



Alternative splicing generates HER2 isoform diversity underlying antibody–drug conjugate resistance in breast cancer

Gabriela D.A. Guardia, Carlos H. dos Anjos, Aline Rangel-Pozzo, et al.

Genome Res. 2025 35: 1942-1958 originally published online July 15, 2025

Access the most recent version at doi:[10.1101/gr.280304.124](https://doi.org/10.1101/gr.280304.124)

References This article cites 78 articles, 18 of which can be accessed free at:
<http://genome.cshlp.org/content/35/9/1942.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Alternative splicing generates HER2 isoform diversity underlying antibody–drug conjugate resistance in breast cancer

Gabriela D.A. Guardia,^{1,5} Carlos H. dos Anjos,^{1,5} Aline Rangel-Pozzo,² Filipe F. dos Santos,^{1,3} Alexander Birbrair,⁴ Paula F. Asprino,¹ Anamaria A. Camargo,¹ and Pedro A.F. Galante¹

¹Centro de Oncologia Molecular, Hospital Sírio Libanês, São Paulo 01308-050, Brazil; ²Department of Physiology and Pathophysiology, CancerCare Manitoba Research Institute, University of Manitoba, Winnipeg MB R3E 0J9, Canada;

³Department de Bioquímica, Universidade de São Paulo, São Paulo 05508-000, Brazil; ⁴Department of Dermatology, University of Wisconsin–Madison, Madison, Wisconsin 53715, USA

Breast cancer (BC) is a heterogeneous disease that can be molecularly classified based on the expression of the ERBB2 receptor (also known as *HER2*) and hormone receptors. Targeted therapies for *HER2*-positive BC, such as trastuzumab, antibody–drug conjugates (ADCs) and tyrosine kinase inhibitors, have improved patient outcomes, but primary/acquired resistance still poses challenges that can limit treatments' long-term efficacy. Addressing these obstacles is vital for enhancing therapeutic strategies and patient care. Alternative splicing, a post-transcriptional mechanism that enhances transcript diversity (isoforms), can produce proteins with varied functions, cellular localizations, or binding properties. Here, we comprehensively characterize the *HER2* alternative splicing isoforms, assess their expression in primary BC patients and cell lines, and explore their role in resistance to anti-*HER2* therapies. We expand the catalog of known *HER2* protein-coding isoforms from 13 to 90, revealing distinct patterns of protein domains, cellular localizations, and protein structures, along with their antibody-binding sites. By profiling expression in 561 primary BC samples and mass spectrometry data, we discover a complex landscape of *HER2* isoform, revealing novel transcripts that were previously unrecognized and are not assessed in routine clinical practice. Finally, the assessment of *HER2* isoform expression in BC cell cultures sensitive or resistant to trastuzumab and ADCs reveals that drug-resistant cells shift their expression toward isoforms lacking antibody-binding domains. Our results broaden the understanding of *HER2* isoforms, revealing distinct mechanisms of potential resistance to anti-*HER2* therapies, particularly ADCs. This expanded landscape of *HER2* isoforms emphasizes the crucial role of alternative splicing investigations in advancing precision-targeted cancer therapies.

[Supplemental material is available for this article.]

Breast cancer remains one of the most prevalent and challenging malignancies worldwide, with an estimated 2.3 million new cases diagnosed in 2022 (Bray et al. 2024). The heterogeneity of breast cancer at both the clinical and molecular levels has compelled its classification into distinct groups, enabling more tailored treatment approaches for improving patient outcomes (Perou et al. 2000). The molecular classification of breast cancer, primarily based on the expression of hormone receptors (HRs; estrogen and progesterone) and the human epidermal growth factor receptor 2 (*ERBB2*; also known as *HER2*), has revolutionized our understanding of this disease and guided the development of *HER2*-targeted therapies (Sørlie et al. 2001; Prat et al. 2017).

