PRS_{313}在洛杉矶生物样本库中对非欧洲裔人群乳腺癌风险的可推广性

作者

Helen Shang*[1,2], Yi Ding[3], Vidhya Venkateswaran[4], Kristin Boulier[5], Nikhita Kathuria-Prakash[6], Parisa Boodaghi Malidarreh[7,8], Jacob M. Luber,[7,8,9]Bogdan Pasaniuc[,10,11,12,13,14]

所属机构

[1]内科部，罗纳德·里根加州大学洛杉矶分校医疗中心，美国加利福尼亚州洛杉矶市 ，美国。[2]德克萨斯大学阿灵顿分校计算机科学与工程系，美国德克萨斯州阿灵顿。[3]加利福尼亚大学洛杉矶分校（UCLA）生物信息学跨系项目，美国加利福尼亚州洛杉矶。[4]加州大学洛杉矶分校牙科学院口腔生物学系，美国加利福尼亚州洛杉矶。[5]罗纳德·里根加州大学洛杉矶分校医疗中心医学系心脏病学部，美国加利福尼亚州洛杉矶。[6]罗纳德·里根加州大学洛杉矶分校医疗中心医学系血液肿瘤学部，美国加利福尼亚州洛杉矶。[7]德克萨斯大学阿灵顿分校计算机科学与工程系，美国德克萨斯州阿灵顿。[8]德克萨斯大学阿灵顿分校多专业健康信息学中心，美国德克萨斯州阿灵顿。[9]德克萨斯大学阿灵顿分校生物工程系，美国德克萨斯州阿灵顿。[10]加利福尼亚大学洛杉矶分校（UCLA）生物信息学跨系项目，美国加利福尼亚州洛杉矶。[11]加利福尼亚大学洛杉矶分校大卫·格芬医学院病理学与实验室医学系，美国加利福尼亚州洛杉矶。[12]加利福尼亚大学洛杉矶分校大卫·格芬医学院人类遗传学系，美国加利福尼亚州洛杉矶。

[13]加利福尼亚大学洛杉矶分校大卫·格芬医学院计算医学系，美国加利福尼亚州洛杉矶。
[14]加利福尼亚大学洛杉矶分校精准健康研究所，美国加利福尼亚州洛杉矶。

***通讯作者：hshang@mednet.ucla.edu**

摘要

多基因风险评分（PRS）汇总了常见风险变异的综合效应，并且与未发现可识别单基因风险因素的患者的乳腺癌风险相关。迄今为止，乳腺癌领域经过验证最充分的多基因风险评分之一是PRS313，该评分基于北欧生物样本库开发，但在非欧洲祖源人群中的表现有所减弱。我们进一步研究了PRS313在同一机构内欧洲裔（EA）、非洲裔（AFR）、亚洲裔（EAA）和拉丁裔（HL）美国女性中的可推广性；该机构采用统一的电子健康记录系统、基因分型平台和质量控制流程。我们发现，PRS313在拉丁裔女性（曲线下面积，0.68；95%置信区间，0.65–0.71）与欧洲血统女性（曲线下面积，0.70；95%置信区间，0.69–0.71）中的受试者工作特征曲线下面积（AUC）相近，但在非洲人群和东亚人群中的AUC较低（非洲人群：AUC，0.61；95%置信区间，0.56–0.65；东亚人群：AUC，0.64；95%置信区间，0.60–0.680）。虽然PRS313与欧洲裔美国人的激素受体阳性（HR+）疾病相关（比值比，1.42；95%置信区间，1.16–1.64），但对于拉丁裔女性，它可能与人表皮生长因子受体2（HER2+）疾病相关（比值比，2.52；95%置信区间，1.35–4.70）；不过，由于病例数较少，还需要进一步研究。总之，我们发现，在洛杉矶同一医疗卫生系统内，PRS313与乳腺癌显著相关，但在非洲裔和亚洲裔女性中的准确性有所降低。我们的研究进一步凸显，在临床应用多基因风险评分之前，需在多样化队列中开展更多验证。

引言

美国预防服务工作组（USPSTF）建议从50岁开始进行乳腺癌筛查；研究显示，90%的乳腺癌病例在这一年龄之后确诊[1]。遗憾的是，这也意味着按照传统指南，每年约有10%的病例会被漏诊，相当于美国每年约漏诊10,000例[2]。因此，研究人员一直在开发新的方法，以识别有早发性乳腺癌风险的患者。一种广为人知的方法是Gail模型，该模型利用临床信息、家族史和人口统计学信息计算个体乳腺癌风险，但其准确性较低；一项涵盖26项研究和29个数据集的荟萃分析显示，Gail模型的一个改良版本在美国女性、亚洲女性和欧洲女性中的AUC分别为0.61（95%置信区间0.59–0.63）、0.55（95%置信区间0.52–0.58）和0.58（95%置信区间0.55–0.62）[3,4]。

另一种途径是识别携带 BRCA1、BRCA2、PTEN、TP53、CDH1 和 STK11 等基因变异的人群；这些变异与患者一生中罹患乳腺癌的风险增加两至三倍有关[5]。然而，绝大多数（>75%）乳腺癌患者并无可识别的单基因风险因素，因此无法从这种方法中获益。过去十年间，多基因风险评分（PRS）被提出作为一种潜在解决方案，通过汇总多个常见风险变异的综合效应；这些变异各自与乳腺癌风险小幅升高有关[6]。目前已开发出多种多基因风险评分，用于预测和分层评估乳腺癌风险；近期一篇论文显示，某种多基因风险评分中处于最高 5% 百分位数的人群，其遗传风险与某些单基因病因所致的风险相当[7]。既往研究估计，如果所有常见单核苷酸多态性的效应大小都能确定无疑，那么理论上最佳的多基因风险评分可解释乳腺癌家族性风险的 ~41%[8]。

