Received: 13 August 2023 | Accepted: 29 October 2023

DOI: 10.1002/imed.1051

RESEARCH ARTICLE

WILEY

The use of transcriptomic data in developing biomarkers in breast cancer

Maher Albitar MD[1] | Andre Goy MD[2] | Andrew Pecora MD[2] | Deena Graham MD[2] | Donna McNamara MD[2] | Ahmad Charifa MD[1] | Andrew IP MD[2] | Wanlong Ma MS[1] | Stanley Waintraub MD[2]

[1]Genomic Testing Cooperative, LCA, Irvine, California, USA

[2]John Theurer Cancer Center, Hackensack University Medical Center, Hackensack, New Jersey, USA

Correspondence
Maher Albitar, Genomic Testing Cooperative, 175 Technology Drive #100, Irvine, CA 92618, USA.
Email: malbitar@genomictestingcooperative.com

Abstract

HER2 and hormone receptors are biomarkers for selecting breast cancer therapy and predicting outcomes. In the era of antibody-drug conjugates (ADC), a relatively low HER2 expression level is adequate for targeting tumor cells. We explored the potential of RNA profiling, determined by next generation sequencing (NGS), to provide more flexible clinical biomarkers as compared with immunohistochemistry (IHC) or fluorescent in situ hybridization (FISH). Data from 57 breast cancers was used to study biomarker levels as detected by routine clinical transcriptomic tests. HER2 (ERBB2), estrogen receptor alpha (ESR1), and androgen receptor (AR) mRNA levels were compared with IHC and FISH results. There was a significant overlap in the levels of ERBB2 mRNA between cases scored by IHC as zero, 1+, and 2+. This variation correlated with progression-free survival (PFS). Similarly, the ESR1 RNA accurately reflected estrogen receptor (ER) status. Patients with high AR mRNA had better PFS ($p = 0.05$). Patients expressing high ER and AR levels had better PFS than those expressing low ESR1 and AR ($p = 0.03$). These findings suggest that RNA analysis can be an alternative to IHC and FISH and provides continuous data that can better determine cut-off points for predicting response to ADC.

KEYWORDS
androgen receptor, breast cancer, ERBB2, estrogen receptor, HER2, next generation sequencing, progression-free survival

1 | INTRODUCTION

Recently, the FDA approved trastuzumab deruxtecan (T-DXd) as the first targeted therapy for patients with HER2-low metastatic breast cancer.[1–5] T-DXd is an antibody-drug conjugate (ADC), in which humanized anti-HER2 antibody is linked to a maleimide peptide linker attached to a cytotoxic payload (DX-8951f).[6] Binding to HER2-positive tumor cells leads to drug endocytosis. Upon entering the cells, the linker is clipped, releasing the cytotoxic payload in the cells. Linker clipping is performed mainly by cathepsin. Cathepsins are overexpressed in breast cancer; therefore, selective targeting of breast cancer cells and limited systemic toxicity are expected. The payload-to-antibody ratio is particularly high (~7–8).[6] This class of drugs is changing the conventional ADC-based therapy and suggests that the level of

expression of HER2, and not gene amplification, is adequate for predicting response to the new class of therapeutic ADCs.[1–6]

Measuring the targeted biomarker for such an approach requires more precise methods than immunohistochemistry (IHC) or gene amplification.[7] IHC is reliable in differentiating positivity from negativity, but its performance as a quantitative test is very poor and depends on many factors, including sample processing, fixative type, antigen retrieval, time in processing and post-processing, and the type of antibody used.[8,9] Recent advances in next generation sequencing (NGS) allow us to evaluate the transcriptome and quantify RNA in fresh and formalin-fixed paraffin embedded (FFPE) cells with relatively high accuracy and reproducibility.[10–13] RNA sequencing and quantification using NGS is more reliable and reproducible than older technologies such as microarrays or polymerase chain reaction (PCR)-based RNA quantification. Targeted RNA sequencing (targeted transcriptome) rather than the whole transcriptome sequencing allows us to focus on relevant oncogenic markers and to sequence at a deeper level for better quantification of low-level expressor genes that might be major regulators of complex cell biology.[11–13]

In this paper, we explored the value of using targeted transcriptome to measure the main breast cancer biomarkers and compared our findings with IHC and fluorescent in situ hybridization (FISH).