HER2-positive breast cancers, accounting for ~20% of all breast cancers, are characterized by the overexpression or amplification of the *HER2* gene (Wolff et al. 2013). Located on Chromosome 17q12, the *HER2* transcripts typically encode a 185 kDa transmembrane tyrosine kinase receptor that belongs to

the epidermal growth factor receptor (EGFR) family (Moasser 2007). Regarding functionalities, *HER2* overexpression leads to the constitutive activation of downstream signaling pathways, including PIK3CA/AKT and MAPK, promoting tumor cell proliferation and survival (Yarden and Sliwkowski 2001). The *HER2*-positive breast cancer subtype is associated with aggressive tumor behavior and poor prognosis in the absence of anti-*HER2*-targeted therapy (Slamon et al. 1987).

The advent of *HER2*-targeted therapies has significantly improved the outcomes for patients with *HER2*-positive breast cancer. Trastuzumab, a humanized monoclonal antibody targeting the extracellular domain of *HER2*, was the first targeted therapy approved for both metastatic and early-stage *HER2*-positive breast cancer (Slamon et al. 2001; Piccart-Gebhart et al. 2005). Since then, additional therapeutic classes have expanded the treatment landscape for *HER2*-positive disease, including tyrosine kinase

⁵These authors contributed equally to this work.

Corresponding author: pgalante@mochsl.org.br

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.280304.124>.

© 2025 Guardia et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

inhibitors such as lapatinib, neratinib, and tucatinib, and, more recently, HER2-targeted antibody–drug conjugates (ADCs) like trastuzumab emtansine (T-DM1) and trastuzumab deruxtecan (T-DXd), which combine the targeting precision of trastuzumab with potent cytotoxic agents (Verma et al. 2012; Modi et al. 2020). Altogether, these new therapeutic options have had a profound impact on the management of both metastatic and early-stage HER2-positive breast cancer, extending disease control and significantly reducing the risk of recurrence (Swain et al. 2015; von Minckwitz et al. 2017).

Therapeutic advancements have greatly improved outcomes for HER2-positive breast cancer, with most patients diagnosed at early stages now achieving a cure and experiencing fewer disease recurrences (von Minckwitz et al. 2017). However, in the metastatic setting, although ~16% of patients attain long-term disease control (Swain et al. 2015), the majority eventually develop resistance to anti-HER2 therapies, whether through primary resistance, in which the disease progresses shortly after treatment initiation, or acquired resistance, in which resistance emerges following an initial period of response.

The mechanisms underlying the resistance to HER2-targeted therapies are complex and multifaceted. *HER2* mutations in the kinase domain can reduce the effectiveness of treatments by altering the receptor's structure (Marín et al. 2023). Compensatory signaling pathways, such as upregulation of *ERBB3* (also known as *HER3*) and *IGFR* or mutations in the PIK3CA/AKT pathway, allow cancer cells to bypass HER2 inhibition (Nagata et al. 2004; Cizkova et al. 2013; Mishra et al. 2018). Tumor cells can also evade the immune response, particularly antibody-dependent cellular cytotoxicity, by downregulating immune-recognition molecules or recruiting immunosuppressive cells (Loi et al. 2013). The tumor microenvironment, characterized by fibrosis or hypoxia, can also act as a physical barrier to drugs like T-DM1 and T-DXd (Sonnenblick et al. 2020). Finally, some studies suggest that *HER2* transcripts generated by alternative splicing code protein isoforms with enhanced dimerization and modification in their antibody (trastuzumab)-binding domain, which are associated with drug resistance (Scaltriti et al. 2007; Turpin et al. 2016). Thus, a deeper understanding of these resistance mechanisms remains a critical challenge in overcoming and managing breast cancer tumors expressing *HER2*.

Alternative splicing, a fundamental post-transcriptional process in eukaryotic gene expression, allows a single gene locus to produce multiple distinct mRNA transcripts, increasing protein diversity (Nilsen and Graveley 2010). In humans, >90% of multi-exon genes undergo alternative splicing, generating roughly seven mRNA isoforms per gene on average, although only a fraction of these yield functionally distinct proteins (Pan et al. 2008; Uhlén et al. 2015; Reixachs-Solé and Eyra 2022). In the context of cancer, aberrant splicing has been implicated in various aspects of tumor biology, including drug resistance (Sveen et al. 2016; Marcelino Meliso et al. 2017; Dvinge et al. 2019).