迄今，乳腺癌领域经过验证最多的多基因风险评分之一是Mavaddat等人提出的PRS_{313}，该评分基于北欧生物样本库开发（n = 33,673例病例和n = 33,381例对照）。纳入家族史和诊断年龄后，PRS_{313}的比值比为1.61（95%置信区间，1.57至1.65），曲线下面积为0.63（95%置信区间 =0.63-0.65）[9]。后续研究表明，PRS_{313}在

非洲裔女性中的效应减弱；例如，Cong等人发现，在9家机构的33,594名欧洲血统女性和2,801名非洲血统女性中，仅使用PRS_{313}时，欧洲裔女性的曲线下面积（0.60，95%置信区间，0.59-0.61）高于非洲裔女性（0.55，95%置信区间，0.51-0.58）[10,11]。

本文旨在进一步考察多基因风险评分$_{313}$在同一机构内对非洲裔、亚洲裔和拉丁裔血统美国女性的可推广性；该机构采用统一的电子健康记录系统、基因分型平台和质量控制流程。

材料与方法

研究参与者

本队列研究纳入了出生时为女性、来自 UCLA ATLAS 生物样本库（N=18,627）的参与者；该生物样本库与电子病历（EMR）数据相关联，且此前已有描述[12]。研究人员使用全基因组芯片对单核苷酸多态性进行基因分型，并将其填补至 TOPmed 参考面板。我们根据对应于 Phecode X 105.1 的 ICD-9 和 ICD-10 编码，识别乳腺癌病例和对照；该 Phecode 指定“女性乳腺恶性肿瘤”，用于将 ICD 编码映射为具有临床意义的表型 [12]。

我们此前基于主成分分析（PCA）和k均值聚类，确定了五种遗传推断祖源（GIA），包括非裔美国人（AA）、西班牙裔拉丁裔美国人（HL）、东亚裔美国人（EAA）、欧洲裔美国人（EA）和南亚裔美国人（SAA）[13]。由于病例数不足以支持有统计效力的分析，因此未将SAA人群纳入后续分析。由于无法获取个体层面的病理结果，我们根据与相应亚型相关的处方数据，确定激素受体阳性（HR+）或人表皮生长因子受体2阳性（HER2+）乳腺癌。

PRS 模型

多基因风险评分（PRS）采用加性模型计算，使用Mavaddat等人最初开发的多基因评分的效应值估计；该评分可在多基因评分（PGS）目录中的条目PGS000004下获取[14]：

公式1：

$$\mathrm{PRS} = \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k \ldots + \beta_n x_n$$

其中，β_k是与单核苷酸多态性 k 的次要等位基因相关的乳腺癌每个等位基因对数优势比（OR），而 x_k是同一单核苷酸多态性的次要等位基因数量（0、1 或 2），n = 313 是单核苷酸多态性的总数。由于存在歧义，排除了 30 个单核苷酸多态性，另有 45 个单核苷酸多态性未能匹配。与原始研究一样，将欧洲人群的原始多基因风险评分标准化为标准差为 1、均值为 0。我们

随后使用欧洲裔样本的均值和标准差对其余基因组推断祖源进行标准化，方法与先前检验多基因风险评分$_{313}$ [10,11]可推广性的研究相同。

基因分型

关于本队列基因分型、基因型填补和质量控制流程的详细信息此前已有描述[13]。在本研究中，符合以下3项标准的变异被保留用于计算PRS：(1) 各基因分型芯片批次中的填补质量R2均值大于0.3；(2) 在祖源特异性的哈代–温伯格平衡检验中，P值大于 1×10^{-6}；(3) 次要等位基因频率（MAF）大于0.005。随后，我们使用plink2（--indep-pairwise 1000 50 0.05）进行连锁不平衡剪枝，并排除了长程连锁不平衡区域。使用flashpca2软件计算了前九个主成分[15]。

统计学方法

我们使用逻辑回归模型，根据以下公式估计PRS与乳腺癌相关的优势比（OR），并将年龄和前九个主成分作为协变量：

公式2：

$$\log(\text{乳腺癌}) = \beta_0 + \beta_1(\text{多基因风险评分}) + \beta_2(\text{年龄}) + \beta_3(\text{PC1}) + \ldots + \beta_{11}(\text{PC9})$$

我们没有采用 Mavaddat等人使用的确诊年龄，而是使用患者病历中首次出现与乳腺癌对应的 ICD-9 或 ICD-10 编码时的年龄。这是因为我们的电子健康记录（EHR）并未将诊断年龄存储为结构化数据。出于同样的原因，我们也无法将家族史纳入协变量。为了与 Mavaddat等人的研究保持一致，我们还将前九个主成分纳入协变量，以控制不同祖源群体之间群体结构可能存在的差异。

采用包含相同协变量的逻辑回归，估计按多基因风险评分十分位数划分的乳腺癌比值比，并以中位数（第50百分位数）作为参照。分别针对每个GIA计算百分位数及其优势比。为评估每个GIA对应的多基因风险评分的判别能力，我们使用标准化多基因风险评分、生物样本库采样时年龄和前九个主成分作为预测变量，估计受试者工作特征曲线下面积（曲线下面积）。

生存分析

使用多基因风险评分作为总生存期（OS，单位为天）的唯一预测变量，并将其作为连续变量，进行了Kaplan-Meier分析。总生存期的计算方法是：从患者病历中首次出现与乳腺癌对应的ICD-9或ICD-10代码的日期，计算至死亡日期或当前日期。分析中剔除了没有总生存期数值的患者（N=12）。

结果

本研究共纳入18,627名女性，其中1,156名具有非洲血统（非洲裔美国人），11,873名具有欧洲血统（欧洲裔美国人），2,010名具有东亚血统（东亚和非洲裔美国人），3,303名具有西班牙裔血统（西班牙裔/拉丁裔），详见表1。大多数病例被鉴定为HR+，各祖源人群中的患病率与最新SEER登记数据相当；该数据显示，70%的乳腺癌亚型为HR+[16]。同样，我们发现HER2+ 疾病的患病率也与SEER登记数据相当，后者显示患病率为10.8%。在2,080个病例中，有1,285个没有处方数据，因为这些患者可能在加州大学洛杉矶分校医疗网络之外接受了癌症治疗。