2 | METHODS

2.1 | Patients and data

The case descriptions were collected by COTA, Inc. This included 57 patients with breast cancer evaluated clinically at a referral cancer center (John Theurer Cancer Center, Hackensack, NJ) and the DNA and RNA of their tissue samples were evaluated by the Genomic Testing Cooperative using NGS. Patient characteristics are listed in Table S1. All patients were treated with standard of care therapy by a subspecialized oncologist. NGS testing was performed on stored FFPE primary diagnostic samples. Sections of the same FFPE block were used for extracting RNA/DNA and for IHC evaluation. Tumor cells were enriched by macrodissection before extracting RNA/DNA for sequencing.

2.2 | Next generation sequencing

DNA and RNA were extracted from FFPE tissues using an Agencourt FormaPure Total 96-Prep Kit and automated KingFisher Flex, following the manufacturer's recommendations. The Agencourt FormaPure Kit provided a split protocol to extract both DNA and RNA from the same FFPE lysate. The study protocol was approved by the Institutional Review Board (IRB) of the Western Copernicus Group (New England IRB, Aspire IRB, and Midlands IRB) (Number 1-1476184-1). The need for informed consent was waived due to the incidental collection and lack of risk. This study was conducted in accordance with the principles of the Declaration of Helsinki and its amendments.

The samples were selectively enriched for 1408 cancer-associated genes using reagents provided in the Illumina® TruSight® RNA pan-cancer panel (Illumina, San Diego, CA) (Table S1). cDNA was generated from the cleaved RNA fragments using random primers during first- and second-strand synthesis. The sequencing adapters were ligated into the resulting double-stranded cDNA fragments. The coding regions of expressed genes were captured from this library using sequence-specific probes to create the final library. Sequencing was performed using an Illumina Novaseq (Illumina, San Diego, CA). A minimum of ten million reads per sample was obtained in a single run, and the read length was 2 × 75 bp. An expression profile was generated from the sequencing coverage profile of each sample using the Cufflinks software. Expression levels were measured as fragments per kilobase of transcripts per million. Reproducibility of quantification was established by testing a control sample (Universal Human Reference RNA) 71 times and demonstrating Pearson correlation coefficient of 0.9980. Gene amplification was determined based on DNA copy number evaluation. Gene amplification was called when four copies or more were detected in the sample. The mRNA level was determined to be corresponding to gene amplification based on the DNA findings.

3 | RESULTS

3.1 | Marked variation in *ERBB2* (HER2) mRNA between IHC groups

The evaluation of HER2 mRNA using NGS showed significant variation between samples. As shown in Figure 1, all cases (40 of 57, 70%) classified by IHC and FISH studies as negative for gene amplification (<+2 by IHC) showed significantly lower levels than cases classified as amplified by FISH study (Wilcoxon matched pairs test, $p = 0.001$). Two of the 13 (15%) cases classified by IHC as equivocal (+2) showed gene amplification and overexpression, consistent with HER2 amplification. As shown in Figure 1, a significant overlap was observed between 0, 1+, and 2+. Overall, there was a significant ($p = 0.0001$) difference in *ERBB2* mRNA expression between HER2-amplified and HER2-negative cases. Using a cut-off (850 FPKM), all the cases classified as HER2-amplified were also cases classified by NGS as HER2-amplified. More importantly, DNA sequencing data can be used to determine the presence of *ERBB2* gene amplification and can always be combined with RNA data. None of the cases with *ERBB2* mRNA expression <850 FPKM showed evidence of gene amplification. Two cases with *ERBB2* levels greater than 800 but less than 850 FPKM had an extra copy of HER2 but not amplification.

There was significant variation between cases with the same IHC score, as illustrated in Figure 1 and Table S2. The RNA level did not differ significantly (Wilcoxon matched pairs test) between cases with a score of 0 and to 0-to-1 ($p = 0.32$) or between a score of 1 and 2 ($p = 0.31$). There was a significant difference between 0 and 1 ($p = 0.03$) and between 2 and 3 ($p = 0.007$). More importantly, there was also significant variation within each group. Interquartile

