

DOI: 10.3872/j.issn.1007-385x.2020.04.014

· 临床研究 ·

## 乳腺癌 HLA-A2 限制性新抗原表位的生物信息学分析

游紫聪<sup>a</sup>, 周炜均<sup>b,Δ</sup>, 罗云峰<sup>a</sup>, 邓鉴文<sup>a</sup>, 张普生<sup>a</sup>, 冯海湛<sup>a</sup>, 翁俊彦<sup>a</sup>, 俞金龙<sup>a</sup>, 朱卉娟<sup>a</sup>, 李玉华<sup>b</sup>, 史福军<sup>a</sup>(南方医科大学珠江医院 a. 普外中心乳腺外科; b. 血液科, 广东 广州 510282)

**[摘要]** **目的:** 通过生物信息学方法筛选乳腺癌 HLA-A2 限制性新抗原候选表位, 寻找乳腺癌新抗原高频突变位点。 **方法:** 利用 NCBI、GDC 数据库, 从已报道的文献与测序数据中寻找乳腺癌单核苷酸变异形成的错义突变位点, 通过 HLA-A2 抗原表位预测网站 BIMAS、SYFPEITHI 与基于人工神经网络的 NetMHC4.0 预测新抗原表位, 剔除 TAP 结合力低于 Intermediate 的表位, 按突变频率与预测结果对候选表位进行优先级排序。 **结果:** 使用上述方法共筛选出了 BTLA、ERBB2、NBPF12 等 17 个相对高频突变基因, 共 26 个新抗原表位, 不同表位 BIMAS、SYFPEITHI 预测结合力有较大差异 ( $P < 0.05$ ), 高优先级表位如 GSTP1(A114V, 突变频率 5.94%)、BRCA2(N991H, 突变频率 5.40%) 等有望成为新抗原候选位点, 但其突变频率较低, 难以做到真正的“通用”, 其作为乳腺癌通用新抗原疫苗表位的可能性较小。 **结论:** 乳腺癌共有突变低于其他肿瘤, 寻找“通用”乳腺癌新抗原表位较为困难; “个体化”新抗原疫苗也许更有前景, 但仍需要进一步研究来证实。

**[关键词]** 乳腺癌; 新抗原; HLA-A2; 二代测序

**[中图分类号]** R737.9; R730.51 **[文献标识码]** A **[文章编号]** 1007-385X(2020)04-0427-06

## Bioinformatics analysis of HLA-A2 restricted neoantigen epitopes in breast cancer

YOU Zicong<sup>a</sup>, ZHON Weijun<sup>b,Δ</sup>, LUO Yunfeng<sup>a</sup>, DENG Jianwen<sup>a</sup>, ZHANG Pusheng<sup>a</sup>, FENG Haizhan<sup>a</sup>, WENG Junyan<sup>a</sup>, YU Jinlong<sup>a</sup>, ZHU Huijuan<sup>a</sup>, LI Yuhua<sup>b</sup>, SHI Fujun<sup>a</sup> (a. Department of Breast Surgery, Center of General Surgery; b. Department of Hematology, Zhujiang Hospital of Southern Medical University, Guangzhou 510282, Guangdong, China)

**[Abstract]** **Objective:** To screen candidate epitopes of breast cancer HLA-A2 restrictive neoantigen and to identify high frequency mutation sites in breast cancer neoantigen by using bioinformatics method. **Methods:** NCBI and GDC databases were used to search missense mutation sites formed by single nucleotide mutation in breast cancer among reported literatures and sequencing data. The new antigen epitopes were predicted by HLA-A2 antigen epitope prediction website BIMAS, SYFPEITHI and artificial neural network-based NetMHC4.0, and the epitopes with TAP binding power less than Intermediate were eliminated. The candidate epitopes were prioritized by mutation frequency and prediction results. **Results:** A total of 17 high-frequency mutation genes, including BTLA, ERBB2 and NBPF12 etc, were screened by the above-mentioned methods, and a total of 26 neoantigen epitopes were identified. The binding power of epitopes predicted using BIMAS and SYFPEITHI showed great difference ( $P < 0.05$ ), epitopes in high priority as GSTP1 (A114V, mutation frequency of 5.94%) and BRCA2 (N991H, mutation frequency of 5.40%) etc, were expected to be candidate neo-antigen epitopes; however, their mutation frequency was relatively too low to achieve “universal use”. The possibility of these epitopes used as general breast cancer neo-antigen epitopes is less likely. **Conclusion:** The common mutation frequency of breast cancer is lower than that of other tumors; it's difficult to find “universal” new antigen epitopes of breast cancer; the individualized neoantigen vaccine may be of more promise, which needs further research.

