

Análise de coortes genómicas públicas de larga escala revelam *BCL6* como marcador de prognóstico no cancro da mama Luminal A

Analysis of large-scale public genomic cohorts reveals BCL6 as a prognostic marker in Luminal A breast cancer

Patrícia Barros, Peter Jordan, Paulo Matos

paulo.matos@insa.min-saude.pt

Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

_Resumo

O cancro da mama (CM) é a neoplasia maligna mais comum entre mulheres a nível mundial, constituindo uma das principais causas de mortalidade oncológica. Representa uma doença heterogénea, agrupada em subtipos moleculares com implicações prognósticas e terapêuticas distintas. O subtipo Luminal A é o mais prevalente e caracteriza-se por elevada expressão de recetores hormonais (estrogénio e progesterona), baixa taxa de proliferação e prognóstico geralmente favorável. No entanto, há evidências consistentes de risco significativo de recorrência tardia e de novos eventos neoplásicos, o que continua a levantar desafios na definição das estratégias de seguimento clínico.

Neste estudo, analisámos dados genómicos da coorte de CM da base de dados pública TCGA (n=1247) para avaliar o valor prognóstico do gene *BCL6*, um regulador transcricional previamente implicado na progressão tumoral.

Foram recolhidos dados de expressão de *BCL6*, subtipagem molecular (PAM50) e sobrevivência global (OS). Observou-se que, apesar da expressão de *BCL6* estar globalmente diminuída nos tumores em comparação com o tecido normal, esta era significativamente mais elevada nos tumores Luminal A do que nos restantes subtipos, com um subgrupo (44%) a manter níveis semelhantes aos do tecido normal. Verificámos ainda que, exclusivamente neste subtipo, a expressão elevada de *BCL6* se associava a pior sobrevivência ($p=0,041$).

Estes resultados apontam para o potencial de *BCL6* como biomarcador de estratificação de risco dentro do subtipo Luminal A, com possíveis implicações na definição da intensidade e duração do seguimento clínico a longo prazo.

_Abstract

Breast cancer (BC) is the most common malignant neoplasm among women worldwide and remains one of the leading causes of cancer-related mortality. It is a heterogeneous disease, classified into molecular subtypes with distinct prognostic and therapeutic implications. The Luminal A subtype is the most prevalent and is characterized by high expression of hormone receptors (estrogen and progesterone), low proliferation rates, and generally favorable outcomes. However, there is consistent evidence of a significant risk of late recurrence and second primary malignancies, which continues to pose challenges for defining optimal clinical follow-up strategies.

In this study, we analyzed public genomic data from the TCGA breast cancer cohort (n=1247) to evaluate the prognostic value of the *BCL6* gene, a transcriptional regulator previously implicated in tumor progression. We retrieved *BCL6* expression data, molecular subtype classification (PAM50), and overall survival (OS) information. While *BCL6* expression

was globally reduced in tumor tissue compared to normal breast tissue, it was significantly higher in a subset of Luminal A tumors (44%) than in other subtypes. Notably, only within the Luminal A subtype did higher *BCL6* expression correlate with worse survival outcomes ($p=0.041$).

These findings support further clinical and molecular validation of *BCL6* usefulness as a biomarker for identifying Luminal A patients at increased risk of late relapse, with potential impact on risk stratification and long-term follow-up decision-making.

_Introdução

A incidência do cancro da mama (CM) tem aumentado globalmente nas últimas duas décadas, tendência associada ao envelhecimento populacional (1). Desde 2022, o CM tornou-se o segundo tumor maligno mais prevalente, com 2,3 milhões de novos casos anuais a nível mundial, representando cerca de 12% de todos os diagnósticos de cancro (2). Nas mulheres, é o tipo de cancro mais frequentemente diagnosticado, sobretudo após os 50 anos (2,3). Em Portugal, o CM é também a neoplasia mais prevalente entre as mulheres, com cerca de 8954 novos casos diagnosticados anualmente, representando aproximadamente 30% de todos os novos casos de cancro. Estimam-se ainda cerca de 2211 mortes por ano associadas a esta neoplasia, correspondendo a cerca de 16% da mortalidade oncológica feminina (2,3). Apesar dos avanços no diagnóstico precoce e tratamento do CM, os custos associados e a morbili-dade permanecem elevados (4). Além disso, prevê-se um aumento significativo da carga global do CM, com mais de 3 milhões de novos casos e 1 milhão de mortes anuais até 2040 (5).

