Commentary

Artificial intelligence-directed prognostication of breast cancer

Azadeh Nasrazadani, Adam M. Brufsky *

University of Pittsburgh, 300 Halket Street, Pittsburgh, PA 15213, USA

Once the diagnosis of breast cancer is established, the conversation in clinic inevitably reaches the patient's most pressing question: prognosis. Clinicians are then able only to offer their best estimates based on few tools at their disposal. Specifically, the standard anatomic staging systems, as well as the classically described Nottingham Prognostic Index (NPI), provide extrapolated approximations based on historical data which give weight to tumor grade, size, and lymph node status [1]. Certainly, discordance between anticipated and actual outcomes are commonplace with vastly different outcomes appreciated between patients diagnosed within the same stage.

More sophisticated instruments, namely Oncotype Dx and MammaPrint, have been developed. They are based on gene expression profiles and offer prognostic information with regards to potential responses to chemotherapy. Interpretation of these assays indirectly allude to outcomes, although each has its own limitations. Oncotype Dx was described in patients with hormone receptor positive, HER2 negative, and node negative disease [2]. Invasive ductal carcinoma (IDC) was the only histology type thought to be well represented in this study, casting a degree of uncertainly on the value of the recurrence score in patients with invasive lobular carcinoma (ILC) and other less common histology types [3]. Mammaprint is more inclusive, including node positive patients – yet only offers an output which delineates patients to high vs low risk disease [4].

In this article of *EBioMedicine*, Shimizu and Nakayama describe a newly developed molecular prognostic score (mPS), derived utilizing inherently unbiased, artificial intelligence (AI)-based methods and verified across multiple gene expression platforms [5]. Through a series of algorithms, all human protein-encoding genes were examined and systematically narrowed to a final list of 23 genes associated with outcomes when either expressed at low or high levels. By incorporating data from TCGA and METABRIC cohorts, mPS is effectively generalizable across race, histology type, hormone receptor status, HER2 receptor status, lymph node status, age, and menopausal status.

Aside from the intriguing potential to identify genes that can reliably predict outcomes, many of these previously understudied genes may serve as effective targets for drug development purposes. The literature to date is sparse for most if not all of the gene panel, revealing entire areas of unchartered territory for future studies.

Quite surprisingly, when the authors employed a hypothesis-driven approach with the selection of MYC as a candidate gene, they were unable to determine a role in predicting prognosis. MYC is well appreciated for its association with poorer prognosis in B cell lymphoma, albeit more so with gene rearrangement vs overexpression [6], and thus reasonably evaluated. While it is naïve to presume similar roles for genes across all cancer types, the lack of an association with survival is still notable in this case given the central role of MYC in proliferation [7]. Nonetheless, persistently dismal survival rates in more aggressive disease underscores the plethora of data building on seminal studies, while many targets and pathways remain undiscovered. It is, of course, possible that the expression levels of these candidate genes are merely a consequence of complex upstream processes where intervention would be more effective.

Since Perou's 2000 initial report applying microarray technology to describe distinctive molecular subtypes based on gene expression signatures [8], the field has evolved dramatically. Major advancements in sequencing technologies culminated in our authors' abilities to perform an exhaustive whole genome sequencing analysis that comprehensively spans the genome for any and all potential candidate genes. Coupled with a machine learning approach, this allows a large volume of data to be processed with extrapolated outcomes determined with statistically powered accuracy. The precision in which outcomes are estimated with mPS invariably sets it apart from currently accepted predictive scores; the closest comparator being Oncotype Dx.

The implications of utilizing AI approaches in cancer research are vast. Beyond prognostic studies, there is a more applicable role for predictive models. The authors of this study propose an integrated classification system, combining mPS and the clinical stage to stratify patients to seven classes. Extrapolation of data utilizing the METABRIC cohort further led to identification of groups determined to be unaffected by cytotoxic chemotherapy. In addition to prognostication, implementation of mPS may then serve to guide clinical management.

Shimizu and Nakayama offer a promising alternative to currently available tools for breast cancer prognostication, by highlighting a role for AI-based approaches. Future large-scale prospective studies will be integral moving forward.

Author contributions

Conception, Writing, Literature Serach: Adam Brufsky MD, PhD, Azadeh Nasrazadani MD, PhD.

Declaration of Competing Interest

References

[1] Todd JH, Dowle C, Williams MR, et al. Confirmation of a prognostic index in primary breast cancer. Br J Cancer 1987;56(4):489–92. https://doi.org/10.1038/bjc.1987.230.

[2] Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. N Engl J Med Jul 12, 2018;379(2):111–21. https://doi.org/10.1056/NEJMoa1804710 [Epub 2018 Jun 3].

[3] Conlon N, Ross DS, Howard J, Catalano JP, Dickler MN, Tan LK. Is there a role for oncotype Dx testing in invasive lobular carcinoma? Breast J 2015;21(5):514–9. https://doi.org/10.1111/tbj.12445.

[4] Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. N Engl J Med 2016;375:717–29. https://doi.org/10.1056/NEJMoa1602253.

[5] Shimizu H, Nakayama K. A 23 gene-based molecular prognostic score precisely predicts overall survival of breast cancer patients. EBioMedicine 2019. https://doi.org/10.1016/j.ebiom.2019.07.046.

[6] Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 2004;103(1):275–82. https://doi.org/10.1182/blood-2003-05-1545.

[7] Carroll PA, Freie BW, Mathsyaraja H, Eisenman RN. The MYC transcription factor network: balancing metabolism, proliferation and oncogenesis. Front Med 2018;12:412–25. https://doi.org/10.1007/s11684-018-0650-z.

[8] Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.