In breast cancer, alternative splicing affects numerous genes involved in key cellular processes, including apoptosis, cell cycle regulation, and signal transduction (Yang et al. 2019). Specifically, *HER2* alternative splicing has gained particular attention in the context of resistance to HER2-targeted therapies. *HER2* splice variants have been identified in large scale (Veiga et al. 2022), but most of the studies have been focused on the p95HER2 and delta16 *HER2* ($\Delta 16$ HER2) isoforms (Scaltriti et al. 2007; Turpin et al. 2016). p95HER2 (also known as CTF611 or HER2-CTF; here and after, P95) is an incomplete isoform of the HER2 protein (molecular

weight of ~95 kDa) lacking the extracellular protein domain and being constitutively active (Molina et al. 2001; Arribas et al. 2011). P95's expression has been associated with poor prognosis and resistance to trastuzumab, as it lacks the antibody-binding site (Scaltriti et al. 2007). $\Delta 16$ HER2 results from skipping the *HER2* exon 16, which encodes a small portion of the extracellular domain (Kwong and Hung 1998). The $\Delta 16$ HER2 isoform assembles stable homodimers and is associated with increased transforming activity and metastatic potential (Turpin et al. 2016). Controversially, studies suggest that $\Delta 16$ HER2 may contribute to trastuzumab resistance (Mitra et al. 2009), whereas others have found that it may enhance sensitivity to specific HER2-targeted therapies (Castagnoli et al. 2014). Thus, the complex interplay between *HER2* splicing isoforms in the context of drug response and resistance requires more extensive and in-depth investigation.

Thus, despite significant advances in HER2-targeted therapies for breast cancer, resistance mechanisms remain incompletely understood, limiting treatment efficacy for many patients. Alternative splicing of *HER2* may represent a critical but underexplored dimension of this challenge. This study aims to comprehensively characterize the landscape of *HER2* alternative splicing isoforms in breast cancer and to investigate their potential role in mediating resistance to anti-HER2 therapies, particularly ADCs. By employing long-read sequencing technology and multidimensional analysis approaches, we seek to identify the structural and functional properties of *HER2* splicing variants and determine how their expression patterns relate to treatment response. Through this investigation, we aim to provide a new framework for understanding resistance mechanisms and potentially improve patient stratification for HER2-targeted therapies.

Results

Assessing the *HER2* splicing isoform diversity in breast cancer

Globally, our study comprised five main steps (Fig. 1A). First, we expanded the known repertoire of *HER2* splicing isoforms using long-read sequence data from breast tumors (Veiga et al. 2022). Second, we used computational models, classical and based on deep learning, to characterize the main features of the proteins encoded by these *HER2* splicing isoforms, including their functional domains, structural elements, and cellular localization. Third, we analyzed the isoform's expression profile and their translational evidence in a large set of primary breast tumors. Fourth, we evaluated *HER2* isoform expression in breast cancer cell cultures that were sensitive or resistant to trastuzumab, T-DM1, and T-DXd. Finally, we compared the expression patterns of *HER2* isoforms before and after the emergence of resistance in cell lines exposed to T-DM1 treatment. To ensure reliable isoform expression data, we carefully selected and stratified samples based on technical considerations (e.g., distinct library preparation strategies or low number of mapped reads), confounding effects (e.g., excluding male samples), and biological information about the tumors (HR and HER2 status). We grouped the samples into HR-positive (HR⁺) and HR-negative (HR⁻) categories and further subgrouped them based on their *HER2* expression status: HER2-high (HER2⁺⁺⁺ in immunohistochemistry or fluorescence in situ hybridization [FISH]/ISH amplified), HER2-low (HER2⁺ or HER2⁺⁺, without FISH/ISH amplification), and HER2-zero (no staining), as shown in Figure 1B. This stratification was necessary to ensure fair comparison among breast cancer subtypes, given the distinct evolution and prognosis expected for different patient groups.