O CM é uma doença multifatorial, resultante da interação entre fatores genéticos e ambientais (6). Entre os fatores modificáveis destacam-se o sedentarismo, o consumo de álcool e tabaco, o excesso de peso e o uso de terapêutica hormonal. Fatores reprodutivos como menarca precoce, menopausa tardia e amamentação reduzida também estão associados ao risco aumentado de CM (7). Cerca de 5–10% dos casos são hereditários, sendo os genes *BRCA1* e *BRCA2* os mais frequentemente implicados (8).

A heterogeneidade molecular do CM reflete-se na sua classificação em quatro subtipos principais, com implicações terapêuticas e prognósticas (9,10). Os tumores luminais representam 70% dos casos, dividindo-se em Luminal A (50–60%) e Luminal B (10–20%). O subtipo Luminal A é geralmente de melhor prognóstico, com baixo índice proliferativo, expressão de recetores hormonais para o estrogénio (ER) e a progesterona (PR), e expressão ausente ou reduzida do recetor do fator de crescimento epidérmico humano 2 (HER2). O subtipo Luminal B, por sua vez, é também geralmente HER2 negativo e ER positivo, mas pode ser PR negativo e apresenta um maior índice proliferativo e prognóstico mais reservado (9). Os tumores de subtipo HER2-enriquecido (HER2+, 15–20% dos casos) são mais agressivos e frequentemente independentes de recetores hormonais. Por último, os triplo-negativos (TNBC) representam 10–15% dos casos e são os mais difíceis de tratar, devido à ausência de expressão de ER, PR e HER2, para os quais existem terapias dirigidas eficazes (10).

Face a esta complexidade, a caracterização molecular rigorosa dos subtipos e a identificação de biomarcadores com valor diagnóstico, prognóstico e preditivo são fundamentais para permitir uma estratificação mais precisa dos doentes, antecipar a progressão da doença e orientar o desenvolvimento de estratégias terapêuticas eficazes e personalizadas (11,12).

O gene *BCL6* (*B-cell lymphoma 6*), que codifica um repressor transcricional previamente caracterizado pelo nosso grupo como um regulador chave da progressão do cancro colorretal (13–15), surgiu como um potencial biomarcador no CM (16). Inicialmente identificado em linfomas B, *BCL6* também é expresso em tecidos epiteliais da mama e cólon (17). Atua regulando genes cruciais para a supressão tumoral, reparação do

DNA, evasão imunitária, inibição da apoptose e controlo do ciclo celular, processos cuja desregulação é crucial no desenvolvimento do cancro (16,18,19). Contudo, o seu papel no CM permanece controverso, com estudos a reportarem tanto associação a pior prognóstico, como perda de expressão em tumores recorrentes e metastáticos (20,21). Estas evidências contraditórias, sublinham a necessidade de clarificar o papel de *BCL6* como biomarcador no CM.

_Objetivos

Neste artigo, demonstramos como é possível recorrer à vasta quantidade de dados genómicos, fenotípicos e clínicos de coortes de grande dimensão, disponíveis em plataformas públicas como o *The Cancer Genome Atlas* (TCGA), e ferramentas acessíveis como a plataforma *UCSC Xena*, para explorar *in silico* a relevância de novos biomarcadores, como *BCL6*, no contexto do CM.

