

# 基于 CTSV 基因表达的三阴性乳腺癌健康管理的研究

徐文聪<sup>1</sup> 王孝娟<sup>1</sup> 黄嘉欣<sup>1</sup>

(1. 广东东软学院 健康医疗科技学院, 广东佛山 528200)

**摘要:** 三阴性乳腺癌发病率不断增加对病人生存质量的影响越来越严重, 在此阶段迫切需要找到一种有效的早期预警方法及健康管理策略。本研究从生物信息学和健康管理学的角度出发, 分别对三阴性乳腺癌 CTSV 基因差异表达情况以及与病人生存状态之间的联系进行了分析, 明确提出肥胖、慢性炎症和 CTSV 基因表达的关系, 并将所获得的成果转化运用到健康管理中。分析结果支持 CTSV 是预测三阴性乳腺癌患者不良预后的有效指标的观点, 且三阴性乳腺癌晚期患者的死亡风险会因为 CTSV 基因的过度表达而突然增加。将 CTSV 在三阴性乳腺癌中的表达与健康管理的三级预防策略(筛查-监测-治疗)相结合, 提出有针对性解决思路可一定程度上优化三阴性乳腺癌患者的全程管理, 为三阴性乳腺癌高危人群制定健康管理策略提供理论支持。

**关键词:** 生物信息学; CTSV; 三阴性乳腺癌; 健康管理

DOI: doi.org/10.70693/cjmsr.v1i3.1616

## 1. 背景

三阴性乳腺癌(Triple-Negative Breast Cancer, TNBC)是一种在年轻女性身上更常见的特殊乳腺癌类型, 其最明显的特点是癌组织中的雌激素受体(ER)、孕激素受体(PR)以及人表皮生长因子受体-2(HER-2)都显示阴性的结果。三阴性乳腺癌能够表现出较高的侵袭性和转移风险, 因此患者的预后通常不太乐观。相对于其余各类乳腺癌, 它与遗传性疾病的联系更为紧密<sup>[1]</sup>。三阴性乳腺癌由于具有高度侵袭性及缺少特定治疗靶点的特殊生物学特性, 导致目前尚缺乏行之有效的标准化治疗方案, 已成为乳腺癌临床研究与治疗领域中一个急需取得突破性进展的重要难题<sup>[2]</sup>。三阴性乳腺癌因缺乏特异性生物标志物使其临幊上诊断和预后面临巨大挑战。

CTSV 的表达有一定的组织选择性, 多发生在胸腺、睾丸和角膜上皮等位置<sup>[3]</sup>。从临床试验中可发现, CTSV 不仅可以影响到正常被试者的正常的生理功能, 也能对患病个体的病理状态有作用。在正常的生理条件下, CTSV 能对抗由血管紧张素 II 引发的动脉重塑及高血压现象<sup>[3]</sup>, 这一有限的特殊作用可以证明它真真切切地参与了人体的某些特定活动。

现阶段学术界对 CTSV 基因的研究一直较为活跃, CTSV 基因研究方面也涉猎甚广, 但探究其在三阴性乳腺癌中的影响的研究较少。而当前乳腺癌的健康管理更多的依赖于对患者生活方式的干预和基于相关风险预测模型进行管理, 所以持续开发新的健康管理的生物标志物具有一定的意义。本文主要探讨 CTSV 基因在三阴性乳腺癌中的表达模式及其在健康管理中的意义, 旨在系统探讨三阴性乳腺癌患者的健康管理模式, 尝试为临床实践提供新的研究视角, 具有一定的理论借鉴意义。

## 2. 研究方法

### 2.1 原始数据获取

TCGA 是一个联合发起的国际性合作项目, 其发起组织为美国国立卫生研究院(NIH)和国立癌症

**作者简介:** 徐文聪(1996—), 男, 硕士, 助教, 研究方向为肿瘤健康管理、健康政策与健康促进;

王孝娟(1991—), 女, 硕士, 助教, 研究方向为生物信息学、医学信息工程;

