

全基因组关联研究在乳腺癌筛查中的应用价值初探

黄育北 宋丰举 陈可欣

天津医科大学肿瘤医院流行病与生物统计室 国家肿瘤临床医学研究中心 天津市肿瘤防治重点实验室 300060

通信作者:宋丰举, Email:songfengju@163.com

【摘要】目的 探索全基因组关联研究(GWAS)发现的单核苷酸多态性位点(SNP)在乳腺癌筛查中的潜在应用价值。**方法** 基于我国女性2013年的年龄构成、年龄别的乳腺癌发病率,以及明确的乳腺癌传统危险因素分布情况,对中国200万35~69岁女性人群进行模拟。进一步模拟GWAS发现的23个与我国女性乳腺癌风险相关的SNP位点的分布情况。依据SNP的遗传风险解释程度及风险再分类准确性的改善程度,初筛出可用于预测乳腺癌高危人群的目标SNP,并进一步探索目标SNP对乳腺癌检出率、乳腺癌风险预测曲线下面积(AUC)、高危人群中乳腺癌发病风险的影响。**结果** 共发现12个SNP可用于预测乳腺癌高危人群。如果将预测风险位于 P_{95} 及更高风险的人群定义为高危人群,并在此类人群中进行筛查,采用目标SNP预测的高危人群中的乳腺癌检出率(146.99/10万)明显低于采用传统危险因素预测的高危人群中的乳腺癌检出率(177.46/10万)($P<0.001$)。在传统危险因素基础上,加上目标SNP进行高危人群预测,高危人群中乳腺癌检出率(229.00/10万)提高29.0%($P<0.001$)。同时乳腺癌风险预测的AUC从64.4%上升至67.8%($P<0.001$),高危人群中乳腺癌发病风险OR值从3.32上升至4.33。**结论** GWAS筛选出的目标SNP可提高乳腺癌检出率、乳腺癌总体风险预测准确性,并有助于在乳腺癌筛查前发现潜在的乳腺癌高危人群。

【关键词】 乳腺肿瘤;全基因组关联研究;单核苷酸多态性

基金项目:国家自然科学基金(81502476);国家重点研发计划(2018YFC1315600)

DOI:10.3760/cma.j.issn.0254-6450.2019.06.021

Application values of genome-wide association studies in screening for breast cancer

Huang Yubei, Song Fengju, Chen Kexin

Department of Epidemiology and Biostatistics, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy of Tianjin, Tianjin Medical University Cancer Hospital, Tianjin 300060, China

Corresponding author: Song Fengju, Email: songfengju@163.com

【Abstract】 Objective To investigate the potential application values of screening on breast cancer, using the single-nucleotide polymorphisms (SNPs) that were identified from the genome-wide association studies (GWASs). **Methods** Two million Chinese women aged 35–69 years were simulated, based on both age distributions, age-specific incidence rates of breast cancer and the distribution of known risk factors, in 2013. Twenty-three SNPs identified from GWAS were further simulated. Both genetic-related risks explained by each SNPs and the improvement on the risks under reclassification, were used to select SNPs for the prediction on breast cancer among the targeted high-risk population. Further analyses were conducted to investigate the following items as: improvements on detection rates of breast cancer among the high-risk populations, areas under the curve (AUC) and the odds ratio (OR) among women at high risk. **Results** A total of 12 SNPs were eligible for targeting the high-risk population of breast cancer. When high-risk populations were defined as women whose predicted risks were higher than the 95th predicted risk of the whole population, the detection rate (146.99/100 000) among high-risked women predicted by 12 SNPs would be significantly lower than 177.46/100 000, which was predicted by the known risk factors ($P<0.001$), among the high-risked women. Among those women at high risk, the detection rate (229.00/100 000) predicted by integrating known risk factors and 12 SNPs was significantly higher than that predicted by known risk factors ($P<$

0.001)。Also, the AUC increased from 64.4% to 67.8% ($P < 0.001$), and the *OR* of increased from 3.32 to 4.33, predicted by integrating known risk factors and 12 SNPs, for women at high risk on breast cancer. **Conclusion** Targeted SNPs that were identified from genome-wide association studies could be used to improve the detection rates as well as the overall accuracy of risk prediction so as to identify the potential high-risk women on breast cancer before carrying on the screening program.