_Materiais e métodos

Os dados de expressão génica referentes ao gene *BCL6* foram obtidos a partir do conjunto de dados da coorte BRCA do TCGA (<https://www.cancer.gov/tcga>), através da plataforma *UCSC Xena* (<https://xena.ucsc.edu/>). Foram recolhidos valores normalizados [$\log_2(\text{normalized count}+1)$] de dados de RNA-seq e correlacionados com o tipo de amostra (*sample type*), subtipo tumoral (*PAM50 mRNA*, (22)), e os status de ER, PR e HER2 (22). Os dados foram exportados em ficheiros *.tsv* e analisados no *GraphPad Prism v5.01*. A comparação da expressão de *BCL6* entre tumores primários, metastáticos e tecidos normais foi realizada por análise *one-way ANOVA*, seguida de teste post-hoc de Newman-Keuls, considerando-se diferenças entre grupos significativas se $p < 0,05$. Para a análise da expressão entre subtipos tumorais, os dados dos tumores primários foram agrupados por subtipo molecular e comparados com tecidos normais usando o mesmo método estatístico. Os tumores triplo-negativos (TNBC) foram identificados dentro do subtipo *Basal-Like*, com confirmação da negatividade para ER, PR e HER2. Os resultados foram representados em gráficos de dispersão vertical (*scatter plots*), indicando a mediana (linha horizontal espessa) e o intervalo interquartil (IQR) como

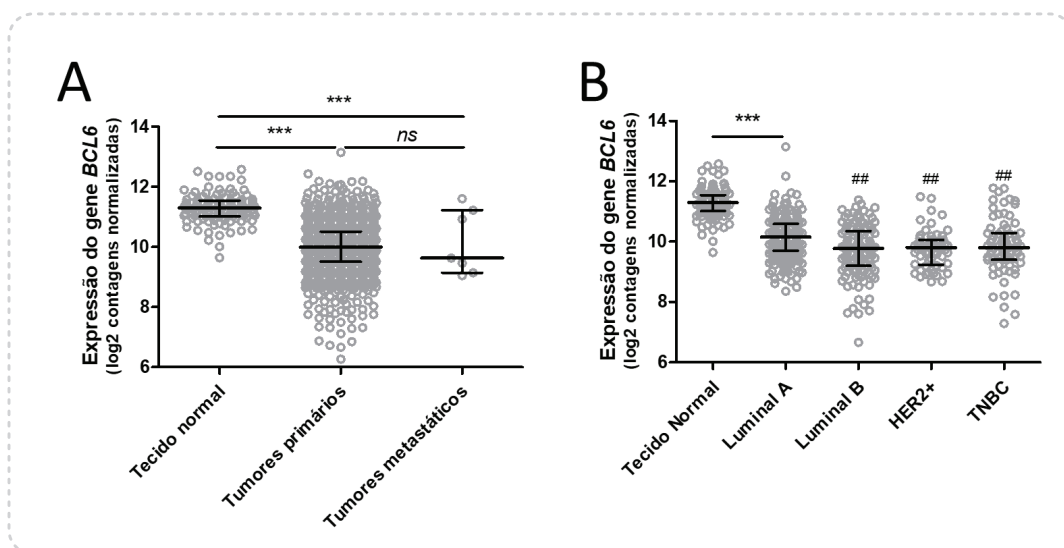
barras de erro. Para a análise de sobrevivência global (OS), os valores de OS e OS.time foram extraídos da plataforma UCSC Xena e filtrados por tipo de tumor ou subtipo. A mediana de expressão de *BCL6* foi usada como ponto de corte para definição dos grupos com expressão elevada e reduzida de *BCL6*. As curvas de sobrevivência Kaplan-Meier foram geradas no GraphPad, e as diferenças entre grupos avaliadas pelo teste Log-rank (Mantel-Cox), considerando-se significativas para $p < 0,05$.

Resultados

Para clarificar o papel da expressão do repressor *BCL6* como biomarcador de prognóstico e progressão do cancro da mama (CM), recolhemos, da base de dados pública TCGA, a informação relativa à expressão de *BCL6* em função do tipo de amostra: tecido normal, tumores primários ou tumores metastáticos (figura 1A). Constatámos que, ao contrário do descrito para outras coortes de menor dimensão (20), a expressão de *BCL6* se encontrava significativamente diminuída nos tumores primários e metastáticos, em comparação com os tecidos normais ($p < 0,001$).