**[Key words]** breast cancer; neoantigen; HLA-A2; next-generation sequencing (NGS)

[Chin J Cancer Biother, 2020, 27(4): 427-432. DOI: 10.3872/j.issn.1007-385X.2020.04.014]

**[基金项目]** 广东省自然科学基金资助项目(No.2014A030313334)。Project supported by the National Natural Science Foundation of Guangdong Province (No.2014A030313334)

**[作者简介]** 游紫聪(1994-), 男, 硕士生, 主要从事乳腺癌新抗原疫苗研究, E-mail: 447478149@qq.com; 周炜均(1987-), 女, 博士, 主要从事血液肿瘤免疫治疗及个体化新抗原疫苗的研究, E-mail: zhouweijun-123@163.com。Δ为共同第一作者

**[通信作者]** 史福军(SHI Fujun, corresponding author), 博士, 副教授, 硕士生导师, 主要从事乳腺癌免疫治疗的基础与临床研究, E-mail: drshifujun@163.com; 李玉华(LI Yuhua, co-corresponding author), 博士, 教授、博士生导师, 主要从事肿瘤生物治疗研究, E-mail: liyuhua2011gz@163.com

乳腺癌是源于乳腺导管上皮的恶性肿瘤, 据统计, 美国2019年女性乳腺癌新发病例为268 600人, 约占所有女性原发肿瘤病例的30%, 遥遥领先于第2位的肺癌和骨肿瘤<sup>[1]</sup>。目前乳腺癌的主要治疗手段包括手术、化疗、放疗、内分泌治疗、靶向治疗等, 使用这些治疗措施已经在早期乳腺癌中取得了相对理想的治疗效果, 但在某些乳腺癌经上述治疗后仍可复发或转移。同时, 随着TNM分期的延后, 乳腺癌的综合治疗效果越来越差<sup>[2]</sup>。鉴于以上治疗方法的局限性, 肿瘤免疫治疗在乳腺癌治疗的研究与日俱增, 乳腺癌新抗原疫苗疗法也成为乳腺癌治疗的新方向<sup>[3]</sup>。新抗原疫苗是一种以位于癌细胞表面主要组织相容性复合物(major histocompatibility complex, MHC)的抗原表位肽为抗原, 诱导T细胞对肿瘤细胞进行攻击, 从而消灭肿瘤的治疗性疫苗。然而, 乳腺癌的新抗原疫苗的研究屈指可数, ZHANG等<sup>[3]</sup>从3例乳腺癌患者的癌及癌旁组织中, 用高通量测序预测并制备了新抗原多肽疫苗, 将此疫苗输注到小鼠乳腺癌模型体内, 使肿瘤体积缩小, 取得了不俗的进展。这充分显示了新抗原疫苗疗法的有效性与良好的临床应用前景, 表明乳腺癌新抗原疫苗疗法是治疗乳腺癌的新突破, 是未来免疫治疗的重要发展方向。本研究旨在通过大数据分析, 寻找乳腺癌高频突变位点, 以探讨制备“通用型”乳腺癌新抗原疫苗的可行性。

## 1 资料与方法

### 1.1 乳腺癌高频新抗原突变位点的获取

SNP4disease数据库是搜索疾病与单核苷酸突变(single nucleotide polymorphism, SNP)关系的数据库, 其数据来源包括dbSNP、Pubmed、OMIM、MESH、Entrez Gene、NHGRI GWAS Catalog等。在SNP4disease (<http://snp4disease.mpi-bn.mpg.de/>)数据库上输入MESH“Breast Neoplasms”, 选择“Entrez Gene ID、Gene Symbol”, 可得出2 313个SNP位点。将搜索结果的dbSNP ID在SNP数据库(<https://www.ncbi.nlm.nih.gov/snp/>)上进行搜索, 选择Functional Consequence中的Missense, 获取其中的突变频率、突变氨基酸序列等信息。在GDC(<https://portal.gdc.cancer.gov/>)数据库上选择“Breast”, Biotype中选择“protein\_coding”, Consequence Type中选择“missense\_variant”, 可得出67 306个错义突变位点。根据相关位点信息于NCBI中Protein(<https://www.ncbi.nlm.nih.gov/>)数据库上获取突变蛋白氨基酸序列。