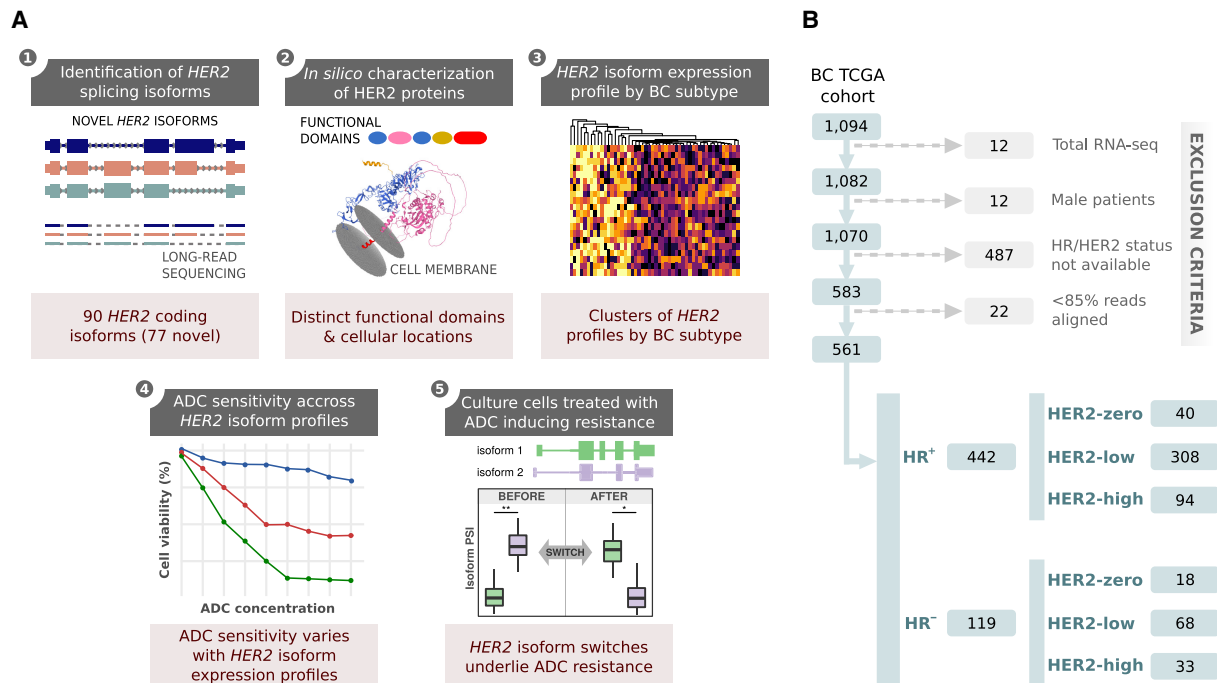


Figure 1. A comprehensive strategy for investigating *HER2* isoform diversity in breast cancer. (A) Five-step approach to characterize *HER2* splicing isoforms in breast cancer patients and cell cultures: (1) identification of *HER2* splicing isoforms, (2) *in silico* characterization of *HER2* protein variants, (3) *HER2* isoform expression profiling by breast cancer subtype, (4) analysis of antibody–drug conjugate (ADC) sensitivity across *HER2* isoform profiles, and (5) examination of *HER2* isoform switches in ADC-induced resistance. (B) RNA-sequenced sample selection and stratification based on technical (e.g., distinct library preparation strategies and low number of mapped reads) and biological characteristics (e.g., male samples were excluded, and tumors with same HR status and *HER2* expression levels were grouped).