Devido à notória dispersão dos dados de expressão no grupo dos tumores primários, procedemos à filtragem dos dados relativos ao subtipo molecular destas amostras, disponível nos metadados fenotípicos do TCGA (parâmetro PAM50, conforme descrito em (22)). Para confirmação, extraímos também a classificação histológica de positividade para os recetores ER, PR e HER2, conforme determinado na mesma referência. Esta análise permitiu comparar a expressão de *BCL6* entre os tecidos normais e os tumores dos quatro subtipos moleculares: Luminal A, Luminal B, HER2+ e TNBC (figura 1B). Observámos que, embora a expressão global de *BCL6* permanecesse significativamente reduzida em todos os subtipos tumorais, esta era significativamente mais elevada nos tumores do subtipo Luminal A, comparativamente aos restantes ($p < 0,01$). Aplicando a regra de Tukey para excluir outliers no conjunto de valores de expressão de *BCL6* em tecidos normais, verificou-se ainda que 44% dos tumores Luminal A apresentam níveis de expressão dentro do intervalo interquartil estendido definido com base nas amostras de tecido normal. Para aprofundar esta observação, procedemos à extração, na plataforma UCSC Xena, dos dados de sobrevivência global (OS) para a totalidade dos tumores primários (figura 2A).

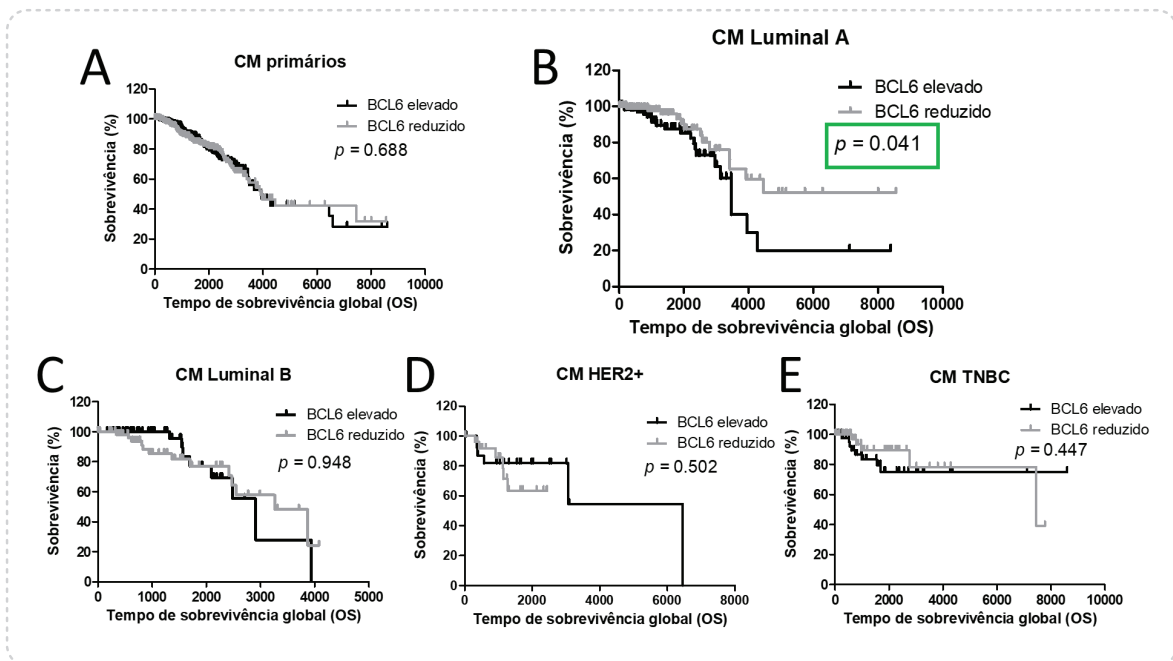
Figura 1: Expressão do gene *BCL6* na coorte de cancro da mama (BRCA) da base de dados TCGA.