黄嘉欣(2002—), 女, 本科在读, 研究方向为健康服务与管理。

**通讯作者:** 赵晓沁

研究所（NCI），主要的作用是分析不同类型癌症的表观组学、基因组学和转录组学数据，其坚实的数据基础为癌症的预防、诊断和治疗等研究提供更多的可能性。该项目通过收集大规模肿瘤和正常组织样本的全基因组序列、DNA 甲基化、基因表达谱、蛋白质组学及临床数据，构建了一个开放的公共数据库平台<sup>[4]</sup>。本研究所使用到的 2222 份的 RNA-seq 数据以及 1087 份的临床信息数据都是属于 TCGA 数据库的乳腺癌（BRCA）相关数据。

## 2.2 数据的筛选与合并

本次研究采用了 R 语言（4.4.2 版本）对 TCGA 数据库中的 mRNA-seq 数据进行了全面的数据标准化处理，将 CTSV 基因的 tpm\_unstranded 表达值数据从原始的 mRNA-seq 数据中抽取出来，通过去除其中异常表达或者空白表达样本的手段清理数据<sup>[5]</sup>。采用 TCGA-key 编码实现临床数据和清洗后的 mRNA-seq 数据的配对连结，并最终筛选出满足分析要求与符合研究需要的患者样本。

## 2.3 差异表达分析

使用 R 语言通过非参数检验的方法 Kruskal-Wallis 检验，评估 CTSV 基因在三阴性乳腺癌与非三阴性乳腺癌患者中的表达差异。利用 stats 包的 kruskal.test() 函数，以病理分型为分组变量（即分为三阴性乳腺癌与非三阴性乳腺癌组）、CTSV 基因的 tpm\_unstranded 表达值作为因变量，分析经过预处理的数据集并检验原假设为两组间表达分布无显著差异（显著性阈值  $\alpha = 0.05$ ）<sup>[6]</sup>。通过计算两组中位表达量的对数比值得出差异倍数，避免极端值影响分析结果。通过使用 ggplot2 包以病理分型为横轴、tpm\_unstranded 表达值为纵轴，绘制箱线图直观展示差异。

## 2.4 生存分析

此项分析利用 Kaplan-Meier 曲线和 Cox 比例风险模型评估 CTSV 的表达如何影响患者的预后情况，依据患者自确诊到死亡或者最后一次随访时间来定义病人的生存时间（OS），用二元变量对生存状态（Status）进行标记。为了平衡样本量减少极端值的干扰，根据中位数将 CTSV 的表达水平分成高、低两组<sup>[5]</sup>。采用 Log-rank 检验法完成单因素分析比较各组之间的生存差异，多因素分析通过 Cox 模型对年龄，病理分期等协变量进行调节，从而确定预后因素并评价 CTSV 表达与生存之间的独立关系。

## 2.5 健康管理因素关联分析

基于差异表达分析与生存分析的结果，以影响癌症患者预后与治疗效果的健康管理因素为切入点，着重对三阴性乳腺癌病人进行健康管理因素关联分析。本研究旨在优化患者预后，综合探讨 CTSV 和健康管理因素之间的相关关系，并针对三阴性乳腺患者和高危人群实施健康风险评估和三级健康管理策略。

## 3. 研究结果

### 3.1 差异表达分析结果

结果表明不同病理分型的 CTSV 基因表达水平有明显差别，三阴性乳腺癌组中 CTSV 基因表达量最多，变异程度、箱体及中位数相对较大（图 1）。从其他的病理分类可以发现 ER 阳性和 HER-2 阳性、ER 阳性、PR 阳性和 HER-2 阳性等组的 CTSV 基因的表达水平较低，变异较小、箱体和中位数均偏低。Kruskal - Wallis 检验的结果显示，统计值达到 171.82，自由度为 7，P 值的大小是 2.2e-16 ( $P < 0.05$ )。CTSV 基因的表达水平 (tpm\_unstranded) 在不同的病理分型之间表现出明显的差异性，这种差异在统计学上获得了高度的认可。