### 1.2 MHC-I类分子结合预测法预测乳腺癌的新抗原表位

MHC-I类分子结合预测方法主要包括定性定量两种, 无论何种筛选方法其筛选的重要标准是MHC-I类分子与肽段的结合力。定性预测通过预测分数判断结合力如BIMAS、SYFPEITHI、MULTIPRED等, 定量的预测包括对结合强度、结合分数等综合的预测, 如Net-MHC、Net-MHCpan、PSSMHCpan、MHCflurry等<sup>[4-6]</sup>。BIMAS与SYFPEITHI主要用于线性多肽结合力的预测, SYFPEITHI预测软件基于文献已报道的7 000个多肽的数据, 对预测短肽进行评分。

本研究利用在线细胞毒性T淋巴细胞(cytotoxic T lymphocyte, CTL)表位预测工具BIMAS ([https://www-bimas.cit.nih.gov/molbio/hla\\_bind/](https://www-bimas.cit.nih.gov/molbio/hla_bind/))、SYFPEITHI (<http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm>)与基于应用基于人工神经网络(artificial neural network, ANN)的在线表位预测软件NetMHC4.0 (<http://www.cbs.dtu.dk/services/NetMHC/>)进行CTL表位结合力预测。选择Bind level $\leq$ 2.0作为候选新抗原表位<sup>[7-9]</sup>。

新抗原多肽需要与抗原处理复合体转运子(transporter associated with antigen processing, TAP)转运到内质网, TAPPred (<http://www.imtech.res.in/raghava/tappred/>)可预测新抗原表位与TAP分子结合力, 可用于筛选有较大概率形成MHC-抗原肽复合物的表位<sup>[10]</sup>。剔除TAP转运亲和力低于Intermediate的候选表位。

### 1.3 统计学处理

采用SPSS25.0统计学软件对数据进行分析, 使用Wilcoxon秩和检验分析BIMAS、NetMHC预测分数的差异, 采用 $t$ 检验分析SYFPEITHI预测分数的差异。以 $P<0.05$ 或 $P<0.01$ 表示差异有统计学意义。

## 2 结果

### 2.1 筛选出乳腺癌17个靶基因和26个HLA-0201限制性新抗原表位

通过上述新抗原表位筛选方法, 筛选出了GDC数据库中8个靶基因、10个HLA-0201限制性新抗原表位(表1)和SNP数据库中9个靶基因、16个HLA-0201限制性新抗原表位(表2)。其中, 突变频率最高为GSTP1(A114V, 5.94%), 其次为BRCA2(N991H, 5.40%), 突变频率最低的为BTLA(R157S)、PALB2(K18R)、PALB2(G1145C)、ATR(M211T)、BTLA(p267L)、LHCGRSTON1-GTF2A1L(N206T)、EGFR(Q787H)、ATM(S707P)、ERBB2(I654L)、ATR(M211T)等共10个表位, 其突变频率均低于0.01%。

表1 GDC数据库中乳腺癌HLA-A2限制性新抗原表位  
Tab.1 HLA-A2 restricted neoantigen epitopes of breast cancer in GDC database

Missense ID	Epitope	Position	Sequence	Frequency [ (n/N)]
d84fa526-0940-5676-be5d-c199036dbb1d	NBPF12(E125Q)	122-130	SLNQHLQAL	1.48 (20/1 353)
11f63f77-cdbb-5f7c-9f7b-ca102f12ce2f	XPO1(E571K)	568-576	KLFKFMHET	0.95 (40/4 195)
d9c13559-9adc-59bb-aeaa-11c5ef8b011c	CLEC18(A24T)	20-28	LLGTTWAEV	0.42 (4/942)
e0d7799d-be6e-56ad-b9ca-b55bc0499409	ERBB2(D769Y)	767-775	ILYEAYVMA	0.41 (5/1 211)
c4e6fe3d-f02c-5a18-ab9c-dca3bf4f1f10	CDC42BPA(P675T)	669-677	KQISYSTGI	0.41 (7/1700)
fd747c5c-3915-5900-a971-2c0025baf366	ERBB2(I767M)	767-775	MLDEAYVMA	0.33 (4/1 211)
c1d0e49a-00ba-57b2-a5a7-c4c3cf688c11	PCDHGB4(D664N)	658-666	LLLVFANSL	0.32 (3/942)
70cd0f78-a517-5545-b82d-91dea457d116	ERBB2(D769H)	767-775	ILHEAYVMA	0.25 (3/1 211)
ed8cbf42-d86f-56bd-9159-d1cda7ca98fa	PTPRZ1(R833C)	830-838	ELFCHLHTV	0.24 (6/2 432)
4e9fe041-26bc-5e25-9b86-9e7c29aac779	GOLGA6L6(Q6H)	1-9	MLMWP <del>H</del> PHL	0.21 (2/942)