(A) Comparação da expressão de *BCL6* entre tecido mamário normal, tumores primários e tumores metastáticos ($n=1247$).

(B) Distribuição da expressão de *BCL6* nos quatro subtipos moleculares definidos de CM: Luminal A, Luminal B, HER2-enriquecido (HER2+) e tripló-negativo (TNBC). Em ambos os gráficos, as barras horizontais representam a mediana e o intervalo interquartil. *** $p < 0,001$, relativamente a Tecido normal; ## $p < 0,01$, relativamente a Luminal A; ns = não significativo.

Figura 2: Análise de sobrevivência global (OS) em doentes com cancro da mama primário da coorte BRCA do TCGA, com base nos níveis de expressão de *BCL6*.



(A) Curva de sobrevivência de Kaplan-Meier dos doentes com base em todos os tumores primários, comparando tumores com expressão elevada versus expressão reduzida de *BCL6*, definida com base na mediana de expressão. (B-E) Curvas de Kaplan-Meier estratificadas por subtipo molecular: Luminal A (B), Luminal B (C), HER2-enriquecido (HER2+; D) e triplo-negativo (TNBC; E). As curvas foram construídas com base nos dados de sobrevivência global (OS) e comparadas entre os dois grupos de expressão de *BCL6* utilizando o teste *log-rank*, com os valores de *p* indicados em cada gráfico. As linhas horizontais representam o tempo de sobrevivência em dias.

Construindo curvas de sobrevivência de Kaplan-Meier, e utilizando a mediana da expressão de *BCL6* como ponto de corte para definir os grupos de expressão elevada e reduzida, verificámos que apenas no subtipo Luminal A (figura 2B), e não nos restantes (figura 2C-E), uma expressão mais elevada de *BCL6* se correlacionava com um pior prognóstico ($p=0,041$). Esta descoberta teria ficado oculta se fosse só considerado o conjunto total de tumores ($p=0,688$).

Discussão

O subtipo Luminal A do cancro da mama (CM) é geralmente associado a um prognóstico favorável, apresentando taxas de recorrência significativamente mais baixas em comparação com outros subtipos, como o Luminal B ou o triplo-negativo (TNBC). Estudos de longo prazo indicam uma taxa de recorrência locorregional de cerca de 1,6% aos cinco anos, com uma taxa global de recaída de aproximadamente 5% no mesmo intervalo (23,24). No entanto, mesmo neste subtipo considerado de baixo risco, continuam a verificar-se

casos de recaída tardia e de ocorrência de novos tumores primários. Estima-se que a recorrência local atinja 3,7% aos dez anos, enquanto a metástase à distância possa atingir os 9,5% no mesmo período (25). Além disso, existem evidências de que os doentes com tumores Luminal A mantêm um risco considerável de recorrência que pode prolongar-se e até aumentar entre os 15 e os 20 anos após o diagnóstico inicial (26). Estes dados estão em consonância com evidências nacionais, nomeadamente um estudo de coorte realizado no norte de Portugal, que reportou uma taxa cumulativa de cerca de 7% de novos tumores primários em sobreviventes de CM ao fim de dez anos (27).

Neste contexto, a identificação de novos marcadores moleculares com valor prognóstico ou preditivo de recaída é essencial para melhorar a estratificação de risco e otimizar o seguimento a longo prazo. A nossa análise demonstrou que, embora a expressão de *BCL6* esteja globalmente diminuída em tumores primários e metastáticos face ao tecido normal, é significativamente mais elevada no subtipo Luminal A.

artigos breves_ n. 11

Uma fração considerável destes tumores apresenta níveis de expressão semelhantes aos do tecido normal, sugerindo a existência de uma subpopulação biologicamente distinta.