多基因风险评分$_{313}$与不同祖源人群乳腺癌风险的关联

如图1所示，各基因组推断祖源组的多基因风险评分均呈正态分布，其中东亚和非洲裔美国人队列的均值较高（非洲裔美国人：N= 1156；均值：-0.02；标准差：0.46；欧洲裔美国人：N= 11,873；均值：-0.06；标准差：0.53；东亚和非洲裔美国人：N= 2010；均值：0.15；标准差：0.47；西班牙裔/拉丁裔：N= 3,303；均值：-0.06；标准差：0.53）。我们发现，多基因风险评分$_{313}$与所有祖源人群的总体乳腺癌风险均存在统计学显著关联。所有基因组推断祖源组的优势比均有重叠（非洲裔美国人：比值比，1.31；95%置信区间，1.05-1.64；欧洲裔美国人：比值比，1.52；95%置信区间，1.23-1.72；东亚和非洲裔美国人：比值比，1.46；95%置信区间，1.23-1.61；西班牙裔/拉丁裔：比值比，1.51，95%置信区间，1.31-1.75）。这些结果也与Mavaddat等人报告的欧洲人比值比（比值比，1.61；95%置信区间，1.57至1.65）重叠。

对于所有基因组推断祖源，PRS$_{313}$的比值比在PRS分布两端相对于第50百分位数最大，且四种基因组推断祖源的95%置信区间彼此重叠（非洲裔美国人：比值比，1.85；95%置信区间，1.08-3.2；欧洲裔美国人：比值比，2.1；95%置信区间，1.9-2.6；东亚和非洲裔美国人：比值比，1.7；95%置信区间，1.1-2.6；西班牙裔/拉丁裔：比值比，1.7；95%置信区间，1.1-2.5：）。对于非洲裔美国人群体，我们发现，与其他基因组推断祖源相比，PRS的效应有所减弱，且在其分布两端（>95 百分位数）时亦如此，不过这项分析受样本量较小的限制（未显示）。

PRS$_{313}$的判别准确性

通过曲线下面积（AUC）衡量的多基因风险评分对任何类型乳腺癌的区分准确性，在欧洲人群中最高，为0.70（95%置信区间 = 0.69至0.71），略高于Mavaddat等人报告的0.63（95%置信区间 = 0.63至0.65）。与欧洲人群相比，西班牙裔/拉丁裔人群的AUC相近，而那些

对于非洲人群和东亚和非洲裔美国人群，曲线下面积均较低（非洲人群：曲线下面积，0.61；95%置信区间，0.56–0.65；东亚和非洲裔美国人群：曲线下面积，0.64；95%置信区间，0.60–0.680；西班牙裔/拉丁裔：曲线下面积，0.68；95%置信区间，0.65–0.71）。

PRS_{313}与不同祖源人群乳腺癌亚型的关联

近期针对欧洲裔女性的研究表明，HR+乳腺癌与PRS_{313}及较低风险的遗传特征[17]相关。与既往研究一致，我们发现，对于欧洲患者，PRS_{313}与HR+ 疾病相关；然而，在患有乳腺癌的西班牙裔、非洲裔和亚裔美国女性中，PRS_{313}与HR+疾病均无关联（表2）。相比之下，仅在西班牙裔美国人中，HER2+疾病与PRS_{313}相关，且其比值比高于总体乳腺癌风险（比值比，2.47；95%置信区间，1.39 - 4.41）。不过，由于我们的队列中兼具西班牙裔身份和HER2+特征的患者仅有18例，因此这项分析受到限制。我们还注意到，HER2+西班牙裔患者的平均诊断年龄明显低于其他族群（非洲裔美国人均值：60（标准差：15）；欧洲裔美国人均值：56（标准差：14）；东亚和非洲裔美国人均值：51（标准差：11）；西班牙裔/拉丁裔均值：44（标准差：7.7）），因此我们考察了诊断年龄是否可能造成这些结果的混杂；在其原始研究中，Mavaddat等人显示，相对风险会随年龄和PRS_{313}而下降。然而，即使考虑诊断年龄，PRS_{313}与HER2+疾病的关联仍然存在（比值比，2.52；95%置信区间，1.35-4.70）。

生存分析

为确认我们的OS估计可靠，我们首先确认OS与患者是否接受过化疗恰当相关，因为化疗是疾病侵袭性更强且通常已发生转移的标志（补充图1）。我们发现，在欧洲患者中，单独以PRS_{313}作为预测因子时，Kaplan-Meier分析无法按生存时间对患者进行分层：比较PRS_{313}高于（N=280）或低于（N=651）第70百分位数的乳腺癌患者，所得对数秩检验P值为0.38（补充图2a）。我们任意选取第70百分位数，是因为我们希望选择一个阈值，使PRS_{313}的比值比高于第50百分位数处的比值比，如图1所示。值得注意的是，我们评估其他阈值时也发现了类似结果，例如比较最高与最低的50个百分位数，以及最高的90个百分位数。

与最低第10百分位数相比。对于欧洲患者，在考虑患者是否接受过化疗、诊断年龄、HER2+以及PR+ 疾病后，使用多变量Cox比例风险模型，PRS_{313}也不再具有总生存期的预测性（补充图2b）。这与近期一项涉及欧洲人（N=98,397）和亚洲人（N=12,920）乳腺癌女性的队列研究结果一致，该研究发现，在欧洲患者中，在根据乳腺癌亚型和肿瘤分级进行校正后，PRS_{313}不再与总生存期相关[17]。