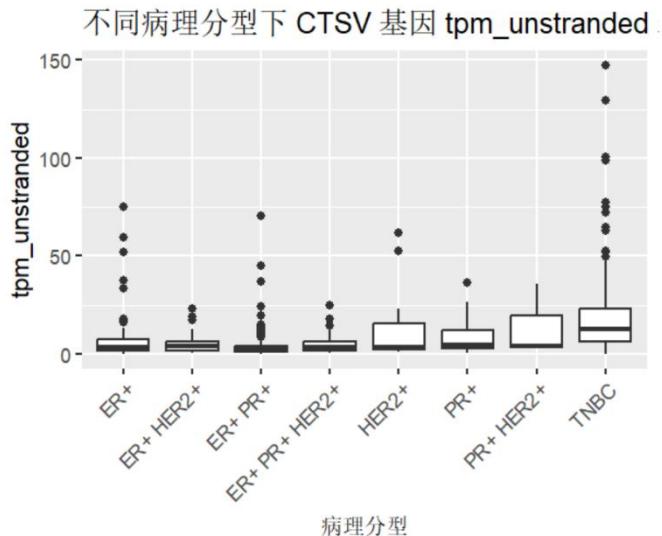


图 1 不同病理分型下 CTSV 基因 tpm\_unstranded 表达水平分布箱线图

注：展示三阴性乳腺癌与非三阴性乳腺癌患者中 CTSV 基因的 tpm\_unstranded 表达水平。箱体表示 25%~75% 分位数范围，箱内横线为中位数，须线延伸至 1.5 倍四分位距，离散点为潜在异常值。结果显示，三阴性乳腺癌组（左侧）的 CTSV 表达中位数显著高于非三阴性乳腺癌组（右侧）（Kruskal-Wallis 检验， $P < 0.05$ ），即 CTSV 基因表达水平与乳腺癌病理分型高度相关，TNBC 患者中 CTSV 表达显著上调，可能为潜在生物标志物或治疗靶点。

### 3.2 生存分析结果

#### 3.2.1 CTSV 表达与生存情况的逻辑回归分析

在单变量分析中  $P$  值达到了  $0.421$  ( $P > 0.05$ )，意味着 CTSV 的表达对生存情况的直接作用是不明显的。考虑到可能有其他的干扰因素，我们决定将临床分型数据纳入到模型中，进行多因素的逻辑回归分析。多因素分析的  $P$  值都小于  $0.05$ ，说明 CTSV 的表达水平会明显影响生存情况，从系数是正值可以看出 CTSV 的表达水平越高，患者的死亡风险越大。`classification_of_tumor3`（原发性）和 `classification_of_tumorother`（其他类型）的系数是负数 ( $P < 0.05$ )，表明此类型患者的死亡风险显著低于参考组。由 OR 值得知肿瘤分类调整后 CTSV 的高表达与病人的死亡风险呈显著正相关。

由上可得，CTSV 基因表达水平(`tpm_unstranded`)升高会增加三阴性乳腺癌患者的死亡情况，CTSV 表达水平的升高对肿瘤患者死亡的结局的影响大小由早期向晚期减少。因此，可以将 CTSV 作为检测三阴性乳腺癌患者不良预后结局的指标，特别是进展期的患者，进行三级健康管理。

#### 3.2.2 Kaplan-Meier 生存曲线分析

为探究 CTSV 表达水平与 TNBC 患者预后的关系，本研究使用了 TCGA 数据集通过 Kaplan-Meier 法对高、低两个不同的 CTSV 表达组的生存率进行对比分析绘制生存曲线，使用 Log-rank 检验以计算该分析的  $P$  值。