Em termos clínicos, não se observou impacto da expressão de *BCL6* na sobrevivência global da coorte total, mas, no subtipo Luminal A, uma expressão elevada de *BCL6* associou-se a pior prognóstico ($p=0,041$). Esta correlação não se verificou noutros subtipos, sugerindo um papel específico de *BCL6* na progressão dos doentes Luminal A.

Conclusões

Os resultados obtidos sugerem que *BCL6* poderá ser um marcador molecular útil para identificar doentes com tumores Luminal A com maior risco de recorrência tardia, mesmo na presença de recetores hormonais acionáveis com terapia dirigida. A sua expressão mais próxima à do tecido normal associa-se a um pior prognóstico exclusivamente neste subtipo, o que aponta para um possível papel na recorrência da doença.

Estes dados destacam a relevância de integrar novos marcadores na avaliação clínica de risco e na personalização do seguimento, incluindo decisões sobre a duração e intensidade da vigilância.

Serão agora necessários estudos clínicos e moleculares adicionais para confirmar o potencial de *BCL6* como biomarcador e esclarecer o seu papel como regulador transcricional na evolução do subtipo Luminal A.

Referências bibliográficas:

- (1) Xu Y, Gong M, Wang Y, et al. Global trends and forecasts of breast cancer incidence and deaths. *Sci Data*. 2023 May 27;10(1):334. <https://doi.org/10.1038/s41597-023-02253-5>
- (2) Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024 May-Jun;74(3):229-263. <https://doi.org/10.3322/caac.21834>
- (3) Filho AM, Laversanne M, Ferlay J, et al. The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. *Int J Cancer*. 2025 Apr 1;156(7):1336-46. Epub 2024 Dec 17. <https://doi.org/10.1002/ijc.35278>
- (4) Łukasiewicz S, Czezelewski M, Forma A, et al. Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. *Cancers (Basel)*. 2021 Aug 25;13(17):4287. <https://doi.org/10.3390/cancers13174287>
- (5) Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *Breast*. 2022 Dec;66:15-23. <https://doi.org/10.1016/j.breast.2022.08.010>
- (6) Tufail M, Cui J, Wu C. Breast cancer: molecular mechanisms of underlying resistance and therapeutic approaches. *Am J Cancer Res*. 2022 Jul 15;12(7):2920-49. <https://e-century.us/files/ajcr/12/7/ajcr0144032.pdf>
- (7) Rudolph A, Chang-Claude J, Schmidt MK. Gene-environment interaction and risk of breast cancer. *Br J Cancer*. 2016 Jan 19;114(2):125-33. <https://doi.org/10.1038/bjc.2015.439>
- (8) Sokolova A, Johnstone KJ, McCart Reed AE, et al. Hereditary breast cancer: syndromes, tumour pathology and molecular testing. *Histopathology*. 2023 Jan;82(1):70-82. Epub 2022 Dec 5. <https://doi.org/10.1111/his.14808>
- (9) Eliyatkin N, Yalçın E, Zengel B, et al. Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. *J Breast Health*. 2015 Apr 1;11(2):59-66. <https://doi.org/10.5152/tjbh.2015.1669>
- (10) Orrantia-Borunda E, Anchondo-Nuñez P, Acuña-Aguilar LE, et al. Subtypes of Breast Cancer. In: Mayrovitz HN (ed). *Breast Cancer (Internet)*. Brisbane (AU): Exon Publications; 2022. <https://doi.org/10.36255/exon-publications-breast-cancer-subtypes>
- (11) Ye F, Dewanjee S, Li Y, et al. Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer. *Mol Cancer*. 2023 Jul 6;22(1):105. <https://doi.org/10.1186/s12943-023-01805-y>
