

乳腺癌易感蛋白 1 序列结构特征及致病性突变的分析

范 燧¹, 郁 芸², 韩新焕^{2*}

(¹南京医科大学临床医学系, ²基础医学院数学与计算机教研室, 江苏 南京 210029)

[摘要] 目的:利用生物信息学方法分析乳腺癌易感基因 1(BRCA1)序列结构特征,并预测 BRCA1 功能编码区突变与致病性遗传效应的关系。方法:采用 Maximum Likelihood、ClustalW、SMART 和 Selecton 等在线生物信息学分析工具,对 BRCA1 进行分子系统发育、保守性、选择压力、结构域、三维结构和错义结构的分析。结果:分析获得 195 个固定残基位点(10.5%)和 393 个保守位点(21.1%),其分布是非随机性的;发现人 BRCA1 序列中的保守结构域 BRCT 的三维结构与其他物种存在着较大的差异,表明 BRCT 结构域在不同物种中具有不同的生物学功能;关联性分析证实发生在保守位点的突变致病性高。结论:基于生物信息学的 BRCA1 序列结构分析,能充分利用相关数据的资源,加深对 BRCA1 基因变异与肿瘤发生的相关性的认识。

[关键词] 乳腺癌易感蛋白 1; 序列分析; 三维同源模型; 生物信息学

[中图分类号] R737.9

[文献标识码] A

[文章编号] 1007-4368(2012)06-805-06

Analysis on the sequence and molecular structure of breast cancer susceptibility protein 1 and its mutagenesis to pathogenicity

FAN Yi¹, YU Yun², HAN Xin-huan^{2*}

(¹Department of Clinical Medicine, ²Department of Mathematics & Computer Science, NJMU, Nanjing 210029, China)

[Abstract] **Objective:** To analyze the structure characteristics of the breast cancer susceptibility protein 1(BRCA1) gene sequence by bioinformatics methods and predict the relationship between mutations of functional coding regions in BRCA1 and the pathogenic genetic effects. **Methods:** The molecular phylogeny, conservation, selection pressure, domains, three-dimensional structure and mis-sense of BRCA1 were analyzed by online Bioinformatics analysis tools such as Maximum Likelihood, ClustalW, SMART, SELECTON and so on. **Results:** Total 195 fixed-residue sites(10.5%)and 393 conservative sites(21.1%), of which distribution is non-randomness, were obtained. A big difference in the BRCT domain was founded between human and other species, which indicated the role of BRCT domain in BRCA1 gene of various species is difference. The high pathogenic mutation rate in conservative sites was confirmed by correlation analysis. **Conclusion:** Taking full advantage of relevant resources, the BRCA1 sequence structure analysis based on bioinformatics methods can deepen our understanding of the correlation of BRCA1 gene mutation and tumor.

[Key words] breast cancer susceptibility protein 1; sequence analysis; three-dimensional homology model; bioinformatics

[Acta Univ Med Nanjing, 2012, 32(6): 805-810]

人类乳腺癌易感蛋白 1(breast cancer susceptibility protein 1, BRCA1) 是与家族遗传性乳腺癌和卵巢癌发生密切相关的抑癌基因^[1]。研究显示, 5%~10%的乳腺癌和卵巢癌患者表现为家族遗传性, 其中 BRCA1 基因突变是上述两种肿瘤发生的重要遗传学事件, 40%~50%的遗传性乳腺癌与 BRCA1 突

变有关, 而在乳腺癌和卵巢癌都高发的家族中, 80%的患者 BRCA1 发生突变。由 BRCA1 突变基因引起的乳腺癌, 其 70 岁以上血缘亲属的发病率高达 80%~90%, BRCA1 突变基因携带者 40 岁以前患乳腺癌和卵巢癌的风险远高于一般人群^[2-3]。

BRCA1 基因全长约 100 kb, 含 24 个外显子, 编码由 1 863 个氨基酸构成的分子量为 220 000 的多功能磷酸化核蛋白, 该蛋白氨基末端含有一锌指结构域, 其茎环结构是介导蛋白与蛋白以及蛋白与