Expanding the range of functional and structural variants of *HER2* splicing isoforms

In the human reference transcriptome GENCODE V36 (Frankish et al. 2019), the *ERBB2* (*HER2*) gene spans ~42.5 kb on Chromosome 17 (Chr 17: 39,687,914–39,730,426; reference genome version, GRCh38), comprising 27 canonical exons and 13 distinct protein-coding isoforms, along with nine noncoding isoforms. Importantly, the annotation of *HER2* isoforms remained unchanged from GENCODE V36 to GENCODE V47 (Kaur et al. 2024), which includes an expanded catalog of transcript annotations based on long-read RNA sequencing (RNA-seq) data. By using full-length mRNA transcripts (see Methods) (Veiga et al. 2022), we have significantly expanded the known repertoire of protein-coding *HER2* splicing isoforms from 13 to 90. In terms of alternative splicing classes, these isoforms primarily result from exon skipping (ES; 40 isoforms), alternative 5' splice sites (A5s; 26 isoforms), alternative 3' splice sites (A3s; 15 isoforms), and other types (12 isoforms) (Fig. 2A). We observed alternative splicing events in specific *HER2* exons, irrespective of the protein domains they encode. Twenty-three exons (85.1%; 23/27) of the canonical isoform showed evidence of alternative splicing (Fig. 2A, colored exons; Supplemental Table S1).

To gain insights into their functional features, we categorized *HER2* splicing isoforms into 13 distinct groups based on key characteristics of the proteins they encode: (1) completeness of *HER2* protein domains (receptor L, furin-like, growth factor receptor, and protein kinase), (2) predicted cellular localization (cell membrane, cytoplasm, extracellular, or Golgi apparatus), (3) presence of the trastuzumab-binding domain, and (4) presence of the immunohistochemical (IHC) ligand domain (for details, see

Methods) (Fig. 2B; Supplemental Table S2). Each group represents isoforms sharing similar functional properties but displaying distinct combinations of these features. This classification system revealed substantial heterogeneity in *HER2* isoform structural and functional characteristics.

The majority of isoforms (57.8%, 52/90; groups 1 to 6) retain cell membrane localization, but 10 (11.1%; groups 5, 6) lack the trastuzumab-binding site. Curiously, 16 (17.8%; groups 4 and 6) isoforms do not contain the binding domain of antibodies used for IHC staining, which may affect *HER2* expression determination by IHC, a critical factor in breast tumor classification. Finally, only 15 isoforms (16.7%) retain all canonical protein domains, suggesting that most alternative isoforms may have some degree of altered functional properties (Fig. 2B). To illustrate how the combination of multiple splicing events (e.g., ES and intron retention) shape *HER2* isoform diversity, Supplemental Figure S1 shows the splicing event combinations of a representative isoform for each isoform group (1 to 13).

HER2 has two well-studied variants, p95 and $\Delta 16$ *HER2*. Here, we took advantage of the full-length transcript analysis to examine beyond individual splicing events (e.g., ES), revealing complete transcripts that combine multiple alternative splicing events and may encode proteins different from those predicted by examining isolated events alone. Using this approach, we identified nine distinct isoforms that lost exon 16 ($\Delta 16$ *HER2* variants): only one reported in GENCODE (ENST00000580074.1) and eight novel isoforms discovered through our long-read sequencing strategy (Supplemental Fig. S2A). These $\Delta 16$ *HER2* isoforms are distinct because they exhibited additional alternative splicing events, including alternative first exon usage, A5 selection, skipping of exons 19 and 24, and, in one case, three additional alternative

Alternative splicing in anti-HER2 therapies

splicing events. For p95HER2 (p95), our analysis revealed eight splicing isoforms that encode p95-like proteins (Supplemental Fig. S2B). These isoforms also demonstrated complex splicing pat-

terns, including alternative first exon usage (one isoform), ES (one isoform), combinations of alternative first exon with other splicing events (six isoforms), and upstream small ORFs (see below).