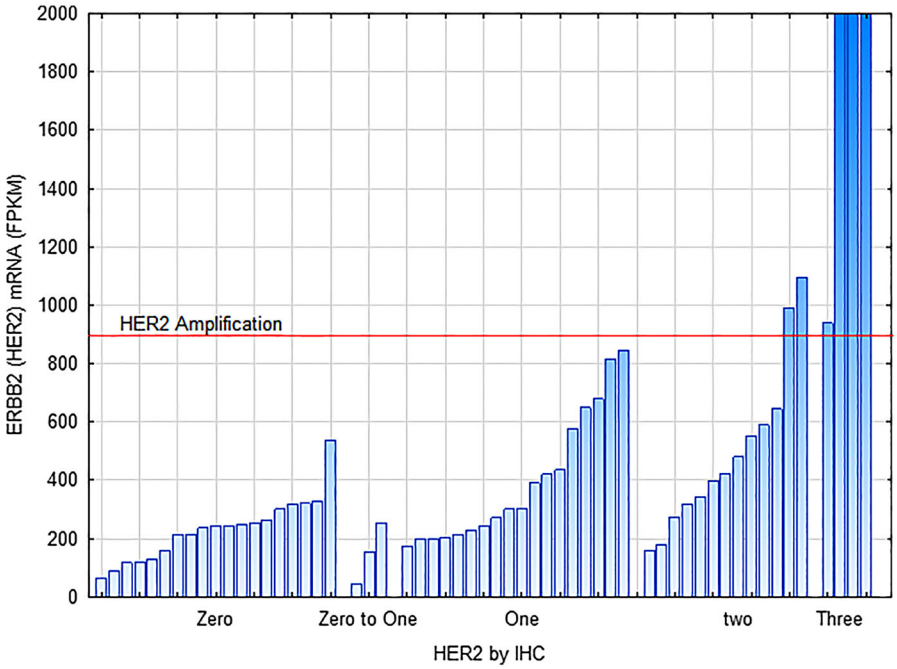

FIGURE 1 *ERBB2* mRNA levels in various HER2 IHC scores. In each score, cases are shown from lowest to highest *ERBB2* level. The cut-off point for expression associated separating cases with gene amplification is shown.

variation and standard deviation were high (Table S2), reflecting significant variation between cases within the amplification-negative cases. To evaluate the clinical relevance of this variation in expression, we studied PFS in these patients. Patients with very low *ERBB2* levels (<211 FPKM, lower quartile) had significantly shorter PFS than patients with higher *ERBB2* mRNA levels, irrespective of whether they were amplified (Figure 2). Multivariate model incorporating ERBB2 mRNA levels and disease stage showed that the positive correlation between ERBB2 and PFS was independent from staging ($p = 0.009$) (Table S3).

3.2 | Marked variation in *ESR1* mRNA within ER-positive breast cancers

Of the 55 cases that had IHC ER data, 32 (58%) were ER-positive and 23 (42%) were ER-negative. The percentage of ER-positive cells correlated significantly with the level of ER mRNA, as detected by NGS ($R = 0.072$, $R^2 = 0.50$, $p < 0.00001$). However, some overlap in RNA levels was noted between ER-positive and ER-negative samples, as classified by IHC. Four cases classified by IHC as positive had RNA levels below 22 FPKM (mean +2 standard deviation), and most likely should have been classified as negative. One case classified as negative by IHC showed *ESR1* mRNA expression at 33 FPKM, and most likely should have been classified as positive based on the overall pattern of expression. This suggests that the positive predictive value (PPV) was 88% (95% confidence interval (CI): 70%–96%) and the negative predictive value (NPV) was 96% (95% CI: 96%–100%). Most importantly, there was a significant variation in *ESR1* RNA within cases classified as ER-positive by IHC (see Figure 3 and Table S4). The degree of variation can be captured by evaluating the percentage of positive cells, but significant variation can be seen between cases classified by IHC as having >90% positivity, as shown in Table S2. There was a marked variation (standard deviation and interquartile range) within this group of cases. However, we could not find any specific clinical relevance of this variation in this group of patients.

3.3 | Androgen receptor correlation with *ESR1* levels

The androgen receptor (AR) is believed to play a significant role in breast cancer pathophysiology.[14–17] More importantly AR can potentially be a therapeutic target in the treatment of breast cancer. A dynamic relationship between the ER and AR has been reported in breast cancer. The two receptors form heterodimers and bind to the same DNA sequence, thereby activating specific pathways. Therefore, these two genes transcriptionally regulate each other. However, AR expression is not routinely evaluated in the management of breast cancer patients. We explored the relationship between *ESR1* and *AR* expression in breast cancer samples that were classified as ER-positive and ER-negative.