[基金项目] 江苏省大学生科技创新基金资助

*通讯作者, E-mail: hxinhuan@163.com

DNA 相互作用的功能域; 羧基末端含有 2 个串联 BRCT(BRCA1 carboxyl terminus, BRCT) 结构域, 多数与肿瘤相关的点突变均发生在这两个保守功能区内^[4]。由于 BRCA1 基因编码区长、突变种类多、无明显的固定热点, 并有新的突变被不断发现, 这给实验检测带来了困难。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, 将为后续的生物实验研究起到引导性的作用。因此, 本文利用生物信息学的方法, 通过对 BRCA1 家族蛋白的氨基酸序列多重比对进行保守性分析, 蛋白质三维同源模型的构建进行结构比对分析并结合 BRCA1 突变数据库与相关的研究报告, 探讨人类 BRCA1 蛋白与乳腺癌相关的致病性突变位点和其氨基酸序列保守位点的关系, 期望为 BRCA1 基因变异与其功能的相关性研究提供有益的资料。

1 资料和方法

1.1 资料

以美国国立生物技术信息中心(NCBI)中人类 BRCA1 蛋白的氨基酸序列为查询序列, 通过数据库检索选取同源性较高的 BRCA1 家族的蛋白序列。参考 Carvalho 等^[5]通过转录分析、结构分析和实验数据分析获得的 BRCA1 突变分型结果。

1.2 方法

1.2.1 BRCA1 进化分析

用 Mega5.0 软件对 11 个物种进行分子进化树的构建, 并获取进化树中各结点处的祖先序列。本研究中进化树的构建采用了 Maximum Likelihood 法^[6]。祖

先序列的获取采用了经验贝叶斯分析法。

1.2.2 BRCA1 多序列比对及保守性分析

应用 ClustalW 在线软件(<http://www.ebi.ac.uk/Tools/msa/clustalw2/>)对获得的各结点祖先序列与 11 个物种 BRCA1 蛋白序列进行多重序列比对分析, 先将多个序列两两比对, 构建距离矩阵, 再根据距离矩阵计算产生系统进化指导树, 软件中各参数值均采用默认值。

1.2.3 BRCA1 结构域分析

应用 SMART 在线软件(<http://smart.embl.de>)对 BRCA1 蛋白序列进行结构域识别^[7]; 应用 Swiss-Model 在线软件(<http://swissmodel.expasy.org/>)对 11 个物种的 BRCT 结构域进行三维同源模型构建, 并用 VMD 软件进行空间结构比对分析^[8-9]。

1.2.4 BRCA1 选择压力分析

基于密码子置换模型计算得到的非同义-同义替换速率比值(dN/dS)常被用来推测作用于蛋白质的选择压力。利用 Selecton 2.4 在线软件(<http://selecton.tau.ac.il/>)计算人类 BRCA1 氨基酸序列各位点 dN/dS 值, 进行选择压力分析。分析选用 M8, beta + w ≥ 1 模型^[10]。

1.3 验证性比较分析

参考已报道的 BRCA1 突变的研究结果, 结合生物信息学预测进行分析。

2 结果

2.1 11 个物种及其 BRCA1 蛋白序列信息

11 个物种包括了灵长目、食肉目、偶蹄目和啮齿目, 序列信息见表 1。

表 1 11 个物种 BRCA1 蛋白的序列信息

Table 1 BRCA1 protein sequence information from 11 species

编号	物种名称	序列长度	相似度(%)	BRCT1	BRCT2	GenBank 号
1	人类(<i>homo sapiens</i>)	1 863	100.0	1 644~1 726	1 758~1 845	AEQ98814
2	黑猩猩(<i>Pan troglodytes</i>)	1 863	99.5	1 644~1 726	1 758~1 845	AAR04849
3	红毛猩猩(<i>Pongo pygmaeus</i>)	1 863	98.3	1 644~1 726	1 758~1 845	AAT44834
4	大猩猩(<i>Gorilla Gorilla</i>)	1 863	98.0	1 644~1 726	1 758~1 845	AAT44835
5	猕猴(<i>Macaca mulatta</i>)	1 863	94.0	1 644~1 726	1 758~1 845	AAT44833
6	家犬(<i>Canis lupus familiaris</i>)	1 878	84.2	1 647~1 729	1 766~1 853	AAC48663
7	牛(<i>Bos taurus</i>)	1 849	79.3	1 638~1 719	1 751~1 838	AAL76094
8	猪(<i>Porcus</i>)	1 863	80.5	1 640~1 722	1 754~1 841	BAF62296
9	家鼠(<i>Mus musculus</i>)	1 812	68.1	1 587~1 669	1 700~1 787	CAM21026
10	大鼠(<i>Rattus norvegicus</i>)	1 817	66.8	1 590~1 672	1 703~1 790	AAC36493
11	负鼠(<i>Opossum</i>)	1 844	53.2	1 619~1 707	1 739~1 826	AAX92675