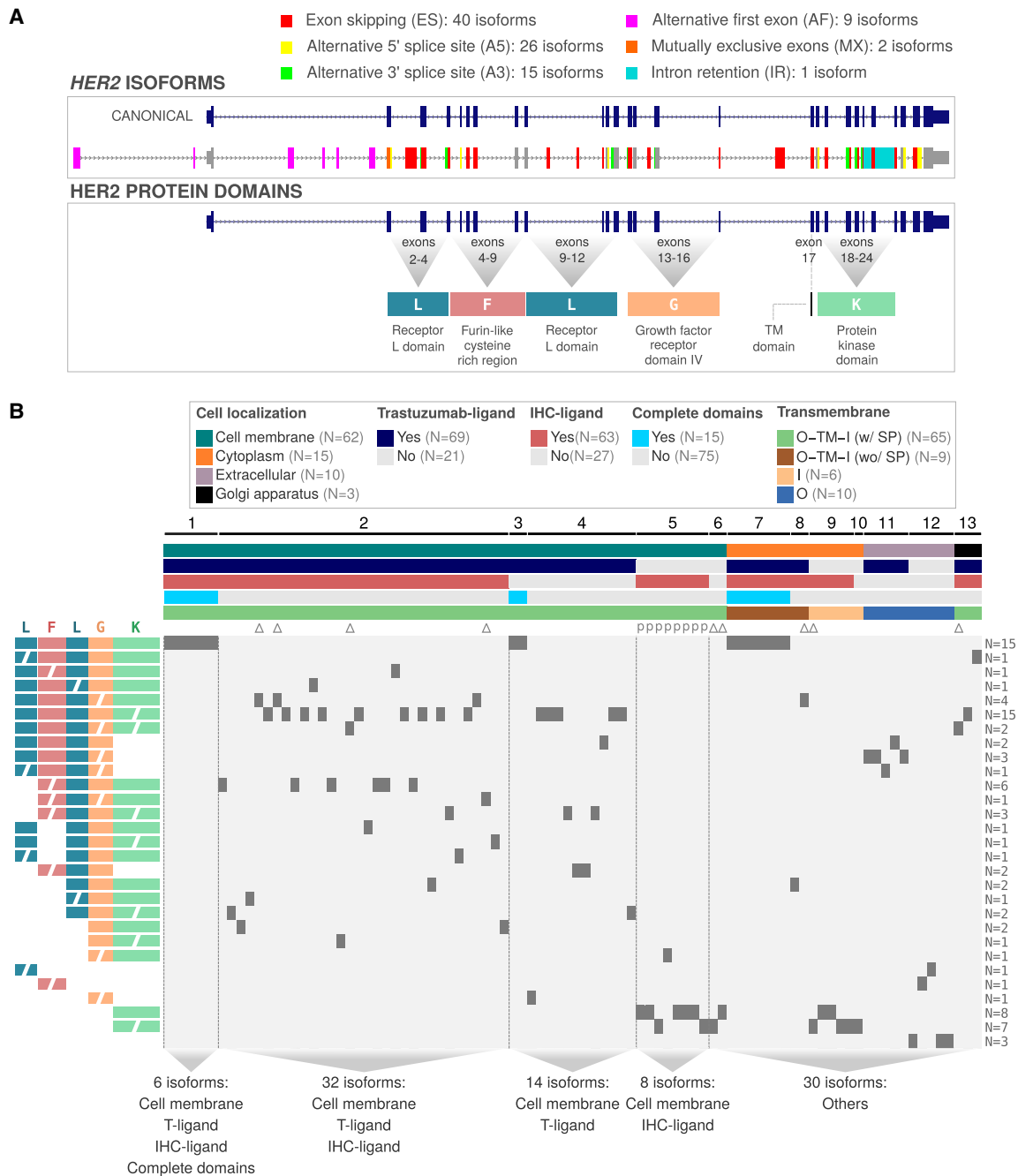


Figure 2. Comprehensive analysis of *HER2* isoforms and their characteristics. (A) Structure of the *HER2* canonical isoform and its splicing variants. The top panel displays the canonical *HER2* isoform and alternative splicing events. The bottom panel illustrates the protein domains encoded by specific exons. (B) Structural and functional properties of *HER2* isoform-encoded proteins. The top bars represent five characteristics of isoform-encoded proteins (e.g., cellular localizations, presence of trastuzumab and immunohistochemistry [IHC] binding regions, presence of complete protein domains and transmembrane topology: [O] outside, [TM] transmembrane region, [I] inside, [SP] signal peptide). The groups of isoform-encoded proteins (from 1 to 13), created based on those five characteristics, are discriminated above the top bars. Below the top bars, the isoforms p95 and $\Delta 16$ are represented by "P" and " Δ ," respectively. The heatmap shows the presence (dark gray squares) or absence (light gray background) of specific protein domain configurations (rows) for each isoform-encoded protein (columns). Colored and lettered protein domain configurations are represented on the left side of the heatmap (L) receptor L domain, [F] furin-like domain, [G] growth factor receptor domain, [K] protein kinase domain), with incomplete domains represented by segmented labels. On the right side of the heatmap, the total number of isoform-encoded proteins with each specific protein domain configuration is presented. Vertical dashed lines help to visualize relevant groups of isoform-encoded proteins, which are described at the bottom of the heatmap.

Collectively, this comprehensive characterization reveals a complex and diverse landscape of *HER2* splicing isoforms, extending even to well-studied variants like p95 and $\Delta 16$. It also provides a robust framework for generating new hypotheses about the functional roles of *HER2* splicing isoforms and their potential impact on resistance to *HER2*-targeted therapies in breast cancer.

Characterization of *HER2* isoform diversity at protein and mRNA levels