As shown in Figure 4, there was no significant correlation between *ESR1* and *AR* ($r^2 = 0.01$, $p = 0.23$). With few exceptions, most *ESR1*-negative cases were also *AR*-negative. However, cases that were *ESR1*-positive varied significantly in *AR* expression. To explore the clinical relevance of AR expression, as measured by RNA sequencing, we evaluated PFS in patients with high AR and compared them with

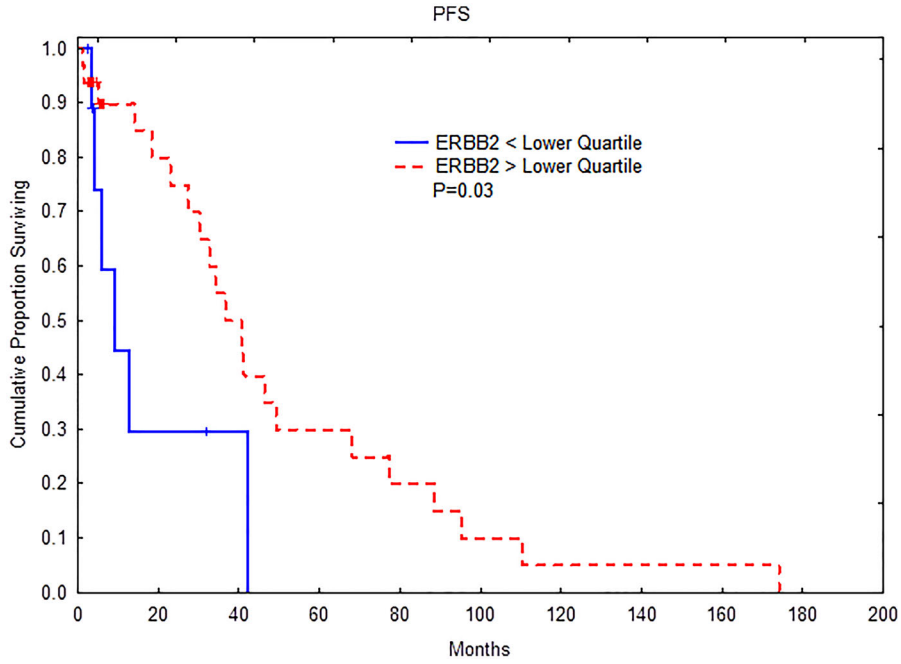

FIGURE 2 Kaplan–Meier plot showing PFS of patients with breast cancer. Patients ($N = 42$) with high *ERBB2* mRNA levels as detected by transcriptomic study have better PFS than patients with low *ERBB2*, irrespective of amplification status. All patients with available PFS data are included without any selection.

patients expressing low AR. As shown in Figure 5, a significantly longer PFS was observed in patients with high AR expression. Multivariate model incorporating AR mRNA levels and disease stage showed that the positive correlation between AR and PFS was independent from staging ($p = 0.02$) (Table S3). This data demonstrates that RNA expression as detected by NGS shows data similar to that reported using IHC and can replace IHC.[14–17]

Based on the concept that both ER and AR may cross-talk to regulate breast cancer growth, we evaluated the clinical relevance of *ESR1* and *AR* when both were high and compared them when both were low. As shown in Figure 6, patients with breast cancer expressing high levels of both *ESR1* and *AR* had significantly better overall PFS than patients with low levels of both *ESR1* and *AR*.

4 | DISCUSSION

Precision oncology requires accurate and reliable biomarkers that can provide precise information on the individual cancer and the patient (host).[18,19] With improvements in the design and development of more targeted therapeutic approaches, old biomarkers that are measured in proximity and not precisely may not be adequate for new classes of therapeutic agents. For example, antibody-drug conjugates such as T-DXd rely on the high load of cytotoxic agents rather than on the high level of the targeted antigen and rely mainly on the specific expression of the antigen to internalize the cytotoxic drug and destroy the cancer cells.[1–5] IHC is a qualitative approach to determine the expression of specific antigens, but it is a very crude way to quantify expression in cells. Furthermore, IHC is influenced by various methods of tissue fixation and processing, making reproducibility very difficult.[7] Recent advances in RNA sequencing and the adaptation of NGS in routine clinical molecular profiling allow for the use of reliable quantitative RNA data from fresh tissue and FFPE tissue.[10–13] Unlike the array approach, NGS practically counts the RNA molecules in the analyzed sample and does not depend on hybridization and colorimetric detection. We explored the potential of using RNA levels measured by NGS as a replacement for common biomarkers used in breast cancer IHC and FISH.

Using real-world data, we evaluated HER2 results as reported by IHC and FISH and compared the results with those of RNA as determined by NGS. As demonstrated in Figure 1, significant variation in the levels of *EBB2* mRNA can be seen between cases classified as negative for HER2 amplification or classified by IHC as grade 0, 1, or 2. Even cases scored as grade 3 (amplified) showed variation in *ERBB2* levels. A significant overlap can also be seen between cases classified as 0, 1+, or 2+. Therefore, IHC 0, 1+, and 2+ did not accurately represent discrete groups. More importantly, this variation in the levels is clinically relevant. Despite the small number of cases, a statistically significant difference in PFS was observed between low *ERBB2* expressers and high expressors. These data suggest that RNA evaluation is an ideal approach for predicting cases that may respond to ADC-based therapy, because it provides a more precise measurement of *ERBB2* mRNA levels.