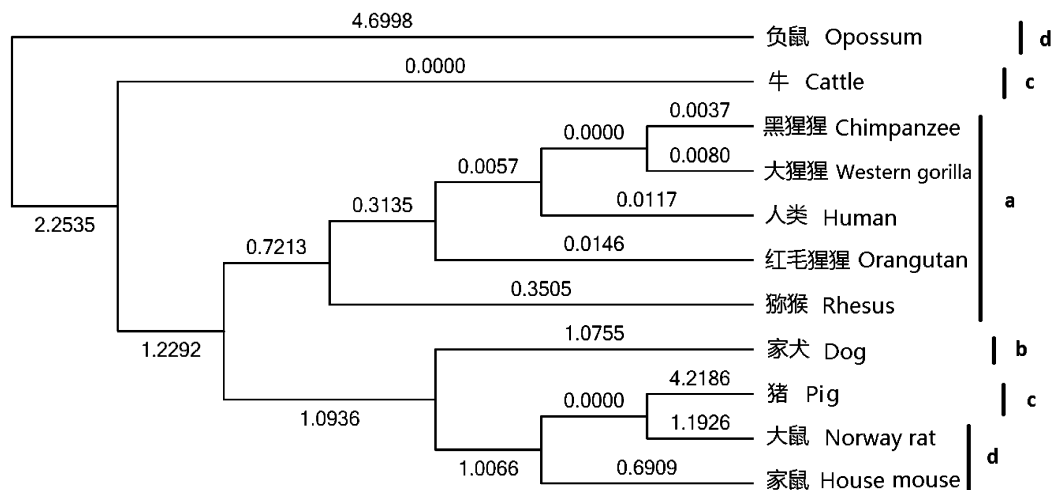
2.2 BRCA1 生物信息学特征

2.2.1 分子进化分析

由图 1 进化树可以看出, 相近物种来源的 BRCA1 蛋白往往具有相近的亲缘关系。例如, 与家鼠

BRCA1 蛋白(BRCA1_House mouse)具有较近亲缘关系的是大鼠(BRCA1_Norway rat);与人类 BRCA1 蛋白最为亲近的是黑猩猩 (BRCA1_Chimpanzee)与大猩猩(BRCA1_Western gorilla);此外,人类 BRCA1

蛋白与鼠类 BRCA1 蛋白进化关系较远,提示在用鼠类作为动物模型来研究人类 BRCA1 功能时,应考虑物种之间的差异性。



a:灵长目;b:食肉目;c 为偶蹄目;d:啮齿目;数据为各分支长度。

图 1 BRCA1 家族蛋白的分子进化树

Figure 1 The phylogenetic tree of BRCA1 proteins

2.2.2 BRCA1 保守区分析

将物种 BRCA1 与其祖先进行多序列比对 (图 2),可以减少物种进化过程中因生物多样性而产生的氨基酸替换的干扰,从而能更好地鉴别重要的功能性位点。用这种方法,研究得到了 195 个固定残基位点(10.5%),393 个保守位点(21.1%)。这些位点并非随机分布,在某些区域有较高的保守性,如 75~112 段氨基酸残基保守度达 76.3%(29/38)。