讨论

在本文中，我们研究了同一机构内非洲裔、亚洲人和拉丁裔血统的美国女性中，PRS_{313}的可推广性。对于欧洲血统的女性，我们得到了与Mavaddat等人（1.61；95%置信区间，1.57-1.65）相近且相互重叠的比值比估计值（1.5；95%置信区间，1.2-1.7）。与既往研究结果一致，我们发现，PRS_{313}在所有祖源人群中仍与乳腺癌相关，但在非洲血统和亚洲血统女性中的效应有所减弱；PRS_{313}在拉丁裔（曲线下面积，0.68；95%置信区间，0.65-0.71）和欧洲血统女性（曲线下面积，0.70；95%置信区间，0.69-0.71）中取得了相当的曲线下面积，而在非洲人群和EAA人群中，曲线下面积则较低（非洲人群：曲线下面积，0.61；95%置信区间，0.56-0.65；EAA：曲线下面积，0.64；95%置信区间，0.60-0.680）。这可能是因为拉丁裔个体的欧洲血统比例高于亚洲人和非洲裔个体；此前，我们发现UCLA ATLAS队列中的拉丁裔个体具有复杂的遗传混合特征，其主成分分析亚结构以欧洲裔样本为中心，并向非洲人群和亚洲人群延伸出分支[13]。

与既往研究一致，我们发现，对于欧洲裔美国患者，PRS_{313}与HR+ 疾病[9,17]相关。对于非洲裔、西班牙裔和亚裔美国人，这种关联不复存在，尽管这些队列中的HR+ 疾病发生率低于欧洲裔美国人队列。令人意外的是，我们发现，仅在西班牙裔美国人中，PRS_{313}与HER2+ 疾病相关，此前尚未有相关报道。西班牙裔患者中HER2+ 疾病患病率较高，并不能完全解释这一发现，因为我们发现，在本队列中，西班牙裔的HER2+疾病发生率低于或相当于其他族群。此外，在校正诊断年龄后，这种关联仍然存在。我们注意到，与其他族群相比，西班牙裔患者确诊乳腺癌的年龄更小；此前已有研究报道这一点。Mavaddat等人在其原始论文中也指出，随着年龄增长，PRS_{313}[18]的相对风险会下降。然而，鉴于我们的队列规模有限，仍有必要对此关联开展进一步研究。

与近期关于 PRS_{313}，的研究结果类似，我们发现它无法对欧洲患者的 OS[15] 进行风险分层。一种可能的解释是，PRS_{313} 对 OS 的影响可能受到其他遗传风险因素的混杂，其中许多因素尚未被发现；

近期多篇论文发现，多基因风险评分$_{313}$能够对*CHEK2*、*PALB2*和*ATM* 携带者的乳腺癌风险进行分层，但无法对*BRCA1/2*携带者进行分层[19,20]。换言之，多基因风险评分$_{313}$的价值可能在于对低外显率风险变异的携带者进行风险分层，但对于携带高外显率变异的人群，其分层能力可能有限，因为无论如何，这些人最终都会罹患乳腺癌。鉴于我们的队列中风险变异携带者人数较少，我们无法在本研究中证实这一点，但希望在未来的研究中对其进行验证。

本研究有许多局限性需要考虑。我们的队列中非欧洲裔样本少于欧洲裔样本，因此无法充分分析极端高值区间及各亚型内的情况。由于缺乏纵向结局数据，我们也无法估算罹患乳腺癌的绝对风险。此外，许多协变量，例如确诊年龄、癌症亚型和总生存期，并未以结构化数据的形式记录在病历中，因此只能通过替代方法进行计算。尽管如此，我们仍能确认这些估算得出了预期的观察结果，表明其具有可靠性，例如乳腺癌各亚型中预期的患病率，以及化疗作为总生存期预测因素时所呈现的预期关联。

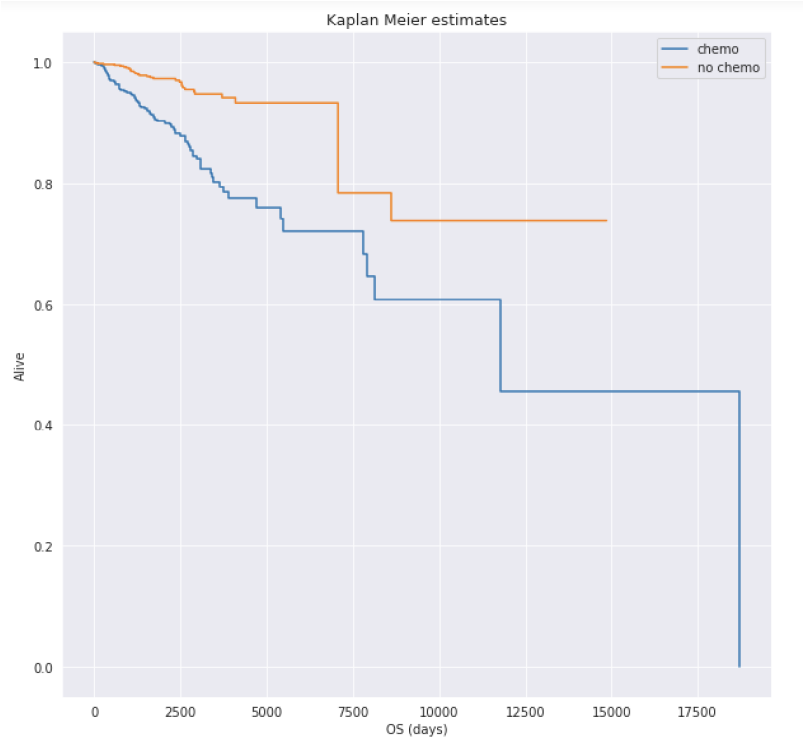
总之，我们发现，在洛杉矶一家规模较小但祖源多样的生物样本库中，多基因风险评分$_{313}$与不同祖源的美国女性乳腺癌显著相关，但在非洲裔和亚洲裔女性中的准确性有所降低。虽然多基因风险评分$_{313}$与欧洲裔美国人的HR+ 型疾病相关，但这种关联在非洲裔、西班牙裔美国人和亚裔美国人中并未观察到。对于西班牙裔美国人，多基因风险评分$_{313}$可能与HER2+ 型疾病相关；不过，由于病例数量较少，仍需开展更多研究来验证这些发现。我们的研究进一步凸显了在临床应用多基因风险评分之前，需在多样化队列中进行更多验证；同时也表明，有必要开发能够解决基因组混合祖源差异的新方法。