Evaluation of ER expression also showed significant variation between cases classified as ER-positive by IHC. Considering the current standard, which indicates that breast cancers with 1%–100% nuclear-positive cells by IHC are categorized as positive for ER, variation in ER levels is expected. At the same time, the difference between

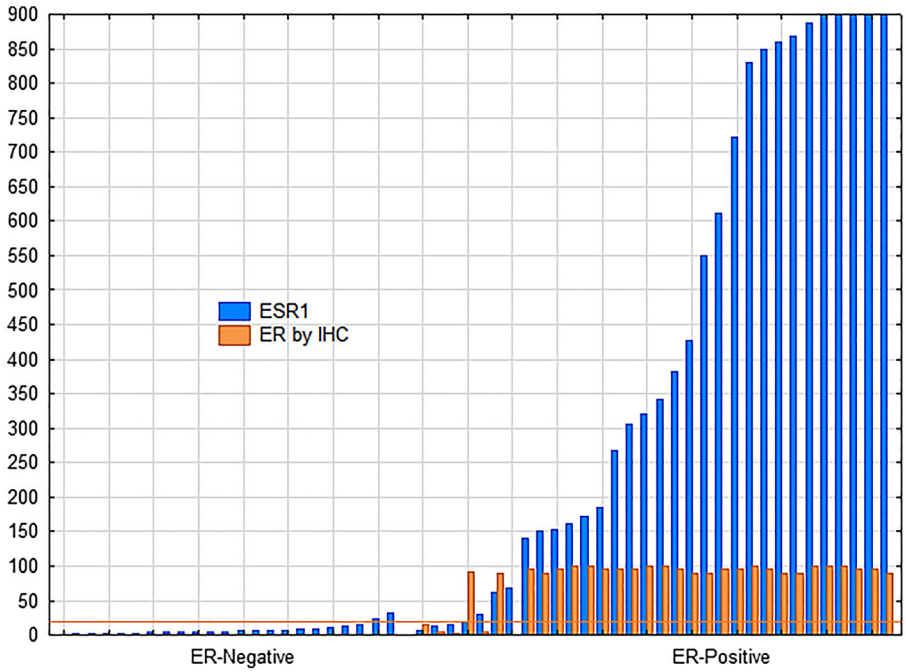

FIGURE 3 *ESR1* mRNA levels in breast cancer patients. ER-negative cases by IHC are shown on the left and ER-positive on the right. In each group, *ESR1* levels are shown from the lowest to highest. The percent of ER positive cells in each case is shown as orange bar. The cut-off point for positive vs negative as determined by *ESR1* expression is indicated with red line.

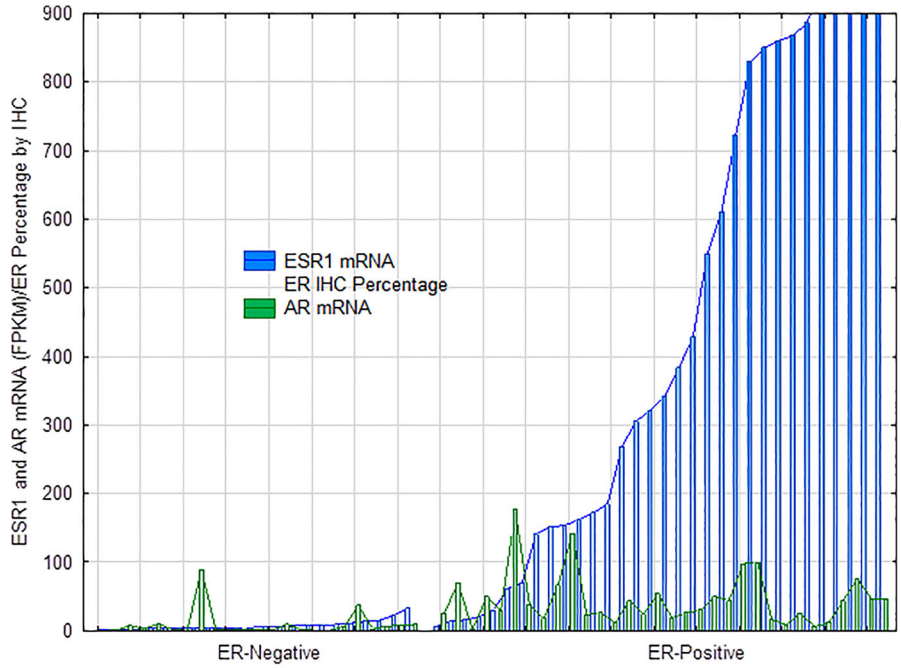

FIGURE 4 *ESR1* and *AR* mRNA levels in breast cancer patients. ER-negative cases by IHC are shown on the left and ER-positive on the right. In each group, *ESR1* levels are shown as blue bars from the lowest to highest. The corresponding *AR* mRNA levels are shown as green bars. Lack of correlation between *ESR1* and *AR* is demonstrated.