2.2.3 结构域分析

对 11 个物种 BRCT 进行三维结构的比对分析发现,人、黑猩猩、猕猴、家犬、牛、猪和负鼠的 BRCT 三维结构几乎完全重叠(图 3)。家鼠和大鼠的 BRCT 三维结构重合(图 4),但是与人有较大的差异,这一结果与进化树中人与鼠类的进化关系较远相符,可进一步推测人 BRCA1 的结构特征与家鼠和大鼠是不相同的。BRCA1 中的保守结构域 BRCT 的三维结构在亲缘关系较远的物种间存在着较大的差异,而 BRCT 是 BRCA1 中的重要功能结构域,差异的存在暗示 BRCT 在不同物种中发挥着不同的功能,这为进一步研究 BRCT 的功能提供了依据。在研究中发现负鼠的 BRCA1-like 蛋白在进化树中与人类 BRCA1 距离较远,但是其 BRCT 三维结构却与人类重合,这表明 BRCA1-like 蛋白功能主要体现在 BRCT

结构域的功能,很好地解释了 BRCA1-like 蛋白不属于 BRCA1 蛋白家系但却具有 BRCA1 蛋白相关功能的问题。

2.2.4 选择压力分析

选择压力分析检测出 3 个正选择位点 (图 5),分别是 926T、1541E 和 1804V,其在上述多序列比对结果中均为非保守位点。在正选择压力下,基因的非同义替代速率远高于同义替代速率。正选择作为分子进化的动力,保留了有利的非同义突变,从而加速同源蛋白分化,因此可以推断发生在正选择位点上的错义突变会改变蛋白质的功能。此外,全序列以紫色和白色位点居多,即序列中净化选择位点居多。在进化选择压力下,基因的非同义替代速率远低于同义替代速率,基因处于较高的保守状态。

2.3 验证性比较分析

参考 Carvalho 等^[5]的研究工作,将 24 种已分型的错义突变位点与上述的多重序列比对的结果相对应 (表 2),发现有害型突变位点的保守性为 60.0%,中立型突变位点的保守性为 21.4%,并且中立型突变位点中未出现固定残基。推测该基因错义突变是否会影响基因的功能与该突变位点是否位于保守区有着密切的关系。