补充列表

补充图1：欧洲患者接受或未接受化疗治疗的总生存期Kaplan-Meier曲线 补充图2：欧洲患者的多基因风险评分313高于或低于第70百分位数时的总生存期Kaplan-Meier曲线 补充图3：欧洲患者总生存期的多变量Cox比例风险模型 补充表1a：用于识别HR+ 型乳腺癌病例的药物 补充表1b：用于识别HER2+ 型乳腺癌病例的药物 补充表2：基因组推断祖源中的多基因风险评分与乳腺癌风险之间的关联

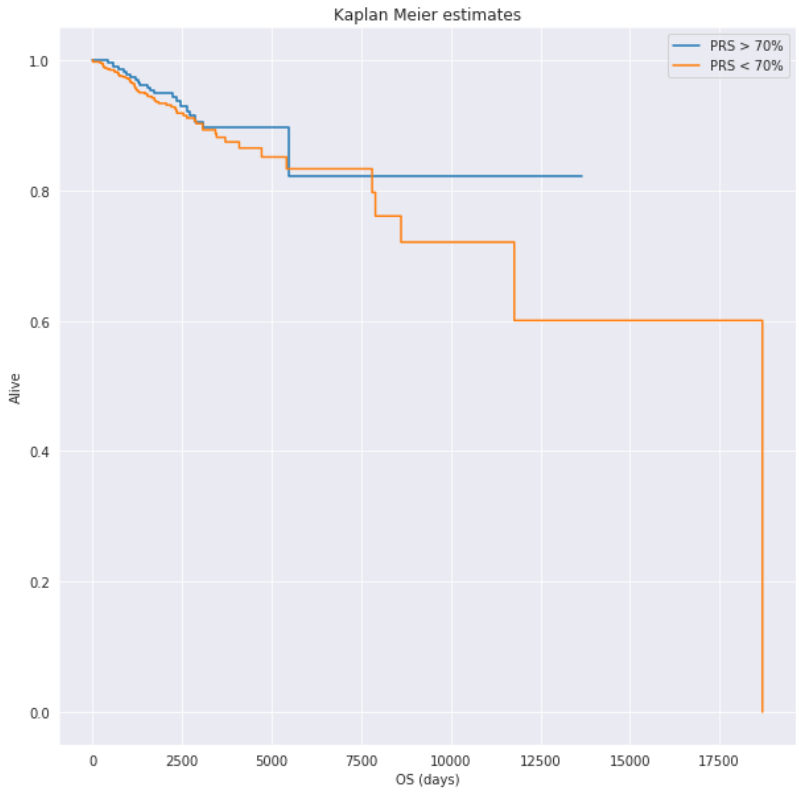
补充图1：欧洲患者接受或未接受化疗治疗的总生存期Kaplan-Meier曲线

对于仍在世的患者，我们用当前日期减去诊断日期；对于已死亡的患者，则用死亡日期减去诊断日期，以此近似计算总生存期（OS），单位为天。为验证这一方法，我们确认，所有接受过乳腺癌化疗的患者，其近似总生存期均比未接受化疗的患者短（P值 < .005）。这是预期的，因为接受化疗的患者通常病情更具侵袭性或患有转移性疾病，因此生存期更短。我们的队列中有503名患者接受过化疗，767名患者未接受化疗。



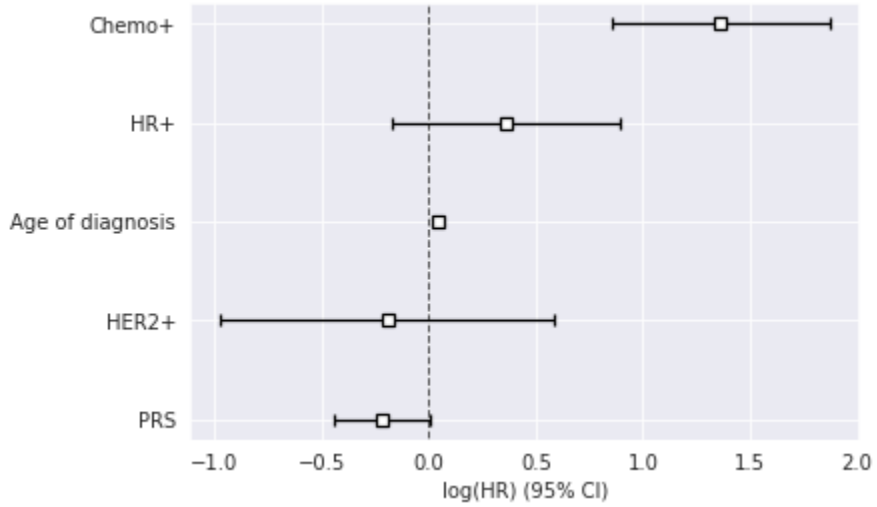
补充图 2：PRS_{313}第 70 百分位数以上或以下欧洲患者的总生存期 Kaplan-Meier曲线

为评估PRS_{313}对总生存期的影响，我们通过Kaplan-Meier分析比较了PRS高于（N=280）和低于（N=651）第70百分位数的欧洲患者的生存时间。通过对数秩检验，我们发现两组的生存时间无差异（p值= 0.38）。比较高于和低于第50百分位数的患者，以及处于最高10%和最低10%百分位数的患者时，生存时间也没有差异。



补充图3：欧洲患者总生存期的多变量Cox比例风险模型

对于欧洲患者，我们最初通过 Cox 比例风险模型发现，标准化 PRS_{313} 对 OS 的预测呈负相关，这表明较低的 PRS_{313} 评分意味着更长的生存时间（HR，0.80；95%置信区间，0.64-0.99，p 值 = 0.04）。然而，在对患者是否接受过化疗、癌症亚型（HR+ 和/或 HER2+）以及诊断年龄等其他变量进行校正后，标准化 PRS_{313}评分不再具有预测 OS 的能力（p 值 = 0.06）。