【Key words】 Breast neoplasms; Genome-wide association study; Single-nucleotide polymorphism

Fund programs: National Natural Science Foundation of China (81502476); National Key Research and Development Plan of China (2018YFC1315600)

DOI:10.3760/cma.j.issn.0254-6450.2019.06.021

近年来,国内外有关乳腺癌全基因组关联研究(GWAS)较多^[1-10],其中5篇来自亚洲地区人群的GWAS^[11-15]共发现了超过90个单核苷酸多态性(SNP)位点与乳腺癌发病风险相关。这些SNP当中有23个SNP位点与我国女性乳腺癌的风险潜在相关^[16]。随着GWAS的不断开展,未来会发现更多与乳腺癌相关的SNP。与此同时,探索这些遗传易感位点在高危人群中的筛查应用价值显得尤为重要^[17]。目前有不少研究开始初步探索这些SNP在乳腺癌筛查中的价值^[18-21]。在开展大规模的人群筛查验证研究之前,为避免资源浪费,并初步探索这些SNP在乳腺癌筛查中的预期效果,有必要开展合理的数据模拟研究。

对象与方法

1. 研究对象:依据2017年中国肿瘤登记报告的2013年乳腺癌年龄别发病率^[22]、2013年的中国女性年龄别的人口分布^[23],以及目前已经明确的乳腺癌传统危险因素(初潮年龄早、初次分娩年龄晚、腰臀比大、乳腺癌家族史、良性乳腺疾病史)分布情况及其与乳腺癌的关联强度^[24],模拟200万35~69岁中国女性人群的乳腺癌患病及危险因素分布情况。

具体模拟过程:①依据《中国卫生统计年鉴2014》报告的2013年35~69岁中国女性不同年龄段(每5岁1个年龄段)人口构成^[23],计算200万女性不同年龄段的人口数,并进行模拟。②依据2013年乳腺癌年龄别发病率及不同年龄段的人口数^[22],计算200万女性不同年龄段的乳腺癌例数和健康人群例数,并进行模拟。③依据Zheng等^[24]报告的已经明确的乳腺癌传统危险因素(初潮年龄早、初次分娩年龄晚、腰臀比大、乳腺癌家族史、良性乳腺疾病史)分别在病例组和对照组的分布概率,采用多项分布或二项分布的方法,分别在乳腺癌患者和健康人群中模拟不同危险因素分布。④由于人群中乳腺癌患者例数明显少于健康人群例数,因此,乳腺癌患者中危险因素的模拟分布与真实分布之间的随机误

差,较健康人群中危险因素的模拟分布与真实分布之间的随机误差要大。因此,分别采用不同的标准对乳腺癌患者和健康人群中危险因素分布的随机误差进行校正。即当某一危险因素在乳腺癌患者中的模拟分布概率与既往患者的真实分布概率正负相差超过2%时,以0.05%的精度对相应危险因素在乳腺癌患者的分布概率进行重新模拟。当某一危险因素在健康人群中的模拟分布概率与既往对照人群中的真实分布概率正负相差超过0.03%时,以0.01%的精度对相应危险因素在非乳腺癌患者的分布概率进行重新模拟。⑤对重新模拟的所有人群,分别计算传统危险因素联合预测乳腺癌的受试者作业特征(ROC)曲线下面积(AUC),并计算所有AUC的均值,以最接近AUC均值的模拟人群作为最终的模拟人群。

模拟人群中各传统危险因素的分布及其与乳腺癌的关联强度见表1。同时基于原始亚洲人群GWAS或亚洲乳腺癌联盟(Asia Breast Cancer Consortium)中所报道的SNP位点在对照组中微小等位基因的分布频率及与乳腺癌关联强度的*OR*值^[11-15,25],以上述传统危险因素的模拟方法进一步模拟人群中23个与我国乳腺癌风险潜在相关的SNP分布。

