*, 2009, 187(2): 99–114. [DOI]
- [11] Kumar A, Purohit R. Computational investigation of pathogenic nsSNPs in CEP63 protein. *Gene*, 2012, 503(1): 75–82. [DOI]
- [12] Carvalho MA, Marsillac SM, Karchin R, Manoukian S, Grist S, Swaby RF, Urmenyi TP, Rondinelli E, Silva R, Gayol L, Baumbach L, Sutphen R, Pickard-Brzosowicz JL, Nathanson KL, Sali A, Goldgar D, Couch FJ, Radice P, Monteiro ANA. Determination of cancer risk associated with germ line *brca1* missense variants by functional analysis. *Cancer Res*, 2007, 67(4): 1494–1501. [DOI]
- [13] Goldgar DF, Easton DF, Deffenbaugh AM, Monteiro ANA, Tavtigian SV, Couch FJ, Breast Cancer Information Core (BIC) Steering Committee. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to *Brca1* and *Brca2*. *Am J Hum Genet*, 2004, 75(4): 535–544. [DOI]
- [14] Karchin R. Next generation tools for the annotation of human SNPs. *Brief Bioinform*, 2009, 10(1): 35–52. [DOI]
- [15] Wang LL, Yang AK, Li Y, Liu JP, Zhou SF. Phenotype prediction of deleterious nonsynonymous single nucleotide polymorphisms in human alcohol metabolism-related genes: a bioinformatics study. *Alcohol*, 2010, 44(5): 425–438. [DOI]
- [16] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*, 2001, 29(1): 308–311. [DOI]
- [17] Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN. The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*, 2014, 133(1): 1–9. [DOI]
- [18] Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. Sift web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*, 2012, 40(Web Server issue): W452–W457. [DOI]
- [19] Ng PC, Henikoff S. Sift: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 2003, 31(13): 3812–3814. [DOI]
- [20] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*, 2010, 7(4): 248–249. [DOI]
- [21] Capriotti E, Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. *Bioinformatics*, 2005, 21(Suppl. 2): ii54–ii58. [DOI]
- [22] Capriotti E, Calabrese R, Casadio R. Predicting the insurgenge of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, 2006, 22(22): 2729–2734. [DOI]
- [23] Li B, Krishnan VG, Mort ME, Xin FX, Kamati KK, Cooper DN, Mooney SD, Radivojac P. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, 2009, 25(21): 2744–2750. [DOI]
- [24] Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res*, 2003, 31(13): 3381–3385. [DOI]
- [25] de Kok YJM, Bom SJH, Brunt TM, Kemperman MH, van Beusekom E, van der Velde-Visser SD, Robertson NG, Morton CC, Huygen PLM, Verhagen WIM, Brunner HG, Cremers CWRJ, Cremers FPM. A Pro51Ser mutation in the *COCH* gene is associated with late onset autosomal dominant progressive sensorineural hearing loss with vestibular defects. *Hum Mol Genet*, 1999, 8(2): 361–366. [DOI]
- [26] Collin RWJ, Pauw RJ, Schoots J, Huygen PLM, Hoefsloot LH, Cremers CWRJ, Kremer H. Identification of a novel *COCH* mutation, G87W, causing autosomal dominant hearing impairment (DFNA9). *Am J Med Genet A*, 2006, 140(16): 1791–1794. [DOI]
- [27] Robertson NG, Lu LY, Heller S, Merchant SN, Eavey RD, McKenna M, Nadol JB Jr, Miyamoto RT, Linthicum FH, Lubianca Neto JF, Hudspeth AJ, Seidman CE, Morton CC, Seidman JG. Mutations in a novel cochlear gene cause DFNA9, a human nonsyndromic deafness with vestibular dysfunction. *Nat Genet*, 1998, 20(3): 299–303. [DOI]
- [28] Kamarinos M, McGill J, Lynch M, Dahl H. Identification of a novel *COCH* mutation, I109N, highlights the similar clinical features observed in DFNA9 families. *Hum*

- Mutat*, 2001, 17(4): 351. [DOI]
- [29] Pauw RJ, Huygen PLM, Collin RWJ, Cruysberg JRM, Hoefsloot LH, Kremer H, Cremers CWRJ. Phenotype description of a novel DFNA9/COCH mutation, I109T. *Ann Otol Rhinol Laryngol*, 2007, 116(5): 349–357. [DOI]
- [30] Hildebrand MS, Gandolfo L, Shearer AE, Webster JA, Jensen M, Kimberling WJ, Stephan D, Huygen PLM, Smith RJH, Bahlo M. A novel mutation in *COCH*—implications for genotype-phenotype correlations in DFNA9 hearing loss. *Laryngoscope*, 2010, 120(12): 2489–2493. [DOI]
- [31] Cho HJ, Park HJ, Trexler M, Venselaar H, Lee KY, Robertson NG, Baek JI, Kang BS, Morton CC, Vriend G, Patthy L, Kim UK. A novel *COCH* mutation associated with autosomal dominant nonsyndromic hearing loss disrupts the structural stability of the vWFA2 domain. *J Mol Med*, 2012, 90(11): 1321–1331. [DOI]
- [32] Gallant E, Francey L, Fetting H, Kaur M, Hakonarson H, Clark D, Devoto M, Krantz ID. Novel *COCH* mutation in a family with autosomal dominant late onset sensorineural hearing impairment and tinnitus. *Am J Otolaryngol*, 2013, 34(3): 230–235. [DOI]
- [33] Wang Z, Moulton J. SNPs, protein structure, and disease. *Hum Mutat*, 2001, 17(4): 263–270. [DOI]
- [34] Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, Thomas N, Cooper DN. Human gene mutation database—a biomedical information and research resource. *Hum Mutat*, 2000, 15(1): 45–51. [DOI]

(责任编辑: 袁慧军)

• 综合信息 •

《遗传》获中国科协学术质量提升项目资助

为了进一步发挥学会办刊的学术和专业优势, 打造一批在本学科和专业领域内有较强影响力和专业辐射力的中文精品科技期刊, 更好地发挥中文学术期刊在引领自主创新、服务科技发展中的作用, 中国科协发布了《中国科协精品科技期刊工程项目实施方案(2015~2017)》。通过项目申报、资格审查及专家评审, 中国科学院遗传与发育生物学所和中国遗传学会主办的《遗传》杂志获得学术质量提升项目资助(3年资助45万元)。

这是《遗传》首次获得中国科协精品科技期刊工程项目的资助, 对于进一步提升《遗传》的期刊质量与学术影响力将起到十分重要的推动作用。