由图 2 可得 CTSV 高表达组和低表达组的生存曲线几乎重合，说明患者预后的区分无法仅按照中位数分组 ( $P=0.55$ )。

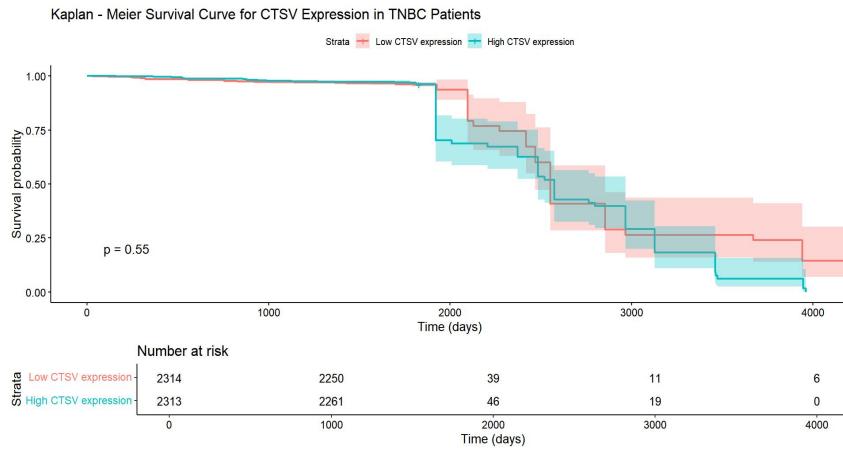


图 2 CTSV 表达在三阴性乳腺癌患者中的 Kaplan-Meier 生存曲线

注：该分析把三阴性乳腺癌患者按 CTSV 基因表达中位数分组，分为高表达组（蓝色）和低表达组（红色），两组的生存概率在随访期内整体趋势相近。生存时间为天，风险表分别列示了 0 天、1000 天、2000 天、3000 天及 4000 天时的存活人数（高表达组：2313, 2261, 46, 19, 0；低表达组：2314, 2250, 39, 11, 6）。尽管生存曲线在后期呈现一定分离，Log-rank 检验结果（ $P=0.55$ ）表明差异无统计学意义，可能与样本量较小或随访时间不足有关，需结合更大队列研究进一步验证。

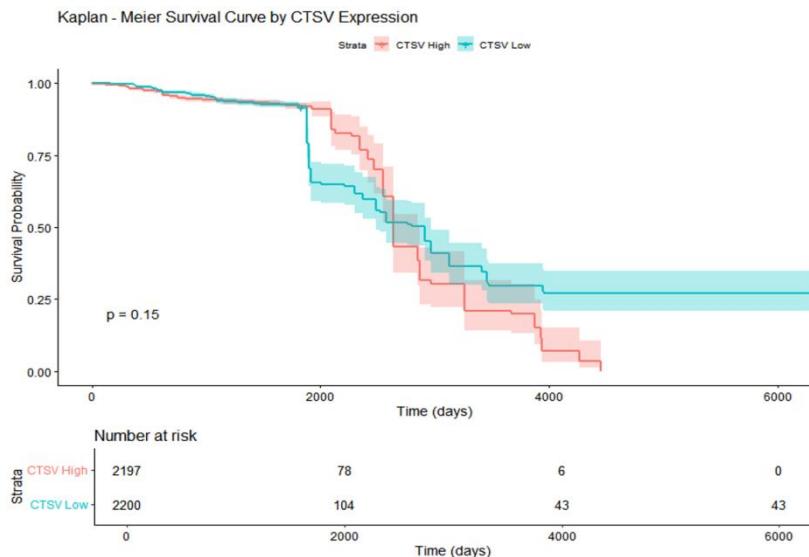


图 3 CTSV 表达在三阴性乳腺癌患者中的 Kaplan-Meier 生存曲线(只考虑原发型和转移型)

由图 3 可得在调整肿瘤分类（原发型和转移型）后，生存差异仍不显著，支持 CTSV 分组的独立性显然不足（ $P=0.15$ ）。

注：该分析在调整肿瘤分类（原发型和转移型）后，把三阴性乳腺癌患者按 CTSV 基因表达中位数分组，分为高表达组（红色）和低表达组（蓝色），随访早期（0-2000 天）生存概率趋于一致，在 2000 天的时间节点呈现分离趋势。生存时间为天，风险表分别列示了 0 天、2000 天、4000 天及 8000 天时的存活人数（高表达组：2197, 78, 6, 0；低表达组：2200, 104, 43, 43）。Log-rank 检验结果（ $P=0.15$ ）表明差异无统计学意义。但原发型显著降低死亡风险，提示肿瘤分类是影响生存的关键因素，CTSV 表达未显示独立预后价值。