2. SNP初筛:采用原始GWAS中所报道的SNP位点在对照组中微小等位基因的分布频率及与乳腺癌关联强度的*OR*值,采用Pharoah等^[20]提出的方法分别计算每个SNP对乳腺癌遗传风险的解释程度(V_i)和解释比例(F_{V_i})。在模拟人群中,以 F_{V_i} 最大的SNP构建遗传风险预测的基础模型,然后按照 F_{V_i} 的大小依次纳入其他的SNP,分别构建23个模型。计算在原有SNP基础上加入新的SNP后,对乳腺癌风险再分类准确性的改善程度,最终选取能够同时带来显著的整合区分改善指数(integrated discrimination improvement, *IDI*)及净再分类改善指数(net reclassification improvement, *NRI*)的SNP,作为可以预测乳腺癌高危人群的目标SNP^[16]。即目标SNP的初筛依据为在基础模型的基础上,加入目标SNP后,对于乳腺癌风险再分类的*IDI*值和*NRI*值所

表1 200万35~69岁模拟人群中乳腺癌传统危险因素的分布及其与乳腺癌之间的关联

变量	病例 (n=1 449)	对照 (n=1 998 551)	OR值(95%CI) ^a
年龄组(岁) ^{b,c}			
35~	99(6.83)	319 948(16.01)	1.00
40~	239(16.49)	393 280(19.68)	1.96(1.55~2.48)
45~	307(21.19)	373 766(18.70)	2.65(2.11~3.33)
50~	233(16.08)	264 770(13.25)	2.85(2.25~3.60)
55~	256(17.67)	269 779(13.50)	3.06(2.43~3.86)
60~	199(13.73)	224 857(11.25)	2.86(2.25~3.64)
65~	116(8.01)	152 151(7.61)	2.46(1.88~3.22)
初潮年龄(岁) ^d			
>15	358(24.70)	644 424(32.24)	1.00
14~	631(43.55)	795 206(39.79)	1.43(1.26~1.63)
≤13	460(31.75)	558 921(27.97)	1.48(1.29~1.70)
初次分娩年龄(岁) ^d			
<25	315(21.73)	519 598(26.00)	1.00
25~	795(54.87)	1 086 308(54.35)	1.21(1.06~1.38)
≥30或未生育	339(23.40)	392 645(19.65)	1.42(1.22~1.66)
腰臀比 ^d			
<0.783	340(23.46)	648 715(32.46)	1.00
0.783~	496(34.23)	679 375(33.99)	1.39(1.21~1.60)
≥0.830	613(42.31)	670 461(33.55)	1.74(1.53~1.99)
乳腺癌家族史 ^d			
否	1 378(95.10)	1 942 067(97.17)	1.00
是	71(4.90)	56 484(2.83)	1.77(1.40~2.25)
良性乳腺疾病史 ^d			
否	940(64.87)	1 521 444(76.13)	1.00
是	509(35.13)	477 107(23.87)	1.73(1.55~1.92)

注:^a所有的OR值均为多因素logistic回归分析的结果;^b原始参数来源于《中国肿瘤登记年报2017》;^c原始参数来源于《中国卫生统计年鉴2014》(《中国肿瘤登记年报2017》报告的是2013年的肿瘤登记结果,因此采用2013年的女性相应年龄段的人口分布情况);^d原始参数来源于文献[24]

相应的P值均<0.05。

3. SNP筛查应用价值评价:依据传统乳腺癌危险因素、目标SNP,及联合传统危险因素和目标SNP,分别构建乳腺癌传统风险预测模型、遗传风险预测模型及综合风险预测模型。进一步系统评价在传统乳腺癌危险因素的基础上,加入目标SNP后,对高危人群中乳腺癌检出率、对乳腺癌总体风险预测准确性及高危人群中乳腺癌患病风险的影响。其中乳腺癌总体风险预测准确性以风险预测AUC来表示,高危人群中乳腺癌的患病风险以OR值及其95%CI表示。

4. 统计学分析:采用SAS 9.3软件进行数据模拟和统计学分析,包括采用 χ^2 分析比较两个检出率差异是否有统计学意义,采用IDI值和NRI值评价SNP对乳腺癌风险再分类准确性改善能力的影响,采用多因素logistic回归计算相应的OR值及其95%CI。采用R 3.5.1软件的ggplot2程序包(2.2.1版本)进行