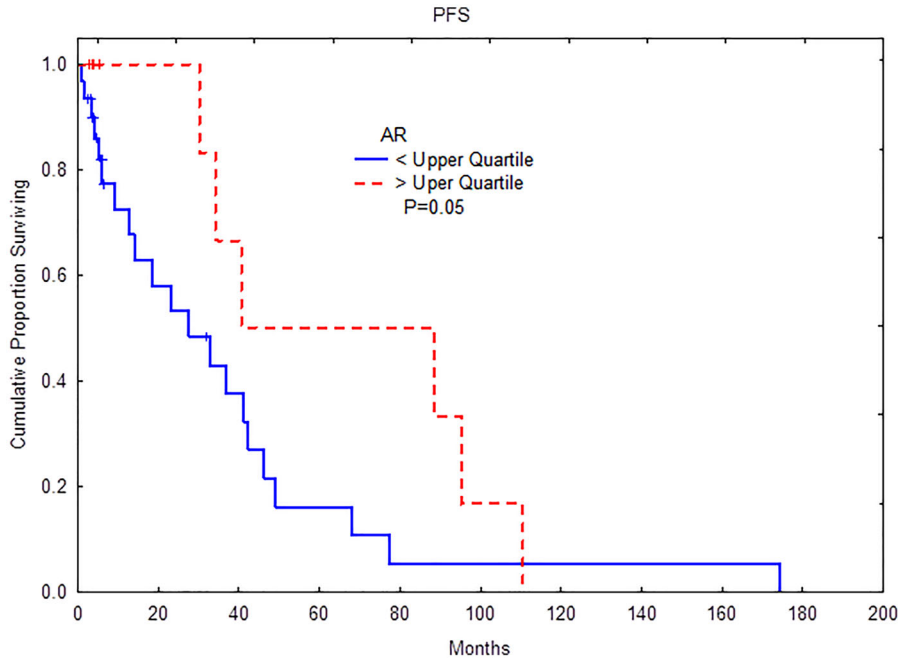

FIGURE 5 Kaplan–Meier plot showing PFS of patients with breast cancer. Patients ($N = 42$) with high *AR* mRNA levels as detected by transcriptomic study have better PFS than patients with low *AR*.

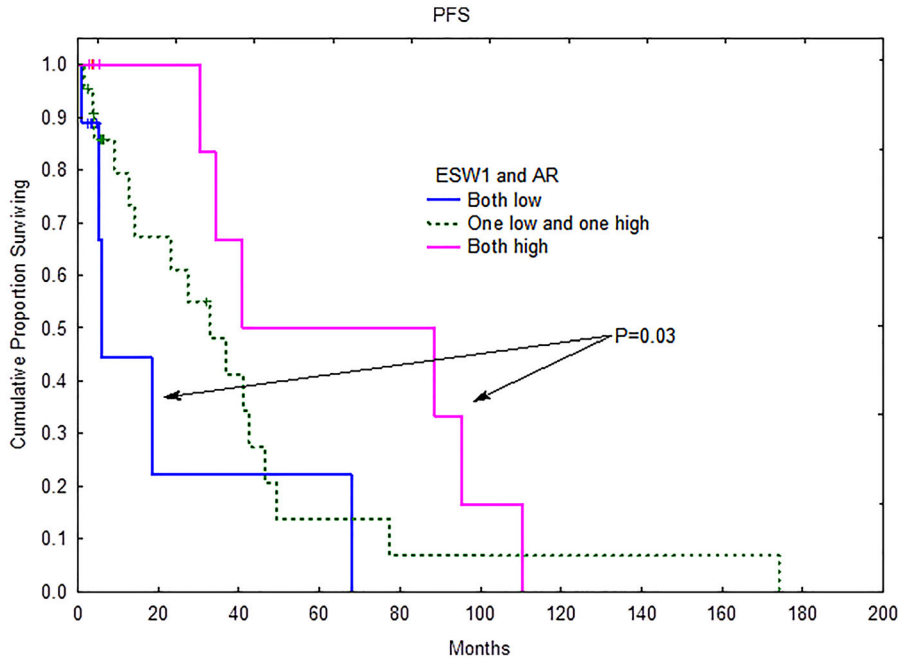

FIGURE 6 Kaplan–Meier plot showing PFS of patients with breast cancer. Patients ($N = 42$) with high levels of both *ESR1* and *AR* mRNA levels show significantly better PFS as compared with patients with low *ESR1* and *AR*.