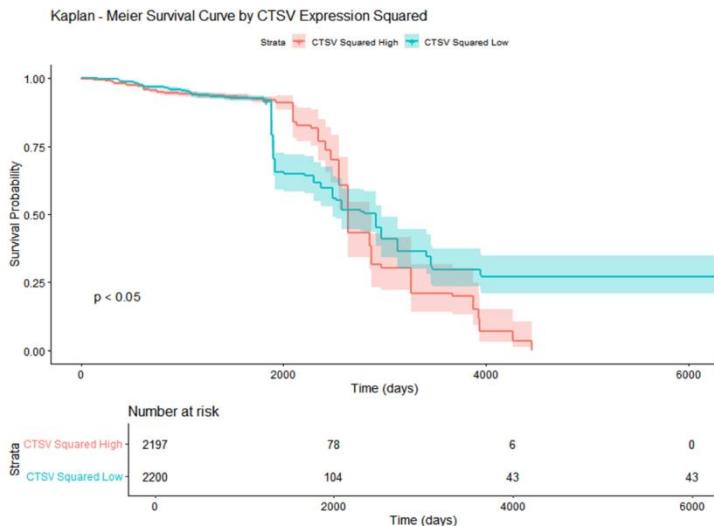


图 4 CTSV 表达二次项分组的 Kaplan-Meier 生存曲线(只考虑原发型和转移型)

注：该图展示了展示三阴性乳腺癌患者按 CTSV 基因表达量二次项中位数分组（高二次项组和低二次项组）的 Kaplan-Meier 生存曲线。生存时间单位为天，风险表分别列示随访时间点（0 天、2000 天、4000 天、6000 天）的存活人数（高二次项组：2197、78、6、0；低二次项组：2200、104、43、43）。Log-rank 检验显示组间生存率存在显著差异 ( $P<0.05$ )。曲线趋势显示高二次项组在长期随访中生存概率下降更明显，需进一步验证 CTSV 表达的非线性剂量效应及其临床意义。

由图 4 可得 CTSV 平方项分组的生存曲线存在显著差异，提示非线性效应的证据充分 ( $P<0.05$ ，二次项分组)。

### 3.2.3 Cox 比例风险模型分析

为了量化多个变量对死亡风险的独立影响，计算风险比 (HR) 及置信区间 (CI)，本研究引入 Cox 比例风险模型进行分析，结果如下：

第一组模型分析结果为，CTSV 分组的 HR 为 1.069 ( $P=0.619$ )，其 95% 置信区间 (95%CI) 包含 1，表明 CTSV 高表达组与低表达组的死亡风险差异无统计学意义。年龄、癌症分期与治疗组的 HR 值分别为 1.000、2.578 和 0.1865，提示年龄增长、癌症晚期等因素与风险增加相关，而化疗能够显著降低风险。可以说明年龄和部分病理分期、治疗方式是影响 TNBC 患者生存的显著因素，而 CTSV 表达水平在调整协变量后对患者生存的影响不显著。

第二组单独以肿瘤类型分别为原发型和转移型为协变量进行分析，得知原发肿瘤的死亡风险  $HR=0.419$  ( $P=3.06e-06$ ) 显著低于转移瘤，在临床中对于原发型肿瘤的治疗和预后评估可能具有重要意义。CTSV 分组仍不显著，模型预测能力较弱。

第三组考虑到 CTSV 表达量是非线性变量，取 CTSV 表达量的二次项分析用于探索变量与风险的非线性关系，并且绘制曲线，得到 CTSV 线性项的 HR 为 1.026 ( $P=0.0038$ ) 即每单位表达量增加风险 2.6%，CTSV 表达量与风险呈近似线性正相关。由此证明 CTSV 高表达能够增加患者的死亡风险，且风险是非线性的与 CTSV 表达的二次项（平方）相关。

## 3.3 健康管理分析结果

### 3.3.1 CTSV 表达与健康管理因素的关联性

肥胖容易导致慢性炎症和缺氧微环境通过释放炎症因子从而上调缺氧诱导因子的水平<sup>[7]</sup>，容易产生炎症反应，而 CTSV 基因表达水平与炎症反应关系密切<sup>[8]</sup>，鉴于上述发现，对于三阴性乳腺癌的高危人群而言，控制体重是预防该疾病发生和发展的重要措施之一。在日常生活中应该将体重控制在健康的范围之内，即体重指数 (BMI) 控制在 18.5 至 23.9 的正常范围内，以此降低缺氧微环境对肿瘤生长的促进作用，减轻疾病发生的风险。