Note: The underlined amino acids in the table are mutated sites

表2 SNP数据库中乳腺癌HLA-A2限制性新抗原表位  
Tab.2 HLA-A2 restricted neoantigen epitopes of breast cancer in SNP database

bSNP ID	Epitope	Position	Sequence	Frequency [ (n/N)]
rs1138272	GSTP1(A114V)	106-114	SLIYTNYEV	5.94 (14 576/245 332)
rs1799944	BRCA2(N991D)	989-997	YMDKWAGLL	5.40 (13 255/245 530)
rs1804645	NCOA1(P1272S)	1266-1274	SLLQQTSSA	1.69 (2 045/121 396)
rs2234997	ATM(D126E)	119-127	ELLN <del>Y</del> IMET	1.57 (3 860/246 020)
rs1800058	ATM(L1420F)	1418-1426	ILFAICEQA	1.12 (2695/241276)
rs4986761	ATM(S707P)	701-709	LLNNYSPEI	0.80 (1 968/245 186)
rs180177137	PALB2(G1145C)	1142-1150	LLL <del>C</del> QCTAL	<0.01 (7/246 232)
rs180177137	PALB2(G1145C)	1143-1151	LL <del>C</del> QCTALL	<0.01 (7/246 232)
rs2931761	BTLA(R157S)	153-161	WLLYSLLPL	<0.01 (11/246232)
rs2227928	ATR(M211T)	208-216	TLL <del>T</del> VLTRI	<0.01 (11/246 232)
rs12470652	LHCGRSTON1-GTF2A1L(N206S)	203-211	TLQSFHSI	<0.01 (6/246 232)
rs9288952	BTLA(P267L)	259-267	SLNHSVIGL	<0.01 (5/246232)
rs1050171	EGFR(Q787H)	781-789	CLTSTV <del>H</del> LI	<0.01 (12/24 6232)
rs2227928	ATR(M211T)	204-212	FIEVTL <del>L</del> TV	<0.01 (11/246 232)
rs138789658	PALB2(K18R)	16-24	KL <del>R</del> EKLAFL	<0.01 (7/246 232)
rs1801201	ERBB2(I654L)	653-661	SLLSAVVGI	<0.01 (4/246 232)

Note: The underlined amino acids in the table are mutated sites

## 2.2 乳腺癌新抗原表位的BIMAS预测评分

BIMAS在线预测软件分析结果(图1)显示,预测分数在BRCA2(N991H)3.69~XPO1(E571K)843.44分之间,均值为(139.95±230.18)分, $\delta^2=52\ 984.58$ 。结果表明,不同的表位其预测结合力有较大差别( $P<0.05$ ),其预测分数可作为选择候选表位的重要参考依据。

## 2.3 乳腺癌新抗原表位的SYFPEITHI预测评分

SYFPEITHI在线预测软件分析结果(图2)显示,预测分数为CDC42BPA(P675T)13~PALB2(K18R)30分之间,均值为(23.11±4.34)分, $\delta^2=18.80$ 。结果表明,不同表位与MHC-I分子亲和力有较大区别( $t=29.13, P<0.05$ ),SYFPEITHI预测分数可用于候选表位优先级排序。

## 2.4 乳腺癌新抗原表位的NetMHC预测评分

NetMHC分数 $\leq 0.5$ 为具有强结合力, $\leq 2.0$ 为具有弱结合力,筛选结果均为具有MHC-I类分子弱结合力以上的新抗原表位<sup>[7]</sup>。NetMHC在线预测软件分析结果(图3)显示,结合力预测分数在ATM(H126Q)1.90~GSTP1(A114V)0.07之间,平均为(0.75±0.56)分。结果表明,不同表位预测结合力以及结合稳定性差异较小( $P<0.05$ ),NetMHC分数 $\leq 2.0$ 的表位均可作为候选表位。