1% and less than 1% was very tight, and cases with 1% positivity were lumped with cases with >90% positivity. Based on RNA data, the PPV of ER by IHC was 88% (95% CI: 70%–96%), and the NPV was 96% (95% CI: 96%–100%). In addition, the variation in levels is huge and RNA profiling provides more precise quantification. Clinicians routinely consider the ER percentage and PR percentage when deciding on neoadjuvant and adjuvant therapy. Providing reproducible measurement of these receptors with a broad dynamic range may improve the selecting patients for such therapeutic approaches.

Similarly, we found significant variations in AR levels in patients with breast cancer. We also demonstrated that there was no correlation between AR and ER. Some studies have suggested that patients with

a high AR may respond to anti-AR therapy.[14,20,21] Determining which patients will most likely respond to anti-AR can be achieved by evaluating AR levels.[14,22] RNA evaluation using NGS provides precise values for AR levels. Using AR as a biomarker in breast cancer not only provides information on the possible response to anti-androgens but may also provide prognostic information.[14,20] Patients with high AR levels showed significantly ($p = 0.03$) better PFS than those with low AR levels, irrespective of ER levels. Furthermore, patients with high levels of both ER and AR showed significantly better PFS than those with low levels of both ER and AR.

In summary, RNA profiling provides a reliable alternative testing option for IHC and FISH testing. This provides more precise information and a better dynamic range that is potentially valuable as a biomarker for new therapeutic approaches. Currently, the cost of molecular profiling using NGS is considerably higher than that of IHC. However, with the increasing utilization and dependence on using DNA and RNA profiling for personalized medicine and selecting targeted or checkpoint therapy, the expression data of various biomarkers used in IHC are practically the byproducts of such molecular profiling and can be used for clinical decisions after developing appropriate standards. The number of cases in this study is too small and these conclusions need to be confirmed by additional studies using significantly more number of cases.

ACKNOWLEDGMENTS

The authors thank the team at COTA (Real-World Data Company) for providing the data and for the careful collection of various data points.

CONFLICT OF INTEREST STATEMENT

MA, AC, and WM work and own stocks in a diagnostic company that offers RNA sequencing. AG consulting/advisory board/honoraria AstraZeneca, SecuraBio, TG Therapeutics, not relevant to this work. AG, AI, DG, DM, and SW have no relevant conflict of interest.

DATA AVAILABILITY STATEMENT

Study protocol and data are available upon request.

PRECIS

The current study demonstrates that targeted transcriptomic data can be used reliably in evaluating hormonal expression and HER2 amplification and expression in breast cancer. Transcriptomic data provides broad dynamic range in evaluating the levels of HER2 and hormone receptors in breast cancer.