To further characterize the spectrum of *HER2* splicing isoform diversity, we analyzed their features at both the mRNA (transcript) and protein dimensions (Supplemental Fig. S3; Supplemental Tables S2, S3). Transcript length analysis revealed two significant sets of isoforms: isoforms with a primary peak length near the canonical *HER2* transcript (ENST00000269571; 4557 nt) and another minor peak around 2600 nt (Supplemental Fig. S3A). Protein length distribution showed a significant peak corresponding to the canonical *HER2* protein (1255 amino acids) and a pronounced left-skewed distribution of shorter proteins (Supplemental Fig. S3B).

To assess the degree of similarity among *HER2* isoforms, we performed pairwise sequence alignments at both the transcript and protein levels (Supplemental Fig. S3C; Supplemental Table S4). The resulting correlogram revealed clusters of isoforms closely corresponding to the 13 functional groups defined in Figure 2B. Isoforms from groups 1, 2, 3, 7, 8, and 13 exhibited high similarity (>80%) at both the nucleotide and amino acid levels, likely representing variants closely related to the canonical *HER2* sequence (ENST0000026957, group 1). In contrast, most of isoforms from groups 4, 5, 6, 9, 10, 11, and 12 displayed lower similarity (<40% in most cases) to the canonical isoform and other groups.

We also evaluated nucleotides and amino acid similarities focused on the set of $\Delta 16$ and p95 *HER2* splicing isoforms (Supplemental Fig. S4B). Although the eight p95 isoforms showed high similarities to each other at both the transcript and protein levels, three of the nine $\Delta 16$ transcripts (33.3%, including the GENCODE-reported isoform ENST00000580074.1) exhibited low similarity to the other five isoforms. Thus, these results highlight the importance of analyzing *HER2* diversity at transcript and protein levels.

HER2 isoform expression patterns across breast cancer subtypes

After our comprehensive characterization of *HER2* isoforms, we sought their expression across 561 breast cancer tumors from The Cancer Genome Atlas (TCGA) (Fig. 1B; The Cancer Genome Atlas Network 2012). To have groups of clinically and biologically similar tumors and patients, we initially stratified tumors into six specific subgroups based on their HR status (HR⁺ or HR⁻) and *HER2* expression levels (*HER2*-high, *HER2*-low, or *HER2*-zero) determined by IHC and/or FISH. The refined cohort was subclassified in 442 samples HR⁺, including 40 with *HER2*-zero, 308 with *HER2*-low, and 94 samples with *HER2*-high status, and in 119 samples ER⁻, divided into 18 *HER2*-zero, 68 *HER2*-low, and 33 *HER2*-high samples (Fig. 1B; Supplemental Table S5).