CTSV 属于半胱氨酸蛋白酶家族，能够影响细胞外基质的重建<sup>[9]</sup>，导致炎症细胞浸润。 CTSV 高

表达者在日常生活中需要避免慢性炎症的状态，在饮食上多摄入抗炎膳食包括谷薯类等富含丰富的碳水化合物的食物；西兰花、胡萝卜、海带等富含膳食纤维及矿物质的食物；猕猴桃、橘子、草莓等含有丰富的抗氧化的花青素和多酚类物质的水果；米、红豆等全谷物与杂豆类食物等。除了注重饮食结构，保持规律的身体活动也是降低身体炎症水平的关键因素之一。而保持规律的运动，每周需要至少150min的中等量运动或者进行75min的高强度运动。摒弃吸烟、熬夜、酗酒等不良的生活习惯，以抑制身体的炎症反应。

### 3.3.2 基于 CTSV 表达的健康风险评估模型

本文采用差异表现分析与生存分析全面揭示了 CTSV 基因的表达水平与三阴性乳腺癌的紧密联系。

CTSV 基因表达水平在三阴性乳腺癌中最高，可在某种程度上预测癌症发生的风险。对于 CTSV 基因表达水平升高但疾病未发生的高危人群，需要采取一级预防（病因预防）手段，针对个体的危险因素采取个性化预防措施如合理膳食、适当体力活动、避免环境暴露、保持良好的精神状态及摒弃不良的生活习惯等措施来减少三阴性乳腺癌的发生几率。

CTSV 基因表达水平的升高对于早期三阴性乳腺癌的死亡风险预测更加精确，因此可以作为早期监测指标辅助早期治疗决策以避免治疗过度或不充分。在早期的筛查中可以结合 CTSV 表达水平和常规的影像学检查来建立动态监测模型，当 CTSV 表达量超过阈值时即刻采取更频繁的随访或者穿刺活检等措施，用这种方式来提高三阴性乳腺癌的早期检出率。早期的三阴性乳腺癌患者需要及时采取二级预防（临床前期预防），以控制三阴性乳腺癌的发展和恶化，减轻三阴性乳腺癌转移和向晚期进展的风险。

以 CTSV 作为不良预后指标能够在极大程度上减少复发与转移发生的死亡风险。对术后三阴性乳腺癌患者进行康复监测，在病人康复期间如果发现 CTSV 基因表达水平增高的情况则说明该患者存在复发风险，需要及时采取三级预防（临床预防）措施，通过手术治疗、放疗、化疗等有效的治疗措施，防止三阴性乳腺癌的进一步恶化。在预后评估环节中还可以通过纳入现成的预后模型 PREDICT，结合患者的肿瘤分期、年龄和治疗方案等因素，建立一个综合的风险评分系统来指导三阴性乳腺癌患者的个性化随访频率和治疗强度。

### 3.3.3 基于 CTSV 表达的健康干预方案

在健康教育方面需要明确降低三阴性乳腺癌的复发的风险、改善患者的生活质量和延长生存周期的核心健康管理目标。知识教育可以从三阴性乳腺癌的生物学特性、治疗难点以及预后特点等方面入手，编写三阴性乳腺癌相关的科普手册和录制针对三阴性乳腺癌的讲解视频或者线上课程能够更加充分全面的向患者及高危人群进行深入的科普教育。同时教会患者怎么使用简单的监测工具，自我监测体重的变化，自行记录异常症状。

生活方式管理是健康干预策略的重点内容。在饮食上要坚持抗炎膳食计划，推荐摄入糙米、薏米等全谷物碳水化合物；菠菜、西兰花等的绿叶蔬菜；黑莓、桑葚、蓝莓等含有丁苯甲酸酯、花青素等多酚类植物化合物的水果。一定要严格限制精制碳水化合物、红肉、加工食品等具有促炎潜力的食物的摄入。根据患者或高危人群的状态制定合理的运动计划，每周需要至少150min的中等量运动或者进行75min的高强度运动，可以适当加入抗阻训练和灵活性训练。通过心理指导减轻患者的患病压力，推荐患者每日记录情绪波动及触发因素方便心理咨询工作的开展。随访的时间、频率根据患者术后复发风险来决定，一般情况按三个月、六个月或一年的时间节点作为随访周期。基于 CTSV 表达的干预方案需要额外注意整合患者的 CTSV 表达数据。