补充表1a：用于识别HR乳腺癌的药物+补充表1b：用于识别HER乳腺癌的药物2+ 乳腺癌我们首先查询了队列电子病历中处方频率最高的50种癌症相关药物。随后对这些药物进行了人工审核，并根据其药物类别及亚型相关性进行分组。补充表1a列出了我们归类的HR+ 相关药物，补充表2a列出了我们归类的HER2+ 相关药物。如果患者至少接受过两份药物清单中每份清单所列的一种相关药物的处方，则将其归类为HR+ 和/或 HER2+ 亚型。

Name	药物类别
依西美坦	激素治疗
阿那曲唑	激素治疗
来曲唑	激素治疗
他莫昔芬	激素治疗
雷洛昔芬	激素治疗
氟维司群	激素治疗
哌柏西利	CDK 4/6抑制剂
瑞波西利	CDK 4/6抑制剂
阿贝西利	CDK 4/6抑制剂
依维莫司	mTOR抑制剂
阿培利司	PI3K抑制剂

Name	药物类别
曲妥珠单抗	HER2 单克隆抗体
帕妥珠单抗	HER2 单克隆抗体
奈拉替尼	酪氨酸激酶抑制剂
拉帕替尼	酪氨酸激酶抑制剂
鲁索替尼	酪氨酸激酶抑制剂
奥希替尼	酪氨酸激酶抑制剂
图卡替尼	酪氨酸激酶抑制剂

尼拉帕利	酪氨酸激酶抑制剂
达沙替尼	酪氨酸激酶抑制剂

补充表2：遗传推断祖源（GIA）中多基因风险评分与乳腺癌风险的关联

采用单变量逻辑回归评估了多基因风险评分[313]与美国非洲裔（AA）、欧洲裔（EA）、东亚裔美国人（EAA）和西班牙裔/拉丁裔（HL）遗传推断祖源（GIA）女性中观察到的所有乳腺癌发病率之间的关联。与此前检验多基因风险评分[313]，可推广性的研究一样，我们依据欧洲裔样本的均值和标准差，对非欧洲裔GIA的原始多基因风险评分进行了标准化。相比之下，Mavaddat等人报告的比值比为1.61（95%置信区间1.57–1.65），与所有GIA的结果均有重叠。

GIA	OR	置信区间下限	置信区间上限
AA	1.31	1.05	1.64
EA	1.52	1.23	1.72
EAA	1.46	1.43	1.61
HL	1.51	1.31	1.75

致谢

我们衷心感谢精准健康研究所、UCLA ATLAS精准健康生物样本库的参与患者、加州大学洛杉矶分校大卫·格芬医学院、加州大学洛杉矶分校临床与转化科学研究所，以及加州大学洛杉矶分校医疗系统。VV的经费由NIH/NIDCR 5K12DE027830-04提供。ATLAS社区健康倡议由加州大学洛杉矶分校医疗系统、大卫·格芬医学院以及加州大学洛杉矶分校临床与转化科学研究所的拨款（UL1TR001881）支持。BP的部分经费由NIH项目R01 HG009120、R01 MH115676、R01 CA251555、R01 AI153827、R01 HG006399、R01 CA244670和U01 HG011715提供。KB由加州大学洛杉矶分校T32项目资助。

利益声明

None

参考文献

1. Siu AL (2016). 乳腺癌筛查：美国预防服务工作组建议声明。《内科学年鉴》*6*，279-96.10.7326/M15-2886。2. Acheampong T, Kehm RD, Terry MB, 等 (2020). 2010至2016年美国不同年龄及种族/族裔群体乳腺癌分子亚型的发病趋势。《JAMA网络开放》，*8*。10.1001/jamanetworkopen.2020. 3. Gail MH, Brinton LA, Byar DP, 等 (1989). 对接受年度检查的白人女性患乳腺癌的个体化概率进行预测。《美国国家癌症研究所杂志》*24*，1879-86。0.1093/jnci/81.24.1879。4. Wang, X., Huang, Y., Li, L. 等 (2018). Gail模型预测乳腺癌风险的表现评估：系统综述、荟萃分析及试验序贯分析。《乳腺癌研究》*20*，18。10.1186/s13058-018-0947-5。5. Shiovitz S, Korde LA (2015). 乳腺癌遗传学：一个不断发展的课题。《肿瘤学年鉴》*26*，1291-9。10.1093/annonc/mdv022。

6. Yanes, T., Young, MA., Meiser, B. 等（2020）。乳腺癌多基因风险的临床应用：新兴领域的批判性综述与展望。《乳腺癌研究》*22*，21。10.1186/s13058-020-01260-3。7. Jia G，Lu Y，Wen W，Long J，等（2020）。评估多基因风险评分在识别八种常见癌症高风险个体中的效用。《JNCI癌症谱》4。10.1093/jncics/pkaa021。8. Michailidou K，Lindström S，Dennis J，等（2017）。关联分析确定65个新的乳腺癌风险位点。《自然》*551*，92-94。10.1038/nature24284 9. Mavaddat N，Michailidou K，Dennis J，等（2019）。用于预测乳腺癌及乳腺癌亚型的多基因风险评分。《美国人类遗传学杂志》*104*，21-34。10.1016/j.ajhg.2018.11.002。10.Liu C，Zeinomar N，Chung WK，等（2021）。具有欧洲人、非洲裔和拉丁裔血统女性的乳腺癌多基因风险评分的可推广性。《JAMA网络开放》，*4*。10.1001/jamanetworkopen.2021.19084 11. Du Z，Gao G，Adedokun B，等（2021）。评估非洲血统女性的乳腺癌多基因风险评分。《美国国家癌症研究所杂志》*113*，1168-1176。10.1093/jnci/djab050。12.Wu P，Gifford A，Meng X，等（2019）。将ICD-10和ICD-10-CM编码映射至表型编码：工作流程开发与初步评估。《JMIR医学信息学》*7*。10.2196/14325。PMID：31553307；PMCID：PMC6911227。13.Johnson R，Ding Y，Venkateswaran V，等（2022）。利用基因组多样性开展电子健康记录关联生物样本库中的发现：加州大学洛杉矶分校ATLAS社区健康倡议。《基因组医学》*1*，104。10.1186/s13073-022-01106-x。14.Samuel A. Lambert，Laurent Gil，Simon J upp，等（2021）。多基因评分目录：作为可重复性研究与系统评估的开放数据库。《自然遗传学》*4*，420-425。10.1038/s41588-021-00783-5。15.Abraham G，Inouye M（2014）。大规模全基因组数据的快速主成分分析。《科学公共图书馆·综合》*4*。doi：10.1371/journal.pone.0093766。16.Howlader N，Altekruse SF，Li CI，等（2014）。由联合激素受体和HER2状态定义的乳腺癌亚型的美国发病率。《美国国家癌症研究所杂志》*5*。10.1093/jnci/dju055。