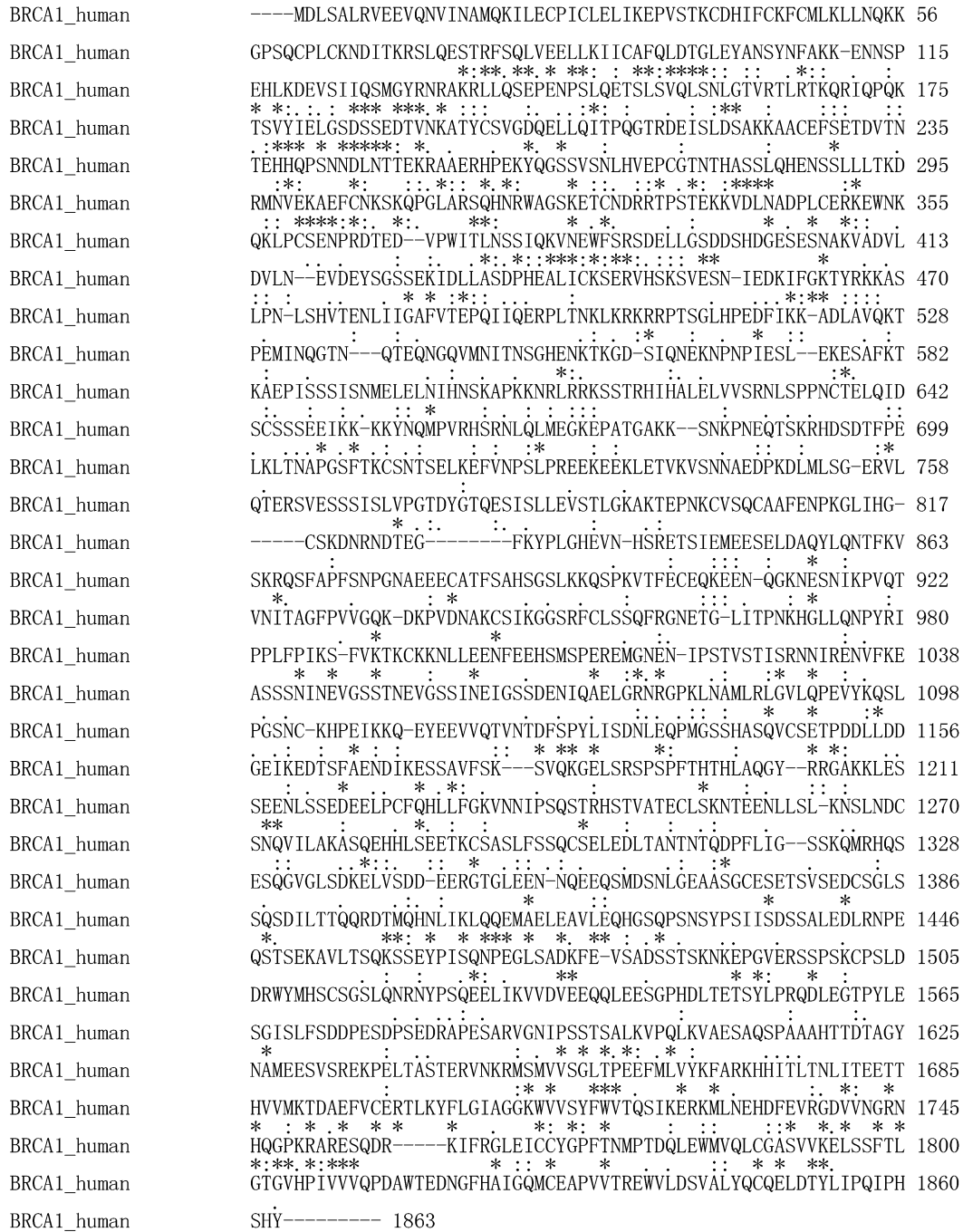


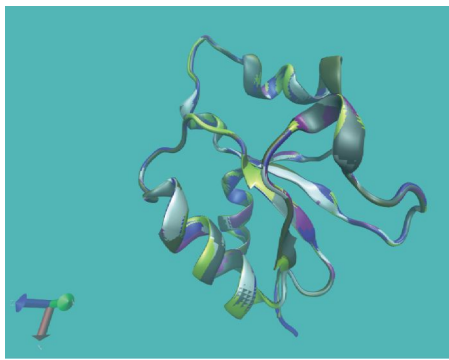
图 2 11 个物种 BRCA1 与进化树中 9 个结点处的祖先序列比对结果

Figure 2 BRCA1 amino acid sequence alignment result including 11 species and ancestors of the tree's nodes

3 讨论

近十年来,一些学者对乳腺癌 BRCA1 基因的研究发现,BRCA1 可发生多形式、多位点的基因突变,突变可分布于整个编码区域,除 Vallon-Christersson 等^[11]报道的斯堪得那维亚妇女乳腺癌家族中普遍存在错义突变 A1669S、C1697R、R1699W、R1699Q、A1708E、S1715R、G1738E 和终止突变 W1837X 外,未见有广泛意义的突变热点的报道。

BRCA1 基因在不同研究及人群中突变的分布以及各自所占比例的报道差异很大,瑞典、北欧犹太妇女中 BRCA1 的突变率较高^[12-13]。在亚洲人群中,有关 BRCA1 的研究结果显示,其突变率明显低于西方人群^[14]。我国的研究起步较晚,结果并不一致,同时还未有中国人乳腺癌患者 BRCA1 突变谱及“始祖突变”的研究报道。BRCA1 基因突变与 BRCT 结构域的缺失有关,其突变形式多样,有 30%~40% 的突变,特别是一大类的错义突变和核苷酸多态性与



红色:人类;紫色:黑猩猩;灰色:猕猴;白色:家犬;蓝色:牛;黄色:猪;赭色:负鼠。
图 3 人、黑猩猩、猕猴、家犬、牛、猪和负鼠 BRCA1 蛋白的 BRCT 结构域三维结构图

Figure 3 The three-dimensional structures of BRCT domain in human, chimpanzee, rhesus, dog, cattle, pig and opossum