ORCID

Maher Albitar MD https://orcid.org/0000-0002-0050-2030

REFERENCES

1. Modi S, Park H, Murthy RK, Iwata H, Tamura K, Tsurutani J, et al. Antitumor activity and safety of trastuzumab deruxtecan in patients with HER2-low-expressing advanced breast cancer: results from a phase Ib study. J Clin Oncol. 2020;38(17):1887–96. https://doi.org/10.1200/JCO.19.02318
2. Ferraro E, Drago JZ, Modi S. Implementing antibody-drug conjugates (ADCs) in HER2-positive breast cancer: state of the art and future directions. Breast Cancer Res. 2021;23(1):84. https://doi.org/10.1186/s13058-021-01459-y
3. Adams E, Wildiers H, Neven P, Punie K. Sacituzumab govitecan and trastuzumab deruxtecan: two new antibody-drug conjugates in the breast cancer treatment landscape. ESMO Open. 2021;6(4):100204. https://doi.org/10.1016/j.esmoop.2021.100204
4. Lee J, Park YH. Trastuzumab deruxtecan for HER2+ advanced breast cancer. Future Oncol. 2022;18(1):7–19. https://doi.org/10.2217/fon-2021-0550
5. Zhang H, Karakas C, Tyburski H, Turner BM, Peng Y, Wang X, et al. HER2-low breast cancers: current insights and future directions. Semin Diagn Pathol. 2022;39(5):305–12. https://doi.org/10.1053/j.semdp.2022.07.003
6. Azar I, Alkassis S, Fukui J, Alsawah F, Fedak K, Al Hallak MN, et al. Spotlight on trastuzumab deruxtecan (DS-8201,T-DXd) for HER2 mutation positive non-small cell lung cancer. Lung Cancer (Auckl). 2021;12:103–14. https://doi.org/10.2147/LCTT.S307324
7. Prat A, Pascual T, De Angelis C, Gutierrez C, Llombart-Cussac A, Wang T, et al. HER2-enriched subtype and ERBB2 expression in HER2-positive breast cancer treated with dual HER2 blockade. J Natl Cancer Inst. 2020;112(1):46–54. https://doi.org/10.1093/jnci/djz042
8. Jensen SG, Thomas PE, Christensen IJ, Balslev E, Hansen A, Høgdall E. Evaluation of analytical accuracy of HER2 status in patients with breast cancer: comparison of HER2 GPA with HER2 IHC and HER2 FISH. APMIS. 2020;128(11):573–82. https://doi.org/10.1111/apm.13076
9. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. Mod Pathol. 2008;21(2):S8–S15. https://doi.org/10.1038/modpathol.2008.34
10. Sarantopoulou D, Brooks TG, Nayak S, Mrčela A, Lahens NF, Grant GR. Comparative evaluation of full-length isoform quantification from RNA-Seq. BMC Bioinformatics. 2021;22(1):266. https://doi.org/10.1186/s12859-021-04198-1
11. Govindarajan M, Wohlmuth C, Waas M, Bernardini MQ, Kislinger T. High-throughput approaches for precision medicine in high-grade serous ovarian cancer. J Hematol Oncol. 2020;13(1):134.
12. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol. 2011;30(1):99–104.
13. Reeser JW, Martin D, Miya J, Kautto EA, Lyon E, Zhu E, et al. Validation of a targeted RNA sequencing assay for kinase fusion detection in solid tumors. J Mol Diagn. 2017;19(5):682–96.
14. Anestis A, Zoi I, Papavassiliou AG, Karamouzis MV. Androgen receptor in breast cancer-clinical and preclinical research insights. Molecules. 2020;25(2):358. https://doi.org/10.3390/molecules25020358
15. Tesei A, Castoria G. Editorial: the androgen receptor in breast cancer. Front Endocrinol (Lausanne). 2020;11:636480. https://doi.org/10.3389/fendo.2020.636480
16. Chen M, Yang Y, Xu K, Li L, Huang J, Qiu F. Androgen receptor in breast cancer: from bench to bedside. Front Endocrinol (Lausanne). 2020;11:573. https://doi.org/10.3389/fendo.2020.00573
17. Hickey TE, Selth LA, Chia KM, Laven-Law G, Milioli HH, Roden D, et al. The androgen receptor is a tumor suppressor in estrogen receptor-positive breast cancer. Nat Med. 2021;27(2):310–20.
18. Bravo-Merodio L, Acharjee A, Russ D, Bisht V, Williams JA, Tsaprouni LG, Gkoutos GV. Translational biomarkers in the era of precision medicine. Adv Clin Chem. 2021;102:191–232. https://doi.org/10.1016/bs.acc.2020.08.002
19. Zou J, Wang E. Cancer biomarker discovery for precision medicine: new progress. Curr Med Chem. 2019;26(42):7655–71. https://doi.org/10.2174/0929867325666180718164712

20. Anestis, A., Karamouzis, MV, Dalagiorgou, G, Papavassiliou, AG. Is androgen receptor targeting an emerging treatment strategy for triple negative breast cancer? Cancer Treat Rev. 2015;41:547–53.
21. Anestis A., Sarantis P., Theocharis S., Zoi I, Tryfonopoulos D, Korogiannos A, et al. Estrogen receptor beta increases sensitivity to enzalutamide in androgen receptor-positive triple-negative breast cancer. J Cancer Res Clin Oncol. 2019;145:1221–33.
22. Cochrane DR, Bernales S, Jacobsen BM, Cittelly DM, Howe EN, D'Amato NC, et al. Role of the androgen receptor in breast cancer and preclinical analysis of enzalutamide. Breast Cancer Res. 2014;16:R7.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Albitar M, Goy A, Pecora A, Graham D, McNamara D, Charifa A, et al. The use of transcriptomic data in developing biomarkers in breast cancer. ImmunoMedicine. 2023;e1051. https://doi.org/10.1002/imed.1051