## 3 讨论

由于医疗技术的发展和多种治疗手段的联合应用,乳腺癌的生存率逐年上升。然而,复发与远处转移的乳腺癌患者的5年生存率只有24%~39%<sup>[2]</sup>。现已公认,转移和复发是导致乳腺癌患者病死的主要原因。目前,对于复发或转移乳腺癌仍有效的治疗

方案。与此同时,患者的生活质量差,长期治疗过程中感染、胃肠道反应等不良反应降低患者治疗依从性,加重医疗负担,给家庭和社会带来极大负担。因此,复发或转移乳腺癌是严重的公共卫生问题,亟需寻找新型、有效的治疗方法解决患者长期生存的难题。乳腺癌新抗原疫苗疗法应运而生。

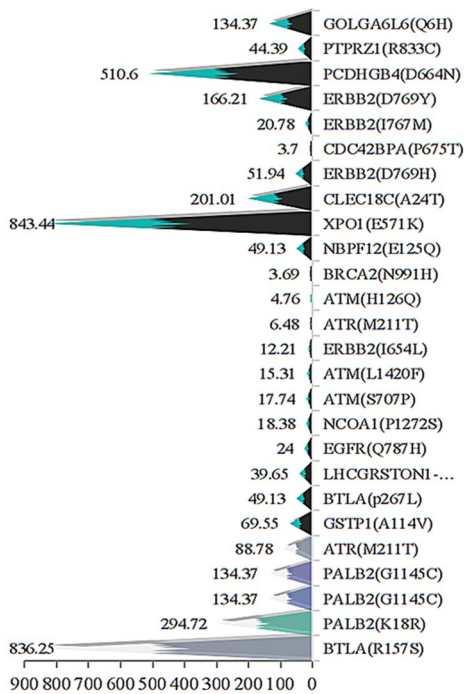


图1 乳腺癌新抗原表位BIMAS评分

Fig.1 BIMAS score for breast cancer neoantigen epitopes

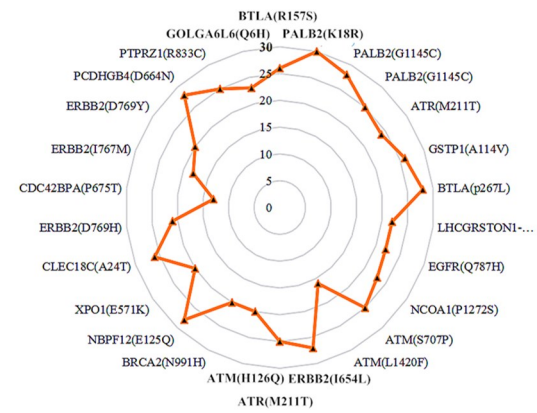


图2 乳腺癌新抗原表位SYFPEITHI评分

Fig.2 SYFPEITHI score of breast cancer neoantigen epitopes

新抗原疫苗又分为个体化新抗原疫苗与“通用型”新抗原疫苗,个体化新抗原疫苗可以做到精准治疗。针对每个人的突变位点不同,研制的疫苗亦具有唯一性。对于高突变负荷的肿瘤,可以做到联合多个治疗靶点,大大减少肿瘤免疫逃逸发生的可能性。2014年, SCHUMACHER 等<sup>[11]</sup>以 IDH1(R132H)位点,体内验证了治疗神经胶质瘤新抗原多肽诱导生成 IDH1 (R132H) HLA-DRB1\*0101 tetramer-positive T cells 可以使小鼠动物模型肿瘤缩小乃至消失。2017年,OTT 等<sup>[12]</sup>在6例晚期黑色素瘤患者中,联合使用新抗原疫苗与抗程序性死亡蛋白1(programmed death protein 1, PD-1)治疗后肿瘤消退。