17.Lopes Cardozo JMN, Andrulis IL, Bojesen SE 等。（2023）。乳腺癌协会联盟和 MINDACT 合作者。《多基因乳腺癌风险评分与肿瘤特征及生存的关联》。临床肿瘤学杂志*10*,1849-1863. 10.1200/JCO.22.01978. 18.Parada H Jr, Vu AH, Pinheiro PS 等。（2021）。美国西班牙裔与非西班牙裔白人的癌症诊断年龄比较。《癌症流行病学生物标志物与预防》*10*,1904-1912. 10.1158/1055-9965.EPI-21-0389. 19.Gallagher S, Hughes E, Wagner S 等。（2020）。多基因风险评分与高风险和中等风险乳腺癌基因女性携带者乳腺癌的关联。《JAMA网络开放》*7*. 10.1001/jamanetworkopen.2020.8501 20.Lakeman IMM, Rodríguez-Girondo MDM, Lee A 等。（2023）。多基因风险评分在家族性病例乳腺癌风险预测中的临床适用性。医学遗传学杂志*4*, 327-336. 10.1136/jmg-2022-108502.

表格和图表

表 1：参与者特征

参与者为来自 UCLA ATLAS 生物样本库的女性（N=18,627），该样本库与加州大学洛杉矶分校自 2013 年至今的病历相连。根据对应乳腺癌的 ICD-9 和 ICD-10 编码确定病例和对照。诊断年龄依据 ICD 编码出现在患者病历中的日期确定，并据此计算总生存期（OS），终点为死亡日期或当前日期。乳腺癌亚型根据所开具的处方确定。

	AA	EA	EAA	HL
对照	1032	10407	1803	3083
病例（占 % 的♀♀♀♀总计）	124（10）	1466（12）	207（10）	220（6.7）
诊断年龄，均值（标准差）	55（17）	55（17）	52（17）	48(17)
病例伴有处方数据（占总数的百分比）	86（52）	1241（58）	188（64）	166（58）
HR+ （病例百分比）	60（71）	979（72）	129（65）	115（65）
使用处方药数据）				
HER2+（占病例伴有处方数据）	11（12）	130（10）	26（14）	18（11）
OS（天），平均值 (SD)	2862 (1922)	2741 (1973)	2569（1427）	2911（2149）

图1：病例和对照中PRS_{313}的分布

按遗传推断祖源（GIA）分组的病例和对照中PRS_{313}评分的核分布估计图。欧洲人群（EA）的原始PRS_{313}评分已标准化至标准差为1、均值为0。其余GIA均按EA样本的均值和标准差进行标准化。

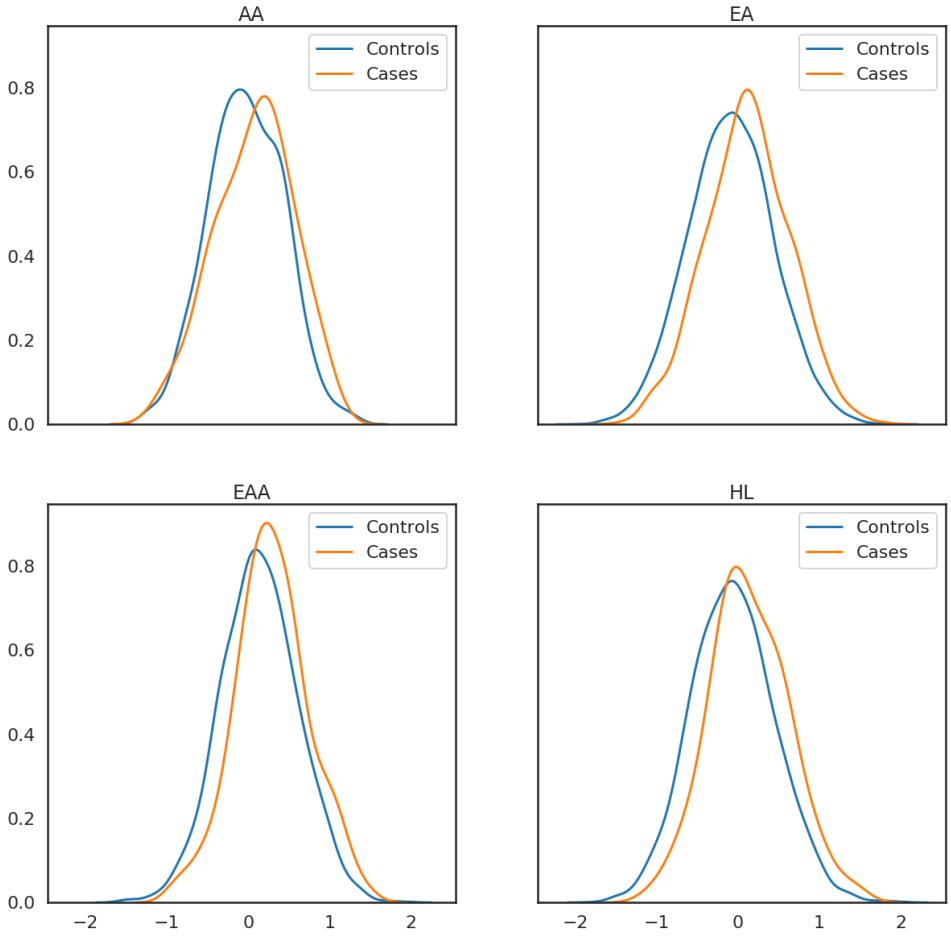


图2：PRS_{313}十分位数与乳腺癌的关联（相对于第50百分位数）

基于遗传推断祖源，非洲裔（AA）、欧洲裔（EA）、东亚裔美国人（EAA）和西班牙裔（HL）美国女性的PRS_{313}与乳腺癌诊断之间的关联。图中显示比值比及其95%置信区间。比值比表示PRS不同十分位数相对于PRS第 50^{th}百分位数的比值。