统计绘图;同时采用pROC1.10.0程序包计算AUC及其95%CI,并比较AUC大小。所有检验均采用双侧检验,以P<0.05为差异有统计学意义。

结 果

1. GWAS发现SNP对女性乳腺癌遗传风险的解释能力:23个SNP共能解释大约11.9%的中国女性乳腺癌遗传风险。23个SNP中,rs2046210所能解释的遗传风险比例最大,占总体可解释乳腺癌遗传风险的21.6%。因此,以rs2046210构建遗传风险预测的基础模型,然后按照 F_v 的大小依次纳入其他SNP。在200万35~69岁中国女性模拟人群中,发现其他11个SNPs(rs4784227、rs1219648、rs10771399、rs9485372、rs4973768、rs10474352、rs17817449、rs9693444、rs616488、rs16857609、rs17356907)可以显著提高乳腺癌风险再分类的准确性,IDI值从0.000 3%~0.003%不等,NRI值从5.8%~19.4%不等(均P<0.05)。因此,共有12个SNPs可以用于预测乳腺癌高危人群。见表2。

2. GWAS发现SNP对模拟人群中女性乳腺癌检出率的影响:如果将预测风险位于 P_{95} 及更高风险的人群定义为高危人群,在以6个传统乳腺癌危险因素预测的高危人群中(共计99 741人),共有177例乳腺癌患者(占所有人群中乳腺癌患者的比例为12.2%),患病率为177.46/10万。以12个目标SNP预测的高危人群中(共计100 005人),共有147例乳腺癌患者(占所有人群中乳腺癌患者的比例为10.1%),患病率为146.99/10万。在6个传统危险因素和12个目标SNP联合预测的高危人群中(共计100 000人),共有229例乳腺癌患者(占所有人群中乳腺癌患者的比例为15.8%),患病率为229.00/10万。也即采用12个目标SNP预测的乳腺癌高危人群中的乳腺癌检出率(146.99/10万)低于采用6个传统危险因素预测的乳腺癌高危人群中的检出率(177.46/10万),差异有统计学意义(P<0.001)。但如果在6个传统危险因素的基础上,加上12个目标SNP预测的乳腺癌高危人群中进行筛查,乳腺癌的检出率(229.00/10万)将提高29.0%,差异有统计学意义(P<0.001)。见图1。

3. GWAS发现SNP对乳腺癌总体风险预测准确性及高危人群中乳腺癌相对风险的影响:基于上述乳腺癌传统风险预测模型、遗传风险预测模型及综合风险预测模型得到的乳腺癌总体风险预测准确性的AUC分别为64.4%(95%CI: 63.0%~65.8%)、

表 2 全基因组关联研究发现 SNP 对乳腺癌遗传风险解释能力及对风险再分类的改善能力

SNP	V_i	F_{vi}	IDI		NRI	
			改善程度 (%)	P 值	改善程度 (%)	P 值
rs2046210(G>A)	0.026	21.60	0		0	
rs4784227(T>C)	0.021	17.85	0.003 0	<0.001	19.4	<0.001
rs1219648(G>A)	0.010	8.60	0.000 3	0.015	5.8	0.028
rs10771399(A>G)	0.006	5.03	0.000 9	<0.001	10.2	<0.001
rs9485372(A>G)	0.006	4.62	0.000 6	0.003	7.5	0.004
rs4973768(C>T)	0.004	3.35	0.000 7	0.003	6.5	0.013
rs10941679(A>G)	0.004	3.12	0.000 1	0.131	2.8	0.286
rs10474352(C>T)	0.004	3.12	0.000 6	0.002	6.9	0.009
rs10822013(C>T)	0.004	3.11	0.000 1	0.364	4.2	0.112
rs17817449(T>G)	0.004	2.98	0.001 1	<0.001	9.9	<0.001
rs2236007(G>A)	0.003	2.77	0.000 2	0.078	4.1	0.115
rs1432679(T>C)	0.003	2.61	0.000 3	0.047	4.6	0.079
rs4951011(G>A)	0.003	2.53	0.000 1	0.051	2.0	0.438
rs2290203(G>A)	0.003	2.49	0.000 1	0.244	3.3	0.204
rs9693444(C>A)	0.003	2.42	0.001 2	<0.001	9.7	<0.001
rs7107217(C>A)	0.003	2.27	<0.000 1	1.000	1.4	0.592
rs1292011(A>G)	0.003	2.23	0.000 2	0.130	2.9	0.274
rs616488(A>G)	0.002	1.91	0.000 6	0.001	7.8	0.003
rs16857609(C>T)	0.002	1.86	0.000 7	0.003	8.9	0.001
rs6472903(T>G)	0.002	1.52	0.000 2	0.026	2.2	0.403
rs4849887(C>T)	0.002	1.49	0.000 1	0.265	3.5	0.182
rs7697216(C>T)	0.002	1.41	0.000 3	0.053	4.7	0.077
rs17356907(A>G)	0.001	1.13	0.000 7	0.002	9.5	<0.001