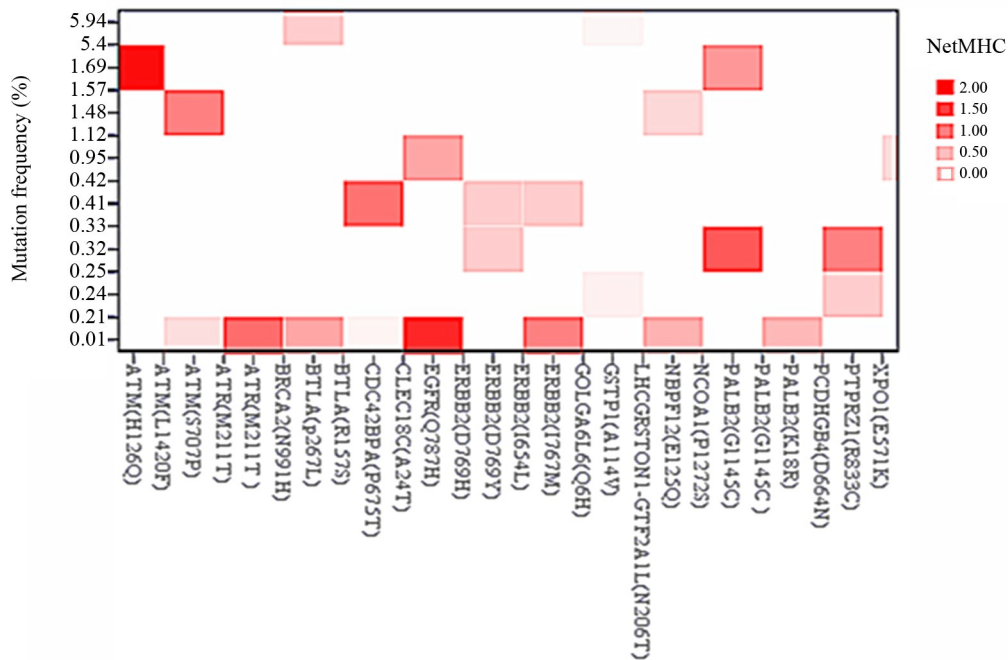


图3 乳腺癌新抗原表位突变频率与NetMHC评分

Fig.3 Mutation frequency and NetMHC score of breast cancer neoantigen epitopes

该研究表明了新抗原多肽疫苗的可行性、安全性及强免疫原性。但是,个体化新抗原疫苗需要联合高通量测序、基因变异检测<sup>[13]</sup>等新兴技术,制作工艺复杂,制作周期较长,可能延误患者的最佳治疗时机,这成为制约个体化新抗原疫苗发展的重要因素。“通用型”新抗原疫苗是指使用高频突变位点作为靶点的疫苗,其具有传统疫苗制作工艺简单、耗时短的特点,同时亦具有新抗原疫苗特异性杀伤肿瘤细胞的优势。但是,相对于传统疫苗的广泛适用性,“通用型”新抗原疫苗适用于高突变负荷的肿瘤,对于突变率低的肿瘤,较难寻找适合的新抗原表位。目前关于乳腺癌的个体化新抗原疫苗研究大部分处于临床试验阶段,根据 ClinicalTrials.gov 网站显示,截至2019年7月,关于新抗原疫苗的临床研究共有122项,关于乳腺癌有13项,其中美国开展最多(共11项),尚未见乳腺癌“通用型”抗原疫苗的报道<sup>[14]</sup>。

“通用型”新抗原疫苗是新抗原疫苗疗法的新发展方向,其将单个体化变为泛个体化。“通用型”新抗原疫苗通过数据库和文献检索查找和发掘新抗原。此法无需获取标本进行测序,该方法的独到之处在于可以发现实体瘤的高频突变位点。得益于如TCGA、NCBI等庞大的测序数据库,在制备“通用型”新抗原疫苗时可根据不同的人种或地区等条件筛选合适的样本,从中寻找高频新抗原。然后通过CTL表位预测软件对新抗原表位进行预测。对优先级较高的新抗原肽进行体内外实验验证其免疫原性与诱导生成CTL的杀伤力。其中,新抗原疫苗研究的重点与难点在于新抗原表位的寻找。目前,尚未见有关乳腺癌“通用型”新抗原疫苗的报道。本研究顺应这一乳腺癌免疫治疗新方向,通过NCBI、TCGA等公共数据库挖掘了乳腺癌的高频突变位点。同时,利用在线CTL表位预测工具BIMAS、SYFPEITHI与基于应用ANN等的在线表位预测软件NetMHC进行CTL表位结合力预测。BIMAS在线预测软件主要用于线性表位的预测,预测结果中分数越高结合力越强;SYFPEITHI是适用结合基序的方法对抗原肽与MHC-I类分子结合力进行评分的方法,其预测准确率可达50%<sup>[15]</sup>,预测结果中分数越高结合力越强。NetMHC是基于神经网络的预测方法,NIELSEN等<sup>[7]</sup>对多种预测方法进行了相关研究,证明了NetMHC预测方法的可靠性。现实中,由于多肽之间的相互影响以及相关酶蛋白的作用,多肽常常折叠形成几何结构。NetMHC算法充分考虑到多肽的空间结构,预测多肽与MHC分子的结合力更为准确<sup>[7-8,16]</sup>。