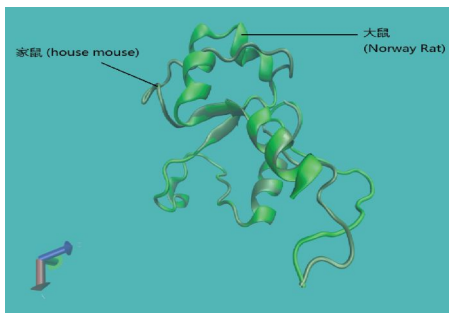
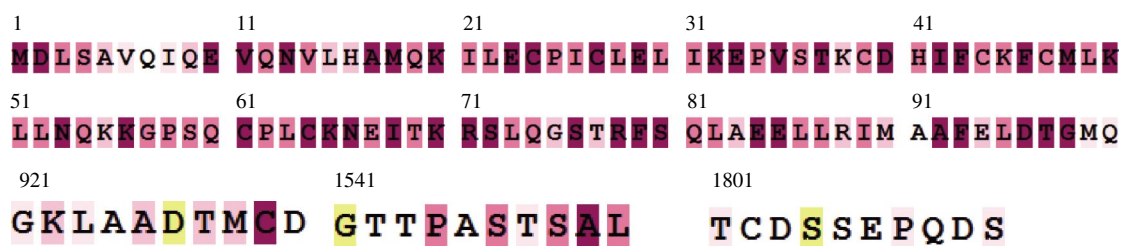


图 4 家鼠和大鼠 BRCA1 蛋白的 BRCT 结构域三维结构重叠图

Figure 4 The three-dimensional structure alignment of BRCA1 protein in house mouse and Norway rat

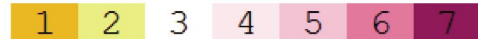
乳腺癌发生的关系尚不清楚。因此,从多层次、多方向开展对中国人乳腺癌发生与 BRCA1 突变效应的相关性研究具有重要的意义。

随着乳腺癌发病率的增加, BRCA1 作为迄今为止与乳腺癌发生关系最密切的易感基因之一越来越被关注,但目前有关 BRCA1 的抑癌机制、致病的错义突变的界定以及如何对致病基因突变携带者判定等问题,仍待深入研究。伴随着高通量测序技术的飞速发展,现代生物信息学在生命科学中的作用已毋庸置疑。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或在分子生物学实验基础上结合生物信息学研究,将极大地拓宽疾病基因组学整体的研究水平。先前, Melissa 等^[15]曾对 57 种真哺乳亚纲的哺乳动物的 BRCA1 基因进行贝叶斯系统发生分析,并按保守程度对 BRCA1 的第 11 外显子区的氨基酸位点进行分类,预测该基因第 11 外显子上潜在的重要功能位点和极有可能干扰基因功能的错义突变位点,研究结果发现 BRCA1 保守位点上的错义突变比非保守位点上的错义突变更能影响基因的功能。本研究以 BRCA1 全基因编码区序列作为待处理数据进行分子进化分析,得到 BRCA1 基因的错义突变与其保守性的关系,是对 BRCA1 全基因序列保守性分析的进一步补充与完善。本文分析获得 195 个固定残基位点 (10.5%) 和 393 个保守位点 (21.1%), 其分布是非随机性的,也提示该基因区域的某些位点的突变是与致病性相关,关联性分析证实发生在保守位点的突



Legend:

The selection scale:



Positive selection

Purifying selection

黄色:正选择位点;紫色:净化选择位点;深黄和深紫代表有显著的统计学意义。

图 5 BRCA1 部分正选择压力检测结果

Figure 5 The partial result of positive selection pressure test of BRCA1

变致病性高。通过多重比对分析对 BRCA1 的重要功能结构域 BRCT 的三维结构进行比较,证实亲缘关系较远的物种间存在着较大的差异,这也表明 BRCT 结构域在不同物种中具有不同的生物学