注:SNP:单核苷酸多态性; V_i 和 F_{vi} 为每个 SNP 对乳腺癌遗传风险的解释程度和解释比例;IDI:整合区分改善指数;NRI:净再分类改善指数

61.1% (95% CI: 59.6% ~ 62.5%) 及 67.8% (95% CI: 66.5% ~ 69.2%)。如果以人群预测风险 $P_{45} \sim P_{50}$ 的人群作为参考人群,根据 12 个目标 SNP 预测的高危人群发生乳腺癌的风险 ($OR=2.45$, 95% CI: 1.82 ~ 3.31) 低于 6 个传统危险因素预测的高危人群发生乳腺癌的风险 ($OR=3.32$, 95% CI: 2.45 ~ 4.49)。联合 6 个传统危险因素和 12 个目标 SNP 综合预测的高危人群发生乳腺癌的风险 ($OR=4.33$, 95% CI: 3.21 ~ 5.84), 明显高于单纯基于传统危险因素预测的高危人群的乳腺癌相对风险。也即在 6 个传统危险因素的基础上,加上 12 个目标 SNP 预测高危人群,将显著提高乳腺癌总体风险预测准确性,并能明显提高高危人群中乳腺癌的相对风险(图 2)。

讨 论

本研究采用数据模拟的办法,初步探索了 GWAS 发现 SNP 在我国女性乳腺癌筛查中的潜在应用价值。研究结果提示:在传统危险因素基础上,加入 GWAS 发现 SNP 用于预测乳腺癌高危人群,可显著提高乳腺癌检出率、乳腺癌总体风险预测准确性,

并能更好地发现真正的乳腺癌患者。之前有类似研究探讨 GWAS 发现的 SNP 在乳腺癌风险预测中的潜在应用价值^[17-20,26],但大部分为理论方法研究。我国人口基数庞大、社会资源非常有限,开展合理的数据模拟研究,对未来更大规模的乳腺癌筛查实践,有很好的指导意义。

本研究在初筛目标 SNP 时,主要采用 SNP 对乳腺癌风险再分类准确性(以 IDI 值或 NRI 值评价)有显著改善的 SNP,而没有直接采用 SNP 能对乳腺癌总体风险预测准确性的 AUC 有显著改善的 SNP 作为目标 SNP。因为之前的研究提示:单一 SNP 对乳腺癌总体风险预测准确性的贡献很有限,因此很可能会遗漏很多潜在有价值的 SNP^[16,21]。如本研究所示,在已知的传统乳腺癌危险因素的基础上,加入 12 个目标 SNP,仅能让乳腺癌总体风险预测准确性的 AUC 提高 3.4%。该结果与既往研究结果非常类似,包括 Wacholder 等^[18]采用的 10 个 SNP 可在乳腺癌传统风险预测模型基础上将 AUC 提高 3.8%, Mealiffe 等^[21]报告的 7 个 SNP 可将乳腺癌传统 Gail 风险预测模型的 AUC 提高 3.7%。同时这些研究采用的 SNP 均为当时发现的所有乳腺癌风险

相关 SNP,并没有对相应的 SNP 进行初步筛查,因此对当前如何从 GWAS 发现的大量 SNP 中筛选出更有潜在人群应用价值的目标 SNP 的指导意义相对欠缺。