利用数据预测软件,本实验筛选出13个强结合力(百分率排位, Bind level $\leq$ 0.5)HLA-A0201限制性

新抗原短肽表位,13个弱结合力表位(百分率排位, Bind level $\leq$ 2.0)。这些新抗原表位中,突变频率最高的为GSTP1(A114V,5.94%)。最低为BTLA(R157S)等10个表位,突变频率均低于0.01%。理论上而言,其他高频突变,如PIK3CA、P53、MUC4等也可能成为肿瘤疫苗的理想靶点。根据GDC数据库,在乳腺癌所有新抗原表位中,突变频率最高为PIK3CA[H1047R,12.1%(115/942)]。但是,根据预测结果,由于其突变位点附近缺乏蛋白酶酶切位点或TAP结合力较低等原因,导致其无法被切割成新抗原短肽进入下一步免疫应答,属于无效表位。“通用型”新抗原疫苗适用于突变负荷高的肿瘤,如神经胶质瘤, IDH1(R132H)新抗原表位在II、III期患者中突变频率超过70%<sup>[11]</sup>,以此靶点研制的新抗原疫苗的有效性与安全性已在相关临床研究中得到验证,在KRAS突变的非小细胞肺癌和结肠癌患者中,KRAS(G12V)突变率达到20%~40%<sup>[17]</sup>;VEATCH等<sup>[18]</sup>通过细胞和动物实验验证了以此为靶点的新抗原疫苗的免疫原性与诱导CTL的杀伤力,具有相当不错的临床应用前景。此外,还有多种肿瘤的共有突变位点:KRAS-G12D、KRAS-G12R、KRAS-G13D、NRAS-Q61R、PIK3CA-H1047R和C-Kit-D816V等<sup>[9]</sup>,以这些突变位点为靶点的新抗原疫苗研究已取得不俗的成果。突变负荷高的肿瘤经过体内外实验筛选后有较大的概率获取高频突变位点,使“通用型”新抗原疫苗在临床的应用成为可能。但是,乳腺癌有效新抗原表位突变频率较低,无法适用于大部分人群,难以做到真正的“通用”。本研究运用多种生物信息学方法,虽然未如期预期的预测到乳腺癌HLA-I类潜在高频CTL新抗原表位,但为下一步乳腺癌“通用型”新抗原疫苗的研制提供了依据。

综上所述,本研究经过上述条件筛选后剩余表位突变频率均较低,针对乳腺癌制备“通用型”疫苗较为困难。所幸的是,越来越多高准确率预测工具的成功编程极大地缩短了从个体突变位点寻找有效抗原表位的过程,并且在新抗原表位数据库的支持下,未来有望快速为肿瘤患者制订个体化疫苗方案,从而为乳腺癌的治疗提供新的手段。

## [参考文献]

- [1] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2019[J]. CA Cancer J Clin, 2019, 69(1): 7-34. DOI:10.3322/caac.21551.
- [2] DESANTIS C E, FEDEWA S A, GODING SAUER A, et al. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women[J]. CA Cancer J Clin, 2016, 66(1): 31-42. DOI:10.3322/caac.21320.
- [3] ZHANG X L, KIM S, HUNDAL J, et al. Breast cancer neoantigens