功能。

本研究表明在对 BRCA1 基因突变与疾病关联性分析时,可以通过同源搜索和多序列比对分析,寻找靶基因的保守区,分析突变位点是否位于保守

表2 24种错义突变位点的保守性分析结果

Table 2 The conservation analysis result of 24 mutation sites

编号	突变	分型	比对结果	编号	突变	分型	比对结果
1	H1402Y	中立型	/	13	T1720A	中立型	/
2	R1443G	中立型	/	14	V1804D	中立型	/
3	S1512I	中立型	/	15	R1699W	有害型	*
4	V1534M	中立型	/	16	R1699Q	有害型	*
5	D1546N	中立型	/	17	G1706E	有害型	/
6	T1561I	中立型	:	18	A1708E	有害型	.
7	L1564P	中立型	/	19	S1715R	有害型	*
8	S1613G	中立型	.	20	G1738R	有害型	*
9	P1614S	中立型	.	21	L1764P	有害型	/
10	M1628T	中立型	/	22	I1766S	有害型	:
11	M1652I	中立型	/	23	M1775R	有害型	/
12	A1669S	中立型	/	24	G1788V	有害型	/

/:不保守位点; *:固定位点; :和.:保守替换位点,基于 Gonnet PAM250 矩阵,前者保守程度强($P > 0.5$),后者保守程度弱($P \leq 0.5$)。

区来验证实验结果的准确性。该研究为 BRCA1 基因突变与乳腺癌分子病因学的实验研究提供了前期的工作基础。

[参考文献]

- [1] Kotsopoulos J, Lubinski J, Salmena L, et al. Breast feeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers [J]. *Breast Cancer Res*, 2012, 14(2): R42
- [2] Vaz FH, Machado PM, Brandão RD, et al. Amilial breast/ovarian cancer and CA1/2 genetic screening; the role immunohisto chemistry as an additional method in the selection opatients[J]. *JHC Express*, 2007, 55(11):1105-1113
- [3] Bergman A, Flodin A, Engwall Y, et al. A high frequency of germline BACA1 /BRCA2 mutations in western Sweden detected with complementary screening techniques [J]. *Familial Cancer*, 2005, 4(2):89-96
- [4] Atipairin A, Canyuk B, Ratanaphan A. Substitution of aspartic acid with glutamic acid at position 67 of the BRCA1 RING domain retains ubiquitin ligase activity and zinc(II)binding with a reduced transition temperature[J]. *J Biol Inorg Chem*, 2011, 16(2):217-226
- [5] Carvalho MA, Marsillac SM, Karchin R, et al. Determination of cancer risk associated with germ line BRCA1 missense variants by functional analysis [J]. *Cancer Res*, 2007, 67(4):1494-1501
- [6] Tamura K., Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. *Mol Biol Evol*, 2011, 28(10):2731-2739
- [7] Letunic I, Doerks T, Bork P. SMART 6: Recent updates and new developments [J]. *Nucleic Acids Res*, 2009, 37(1):D229-D232
- [8] Schwede T, Kopp J, Guex N, et al. SWISS-MODEL: An automated protein homology-modeling server[J]. *Nucleic Acids Res*, 2003, 31(13):3381-3385
- [9] Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: A web-based environment for protein structure homology modeling [J]. *Bioinformatics*, 2006, 22(2):195-201
- [10] Stern A, Doron-Faigenboim A, Erez E, et al. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach [J]. *Nucleic Acids Res*, 2007, 35 (Web Server issue): W506-511
- [11] Vallon-Christersson J, Cayanan C, Haraldsson K, et al. Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families [J]. *Hum Mol Genet*, 2001, 10(4):353-360
- [12] Irminger-Finger I, Siegel BD, Leung WC. The functions of breast cancer susceptibility gene 1 (BRCA1) product and its associated proteins [J]. *Biol Chem*, 1999, 380(2):117-128
- [13] Maillet P, Chappuis PO, Khoshbeen-Boudal M, et al. Twenty-three novel BRCA1 and BRCA2 sequence variations identified in a cohort of Swiss breast and ovarian cancer families [J]. *Cancer Genetic and Cytogenesis*, 2006, 169(1):62-68
- [14] 南润玲, 尚培中, 王铁山. 乳腺癌易感基因 BRCA1 的研究进展[J]. *实用医学杂志*, 2011, 27(12):22284-2286
- [15] Fleming MA, Potter JD, Ramirez CJ, et al. Understanding missense mutations in the BRCA1 gene: an evolutionary approach [J]. *Proc Natl Acad Sci U S A*, 2001, 100(3):1151-1156

[收稿日期] 2012-01-13