风险再分类改善指标,可以认为是风险预测准确性的相对指标。也即在原有预测变量的基础上,加入新的预测变量,只要方向上能让患者的风险重新再分类为更高风险,非患者的风险重新再分类为更低风险,即认为风险再分类发生了改善。相对于预测准确与否的二分类预测方法,这种三分类的风险再分类结果(风险上升、风险不变、风险下降)理论上能发现风险预测上更加细微的改善^[27]。而且本研究提示目标 SNP 对乳腺癌风险再分类的改善强于对乳腺癌总体风险预测准确性的改善。在使用风险再分类改善指标初筛目标 SNP,进而采用目标 SNP 初筛乳腺癌高危人群时,需注意以下问题:当得到初步验证的目标 SNP 相对较多时,须同时兼顾效果和成本两个方面来选择最终需检测的 SNP,通常可进一步严格 SNP 的初筛标准(如将 IDI 值和 NRI 值所对应的 P 值下调至 <0.001)。反之,当得到初步验证的目

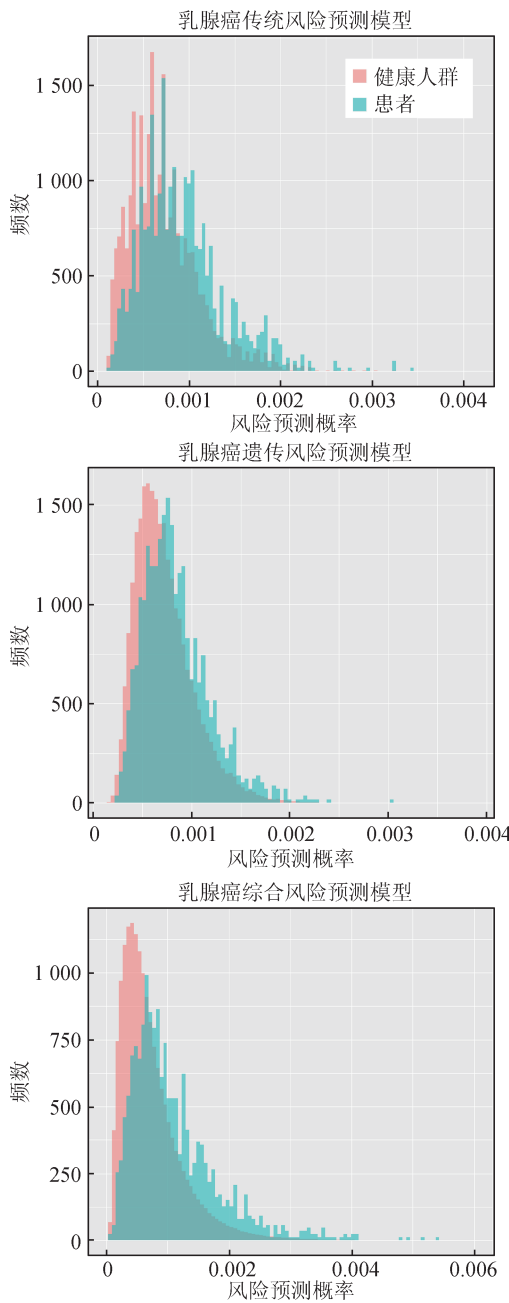
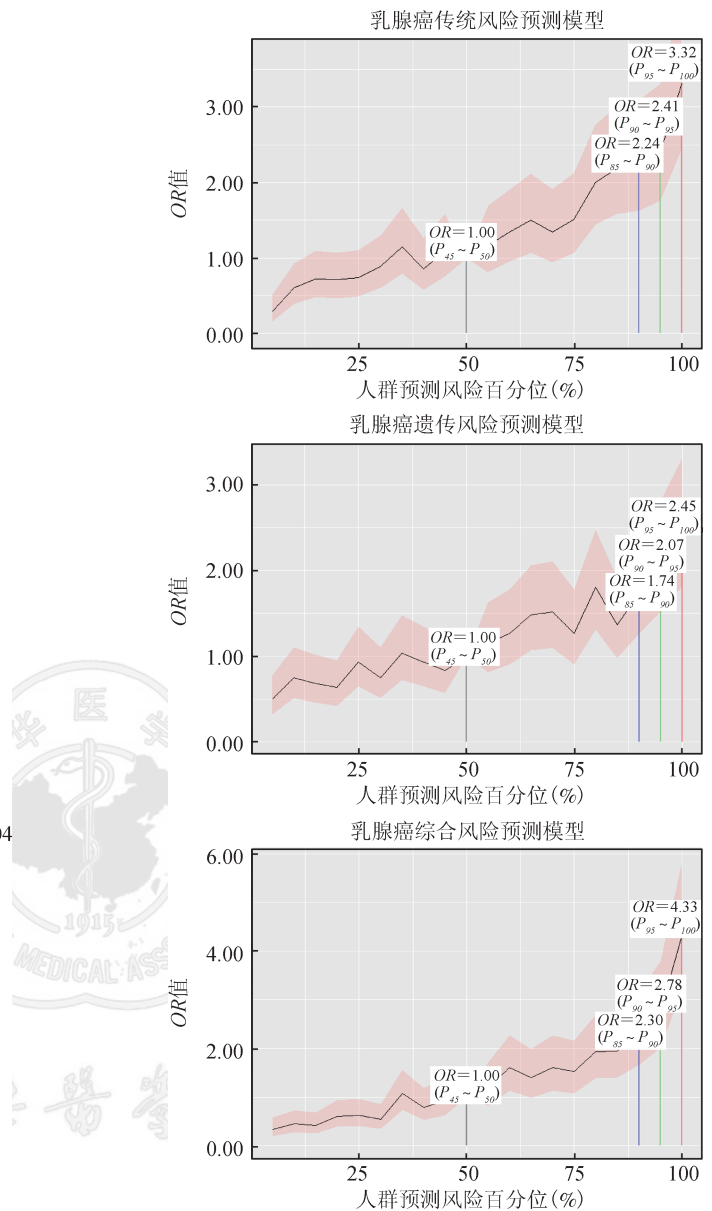