- (12) Hou Y, Peng Y, Li Z. Update on prognostic and predictive biomarkers of breast cancer. *Semin Diagn Pathol*. 2022 Sep;39(5):322-32. <https://doi.org/10.1053/j.semdp.2022.06.015>
- (13) Barros P, Jordan P, Matos P. Rac1 signaling modulates BCL-6-mediated repression of gene transcription. *Mol Cell Biol*. 2009 Aug;29(15):4156-66. <https://doi.org/10.1128/MCB.01813-08>
- (14) Barros P, Lam EW-F, Jordan P, et al. Rac1 signalling modulates a STAT5/BCL-6 transcriptional switch on cell-cycle-associated target gene promoters. *Nucleic Acids Res* 2012;40:7776-87. <https://doi.org/10.1093/nar/gks571>
- (15) Barros P, Amaral AJ, Abrantes LB, et al. 76P - Stimulation of RAC1/PAK1 signalling upregulates DNA damage repair genes via the BCL6/STAT5-switch. *Ann Oncol*. 2017;28(Suppl.5):V19-20. <https://doi.org/10.1093/annonc/mdx361.069>
- (16) Cardenas MG, Oswald E, Yu W, et al. The Expanding Role of the BCL6 Oncoprotein as a Cancer Therapeutic Target. *Clin Cancer Res*. 2017 Feb 15;23(4):885-93. Epub 2016 Nov 23. <https://doi.org/10.1158/1078-0432.CCR-16-2071>
- (17) Logarajah S, Hunter P, Kraman M, et al. BCL-6 is expressed in breast cancer and prevents mammary epithelial differentiation. *Oncogene*. 2003 Aug 28;22(36):5722-78. <https://doi.org/10.1038/sj.onc.1206689>
- (18) McLachlan T, Matthews WC, Jackson ER, et al. B-cell Lymphoma 6 (BCL6): From Master Regulator of Humoral Immunity to Oncogenic Driver in Pediatric Cancers. *Mol Cancer Res*. 2022 Dec 2;20(12):1711-23. <https://doi.org/10.1158/1541-7786.MCR-22-0567>
- (19) Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discov*. 2022 Jan;12(1):31-46. <https://doi.org/10.1158/2159-8290.CD-21-1059>
- (20) Wu Q, Liu X, Yan H. B-cell lymphoma 6 protein stimulates oncogenicity of human breast cancer cells. *BMC Cancer*. 2014 Jun 10;14:418. <https://doi.org/10.1186/1471-2407-14-418>
- (21) Pinto AE, André S, Silva G, et al. BCL-6 oncoprotein in breast cancer: loss of expression in disease progression. *Pathobiology*. 2009;76(5):235-42. <https://doi.org/10.1159/000228899>
- (22) Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012 Oct 4;490(7418):61-70. <https://doi.org/10.1038/nature11412>
- (23) Wang J, Luo J, Jin K, et al. Biological subtype predicts locoregional recurrence after postmastectomy radiotherapy in Chinese breast cancer patients. *Cancer Med*. 2020 Apr;9(7):2427-34. <https://doi.org/10.1002/cam4.2904>
- (24) Shim HJ, Kim SH, Kang BJ, et al. Breast cancer recurrence according to molecular subtype. *Asian Pac J Cancer Prev*. 2014;15(14):5539-44. <https://doi.org/10.7314/apjcp.2014.15.14.5539>
- (25) van Maaren MC, de Munck L, Strobbe LJA, et al. Ten-year recurrence rates for breast cancer subtypes in the Netherlands: A large population-based study. *Int J Cancer*. 2019 Jan 15;144(2):263-72. Epub 2018 Nov 28. <https://doi.org/10.1002/ijc.31914>
- (26) Knauer M, Filipits M, Dubsy P. Late recurrences in early breast cancer: for whom and how long is endocrine therapy beneficial? *Breast Care (Basel)*. 2014 May;9(2):97-100. <https://doi.org/10.1159/000362482>
- (27) Gonçalves E, Fontes F, Rodrigues JR, et al. Second primary cancers among females with a first primary breast cancer: a population-based study in Northern Portugal. *Breast Cancer Res Treat*. 2024 Apr;204(2):367-76. Epub 2023 Dec 27. <https://doi.org/10.1007/s10549-023-07224-3>