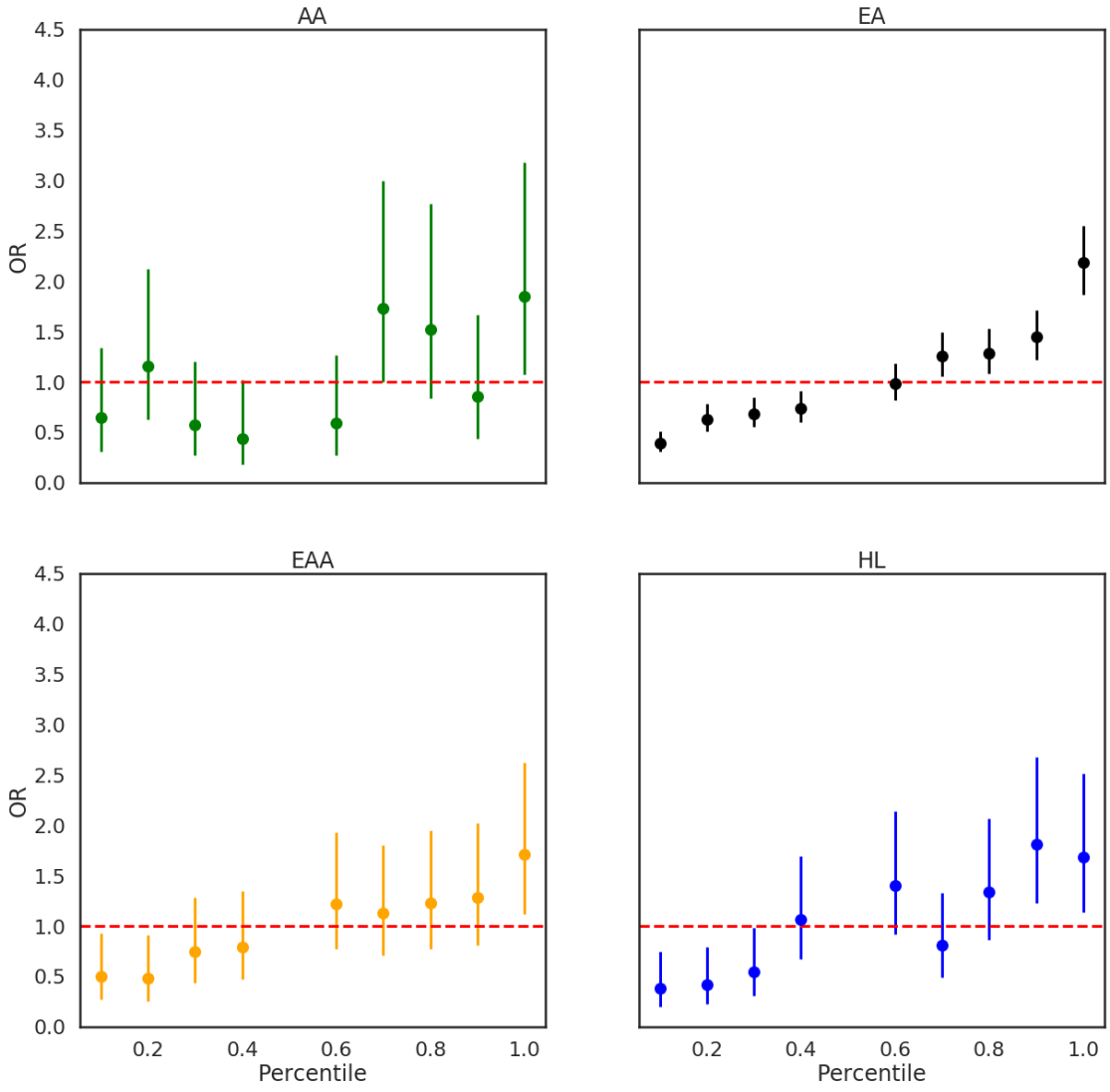


表2：按遗传推断祖源分组的PRS_{313}与HR+ 和HER2+ 乳腺癌的关联。

基于遗传推断祖源，分析非裔（AA）、欧洲裔（EA）、东亚裔美国人（EAA）和西班牙裔（HL）美国女性中PRS_{313}与HR+ 和HER2+ 的关联。上表显示，PRS_{313}仅与欧洲裔美国患者的HR+乳腺癌相关（p值 = 0.0002），与其他祖源人群则无关。相比之下，下表显示，PRS_{313}与HER2+ 乳腺癌相关，但仅见于西班牙裔美国患者（p值 = 0.002）。即使校正诊断年龄，PRS_{313}在西班牙裔美国人中仍能显著预测HER2+ 乳腺癌（比值比，2.52；95% 置信区间，1.35-4.70；p值 = 0.003）。

	HR-	HR+	OR	置信区间下限	置信区间上限
AA	15	51	0.71	0.36	1.40
EA	176	768	1.42	1.16	1.64
EAA	22	114	1.38	0.41	1.20
HL	33	106	0.70	0.91	2.20

	HER2阴性	HER2+	OR	置信区间下限	置信区间上限
AA	55	11	2.26	0.96	5.33
EAS	814	130	0.95	0.79	1.15
EAA	110	26	1.02	0.62	1.66
HL	121	18	2.47	1.39	4.41