图1 200万模拟人群中病例及对照人群乳腺癌风险预测概率的分布

标SNP相对较少时,可放松SNP的初筛标准。

本研究存在局限性,首先,本研究是数据模拟研究,必然与真实世界研究存在一定的差异。其次,本研究选取的目标SNP只是有统计学意义的目标SNP,这些目标SNP之间是否存在生物学上的关联,目前仍在进一步的探讨中。第三,目前尚没有研究报道这些目标SNP与早期乳腺癌之间的关联,因此,本研究也无法探讨目标SNP对早期乳腺癌检出率等筛查效果指标的影响。

综上所述,本研究是一项数据模拟研究,但对我



注:参考人群为预测风险 $P_{45} \sim P_{50}$ 的人群($OR=1.00$)

图2 不同乳腺癌风险预测百分位的人群发生乳腺癌的相对风险

国未来开展更大规模的乳腺癌高危人群筛查,尤其是针对是否应该将GWAS发现SNP用于初筛乳腺癌高危人群,具有一定的参考价值。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] Antoniou AC, Wang XS, Fredericksen ZS, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population [J]. Nat Genet, 2010, 42 (10) : 885-892. DOI: 10.1038/ng.669.

[2] Gaudet MM, Kirchoff T, Green T, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer [J]. PLoS Genet, 2010, 6 (10) : e1001183. DOI:

- 10.1371/journal.pgen.1001183.
- [3] Turnbull C, Ahmed S, Morrison J, et al. Genome-wide association study identifies five new breast cancer susceptibility loci[J]. *Nat Genet*, 2010, 42(6): 504–507. DOI: 10.1038/ng.586.
- [4] Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*) [J]. *Nat Genet*, 2009, 41(5): 579–584. DOI: 10.1038/ng.353.
- [5] Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2 [J]. *Nat Genet*, 2009, 41(5): 585–590. DOI: 10.1038/ng.354.
- [6] Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer [J]. *Nat Genet*, 2008, 40(6): 703–706. DOI: 10.1038/ng.131.
- [7] Gold B, Kirchoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33 [J]. *Proc Natl Acad Sci USA*, 2008, 105(11): 4340–4345. DOI: 10.1073/pnas.0800441105.
- [8] Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer [J]. *Nat Genet*, 2007, 39(7): 870–874. DOI: 10.1038/ng.2075.
- [9] Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer [J]. *Nat Genet*, 2007, 39(7): 865–869. DOI: 10.1038/ng.2064.
- [10] Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci [J]. *Nature*, 2007, 447(7148): 1087–1093. DOI: 10.1038/nature05887.
- [11] Cai QY, Long JR, Lu W, et al. Genome-wide association study identifies breast cancer risk variant at 10q21. 2: results from the Asia breast cancer consortium [J]. *Hum Mol Genet*, 2011, 20(24): 4991–4999. DOI: 10.1093/hmg/ddr405.
- [12] Zheng W, Long JR, Gao YT, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1 [J]. *Nat Genet*, 2009, 41(3): 324–328. DOI: 10.1038/ng.318.
- [13] Cai QY, Zhang B, Sung H, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1 [J]. *Nat Genet*, 2014, 46(8): 886–890. DOI: 10.1038/ng.3041.
- [14] Long JR, Cai QY, Shu XO, et al. Identification of a functional genetic variant at 16q12. 1 for breast cancer risk: results from the Asia breast cancer consortium [J]. *PLoS Genet*, 2010, 6(6): e1001002. DOI: 10.1371/journal.pgen.1001002.
- [15] Long JR, Cai QY, Sung H, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer [J]. *PLoS Genet*, 2012, 8(2): e1002532. DOI: 10.1371/journal.pgen.1002532.
- [16] 黄育北, 宋丰举, 陈可欣. 乳腺癌全基因组关联研究的现状及其应用价值的分析方法 [J]. *中华流行病学杂志*, 2015, 36(10): 1058–1061. DOI: 10.3760/cma.j.issn.0254–6450.2015.10.004.
- Huang YB, Song FJ, Chen KX. Current status of genome-wide association studies (GWAS) on breast cancer and application values of single nucleotide polymorphisms identified from GWAS [J]. *Chin J Epidemiol*, 2015, 36(10): 1058–1061. DOI: 10.3760/cma.j.issn.0254–6450.2015.10.004.
- [17] Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer [J]. *J Natl Cancer Inst*, 2014, 106(11): dju305. DOI: 10.1093/jnci/dju305.
- [18] Wacholder S, Hartge P, Prentice R, et al. Performance of common genetic variants in breast-cancer risk models [J]. *N Engl J Med*, 2010, 362(11): 986–993. DOI: 10.1056/NEJMoa0907727.
- [19] Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model [J]. *J Natl Cancer Inst*, 2009, 101(13): 959–963. DOI: 10.1093/jnci/djp130.
- [20] Pharoah PD, Antoniou AC, Easton DF, et al. Polygenes, risk prediction, and targeted prevention of breast cancer [J]. *N Engl J Med*, 2008, 358(26): 2796–2803. DOI: 10.1056/NEJMSa0708739.
- [21] Mealiffe ME, Stokowski RP, Rhee BK, et al. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information [J]. *J Natl Cancer Inst*, 2010, 102(21): 1618–1627. DOI: 10.1093/jnci/djq388.
- [22] Zuo TT, Zheng RS, Zeng HM, et al. Female breast cancer incidence and mortality in China, 2013 [J]. *Thoracic Cancer*, 2017, 8(3): 214–218. DOI: 10.1111/1759–7714.12426.
- [23] 国家卫生和计划生育委员会. 中国卫生和计划生育统计年鉴 2014 [M]. 北京: 中国协和医科大学出版社, 2014.
- National Health Commission of the People's Republic of China. *China Health Statistics 2014* [M]. Beijing: China Union Medical University Press, 2014.
- [24] Zheng W, Wen WQ, Gao YT, et al. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women [J]. *J Natl Cancer Inst*, 2010, 102(13): 972–981. DOI: 10.1093/jnci/djq170.
- [25] Zheng W, Zhang B, Cai QY, et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls [J]. *Human Mol Genet*, 2013, 22(12): 2539–2550. DOI: 10.1093/hmg/ddt089.
- [26] Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk [J]. *J Natl Cancer Inst*, 2008, 100(14): 1037–1041. DOI: 10.1093/jnci/djn180.
- [27] Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond [J]. *Stat Med*, 2008, 27(2): 157–172. DOI: 10.1002/sim.2929.

(收稿日期: 2018–11–21)

(本文编辑: 万玉立)