## 4 结论

三阴性乳腺癌的特点是 ER、PR 和 HER-2 的表达都呈现为阴性，是一种较为特殊的乳腺癌类型。三阴性乳腺癌因为预后不佳、有高侵袭性和治疗选择受限的特点而受到了很多关注。由于目前尚无有效的靶向治疗手段，三阴性乳腺癌病人的治疗方式多以化疗为主，在很大程度上限制了治疗效果的提高。所以寻找能够预测三阴性乳腺癌患者手术后复发、转移和预测健康风险的生物标志物，对完善三阴性乳腺癌高危人群的健康管理策略和改善病人的预后具有重要的意义。

**参考文献:**

- 郭贝贝. PARP1、OGG1、XRCC1、XRCC2 和 XRCC3 基因多态性与乳腺癌易感性的研究[D]. 甘肃: 兰州大学, 2020.
- OBIDIRO O, BATTOGTOKH G, AKALA E O. Triple Negative Breast Cancer Treatment Options and Limitations: Future Outlook[J]. Pharmaceutics, 2023, 15(7).
- LU Y, SUN X, PENG L, et al. Angiotensin II-Induced vascular remodeling and hypertension involves cathepsin L/V- MEK/ERK mediated mechanism[J]. International Journal of Cardiology, 2020, 298: 98-106.
- 周翰林. 基于生物信息学分析构建铁死亡/铜死亡相关基因在癌症中的预后评估模型[D]. 山东: 曲阜师范大学, 2024.
- Duan S, Gao J, Lou W, et al. Prognostic signature for hepatocellular carcinoma based on 4 pyroptosis-related genes [J]. BMC Med Genomics, 2022, 15: 166.
- Dai H, Ren J, Wang C, et al. Prognostic molecular subtype reveals the heterogeneity of tumor immune microenvironment in gastric cancer [J]. Sci Rep, 2025, 15: 14453.
- 毛婷婷, 何进勇, 李征征等. 肥胖微环境促进脂肪纤维化的发病机制研究进展[J]. 现代生物医学进展, 2017, 17(19): 3798-3800.
- XIA Y, GE M, XIA L, et al. CTSV (cathepsin V) promotes bladder cancer progression by increasing NF- $\kappa$  B activity[J]. Bioengineered, 2022, 13(4): 10180-10190.
- TOSS M, MILIGY I, GORRINGE K, et al. Prognostic significance of cathepsin V (CTSV/CTSL2) in breast ductal carcinoma in situ[J]. J Clin Pathol, 2020, 73(2): 76-82.

## **Health risk management of triple negative breast cancer based on CTSV gene expression**

**Xu, Wencong<sup>1</sup>, Wang, Xiaojuan<sup>1</sup>, Huang, Jiaxin<sup>1</sup>**

*1.School of Health and Medical Science, Neusoft Institute Guangdong, Foshan, China*

**Abstract:** Triple-negative breast cancer (TNBC) rates increase each year and this rising trend will affect the survival quality of patients, so there is a need to find effective early warning methods and health management strategies. In this study, the differential expression of the CTSV gene in TNBC and the association with the survival status of patients were analysed from the perspectives of bioinformatics and health management, respectively, and the relationship between obesity, chronic inflammation and the expression of the CTSV gene was explicitly proposed, and the results obtained were translated and applied to health management. The results of the analyses support the idea that CTSV is a valid predictor of poor prognosis in patients with TNBC, and that the risk of death in patients with advanced TNBC is abruptly increased by overexpression of the CTSV gene. Combining the expression of CTSV in TNBC with the three-stage prevention strategy of health management (screening-surveillance-treatment) and proposing targeted solutions can optimise the whole management of TNBC patients to a certain extent, and provide theoretical support for the development of health management strategies for TNBC high-risk groups.

**Keywords:** Bioinformatics; CTSV; TNBC; Health Management