- can induce CD8<sup>+</sup> T-cell responses and antitumor immunity[J/OL]. *Cancer Immunol Res*, 2017, 5(7): 516-523[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5647648/>. DOI: 10.1158/2326-6066.CIR-16-0264.
- [4] LUNDEGAARD C, LAMBERTH K, HARND AHL M, et al. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11[J/OL]. *Nucleic Acids Res*, 2008, 36(Web Server issue): W509-W512[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2447772/>. DOI:10.1093/nar/gkn202.
- [5] LIU G, LI D L, LI Z, et al. PSSMHCPan: a novel PSSM-based software for predicting class I peptide-HLA binding affinity[J/OL]. *Gigascience*, 2017, 6(5): 1-11 [2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5467046/>. DOI: 10.1093/gigascience/gix017.
- [6] JURTZ V, PAUL S, ANDREATTA M, et al. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data[J/OL]. *J Immunol*, 2017, 199(9): 3360-3368[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5679736/>. DOI:10.4049/jimmunol.1700893.
- [7] NIELSEN M, LUNDEGAARD C, WORNING P, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations[J/OL]. *Protein Sci*, 2003, 12(5): 1007-1017[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2323871/>. DOI: 10.1110/ps.0239403.
- [8] ANDREATTA M, NIELSEN M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system[J/OL]. *Bioinformatics*, 2016, 32(4): 511-517[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6402319/>. DOI:10.1093/bioinformatics/btv639.
- [9] O'DONNELL T J, RUBINSTEYN A, BONSAK M, et al. MHCflurry: open-source class I MHC binding affinity prediction[J/OL]. *Cell Syst*, 2018, 7(1): 129-132.e4[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pubmed/29960884>. DOI:10.1016/j.cels.2018.05.014.
- [10] NIELSEN M, LUNDEGAARD C, LUND O, et al. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage[J]. *Immunogenetics*, 2005, 57(1/2): 33-41. DOI:10.1007/s00251-005-0781-7.
- [11] SCHUMACHER T, BUNSE L, PUSCH S, et al. A vaccine targeting mutant IDH1 induces antitumor immunity[J]. *Nature*, 2014, 512(7514): 324-327. DOI:10.1038/nature13387.
- [12] OTT P A, HU Z T, KESKIN D B, et al. An immunogenic personal neoantigen vaccine for patients with melanoma[J/OL]. *Nature*, 2017, 547(7662): 217-221[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5577644/>. DOI:10.1038/nature22991.
- [13] 张乐吟, 孙磊涛, 沈敏鹤. 基因变异检测技术在恶性肿瘤精准医疗中的应用[J]. *中国肿瘤生物治疗杂志*, 2019, 26(1): 22-28. DOI: 10.3872/j.issn.1007-385x.2019.01.005.
- [14] YAMAMOTO T N, KISHTON R J, RESTIFO N P, et al. Developing neoantigen-targeted T cell-based treatments for solid tumors, 2019[J]. *Nat Med*, 2019, 25(10): 1488-1499. DOI: 10.1038/s41591-019-0596-y.
- [15] SCHULER M M, NASTKE M D, STEVANOVIĆ S. SYFPEITHI: database for searching and T-cell epitope prediction[J]. *Methods Mol Biol*, 2007, 409: 75-93. DOI: 10.1007/978-1-60327-118-9\_5.
- [16] KEŞMİR C, NUSSBAUM A K, SCHILD H, et al. Prediction of proteasome cleavage motifs by neural networks[J]. *Protein Eng Des Sel*, 2002, 15(4): 287-296. DOI:10.1093/protein/15.4.287.
- [17] FORBES S A, BEARE D, BINDAL N, et al. COSMIC: high-resolution cancer genetics using the catalogue of somatic mutations in cancer[J]. *Curr Protoc Hum Genet*, 2016, 91: 10.11.1-10.1110.11.37. DOI: 10.1002/cphg.21.
- [18] VEATCH J R, JESERNIG B L, KARGL J, et al. Endogenous CD4<sup>+</sup> T cells recognize neoantigens in lung cancer patients, including recurrent oncogenic KRAS and ERBB (Her2) driver mutations[J/OL]. *Cancer Immunol Res*, 2019, 7(6): 910-922[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6584616/>. DOI: 10.1158/2326-6066.CIR-18-0402.
- [19] IIZUMI S, OHTAKE J, MURAKAMI N, et al. Identification of novel HLA class II-restricted neoantigens derived from driver mutations[J/OL]. *Cancers (Basel)*, 2019, 11(2): E266[2019-09-08]. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6406322/>. DOI:10.3390/cancers11020266.

[收稿日期] 2019-09-10

[修回日期] 2020-01-31

[本文编辑] 党瑞山