To confirm the splicing changes observed in *HER2* full-length isoforms (Fig. 2), we examined exon–exon junction read counts from TCGA breast cancer samples. This analysis reproduced the key splicing patterns initially identified, including ES, alternative splice site usage, intron retention, and mutually exclusive exons, across both known and novel *HER2* isoforms (Supplemental Table S6). All alternative exons were supported by junction reads in the TCGA data set, demonstrating that these splicing events

are present in the breast cancer transcriptome. This orthogonal validation confirms the existence of the alternative splicing events identified through long-read sequencing. Additionally, analysis of *HER2* gene expression (pooled from all isoforms) using RNA-seq data showed strong concordance with *HER2* status determined by immunohistochemistry (Fig. 3A,B). *HER2*-high tumors exhibited the highest expression levels, followed by progressively lower expression in *HER2*-low and *HER2*-zero tumors. This gradient of expression was observed in both HR⁺ (Fig. 3A) and HR⁻ (Fig. 3B) tumors, with significant differences between all *HER2* status categories (Mann–Whitney *U* test, *P*-value < 0.05).

Next, we analyzed the expression profiles of all individual *HER2* splicing isoforms across the ISO groups and both TCGA (patients) and Cancer Cell Line Encyclopedia (CCLE; cell lines) samples from distinct HR and *HER2* statuses. First, using a threshold of transcripts per million (TPM) > 0, expression evidence was observed for all isoforms (Supplemental Fig. S5). When applying more stringent thresholds (TPM \geq 1 and TPM \geq 10), a high proportion of *HER2* total isoforms (92.2% and 51.1%, respectively) (Supplemental Table S7) and novel isoforms (93.5% and 50.6%, respectively) (Supplemental Table S8) remained expressed. Comparative analysis with *TP53* and *AKAP9* revealed these genes have substantially fewer isoforms expressed at these thresholds (Supplemental Tables S7, S8), suggesting particular biological relevance of *HER2* isoform diversity in these samples. We then examined the expression patterns of *HER2* isoforms in detail (Fig. 3C (HR⁺) and 3D (HR⁻), the top 10 most-expressed isoforms; Supplemental Fig. S6; for all 90 isoforms, see Supplemental Table S9). Overall, *HER2*-high patients presented the highest expression for all isoforms, followed by *HER2*-low. *HER2*-zero samples presented the lowest expression for all isoforms. Accordingly, all breast cancer subtypes presented the same top four most-expressed isoforms, including the canonical isoform (ISO 1: ENST00000269571) (Fig. 3C,D). However, the canonical isoform was not the highest expressed within all groups. In HR⁺ samples (Fig. 3C), the canonical *HER2* isoform exhibited a gradient of expression (median: *HER2*-high, \sim 6 TPM; *HER2*-low, \sim 4 TPM, *HER2*-zero, \sim 3 TPM). ENST00000541774.5 (ISO 7) and PB.14155.831 (ISO 9) showed consistently high expression across all *HER2* categories (median: >6 TPM for all groups). PB.14155.385 (ISO 1) and ENST00000578373.5 (ISO 12) displayed a gradient similar to the canonical isoform but with lower overall expression levels. In HR⁻ samples (Fig. 3D), although overall patterns of *HER2* splicing isoforms were similar to those of HR⁺, some differences were highlighted. The median of expression of *HER2*-high versus *HER2*-low and *HER2*-zero was more pronounced for multiple splicing isoforms, including the canonical. An increased expression variability was noted in the *HER2*-high group for specific isoforms (e.g., ISO 1: PB.14155.385). Although some differences were observed, no distinct expression patterns were evident within patient groups.

Next, because UTR regions play critical roles in post-transcriptional regulation, such as mRNA stability, localization, and translation efficiency, potentially influencing gene expression without altering protein structure (Mayr 2017), we investigated whether UTR length correlated with expression levels in isoforms sharing the same protein-coding region. Our analysis identified six groups comprising 21 isoforms with identical protein sequences but differing 3'-UTR regions (Supplemental Table S10). However, we observed no significant association between UTR structural differences and expression profiles (Supplemental Fig. S7), suggesting that, in this context, UTR variation may not be a primary driver of expression regulation.

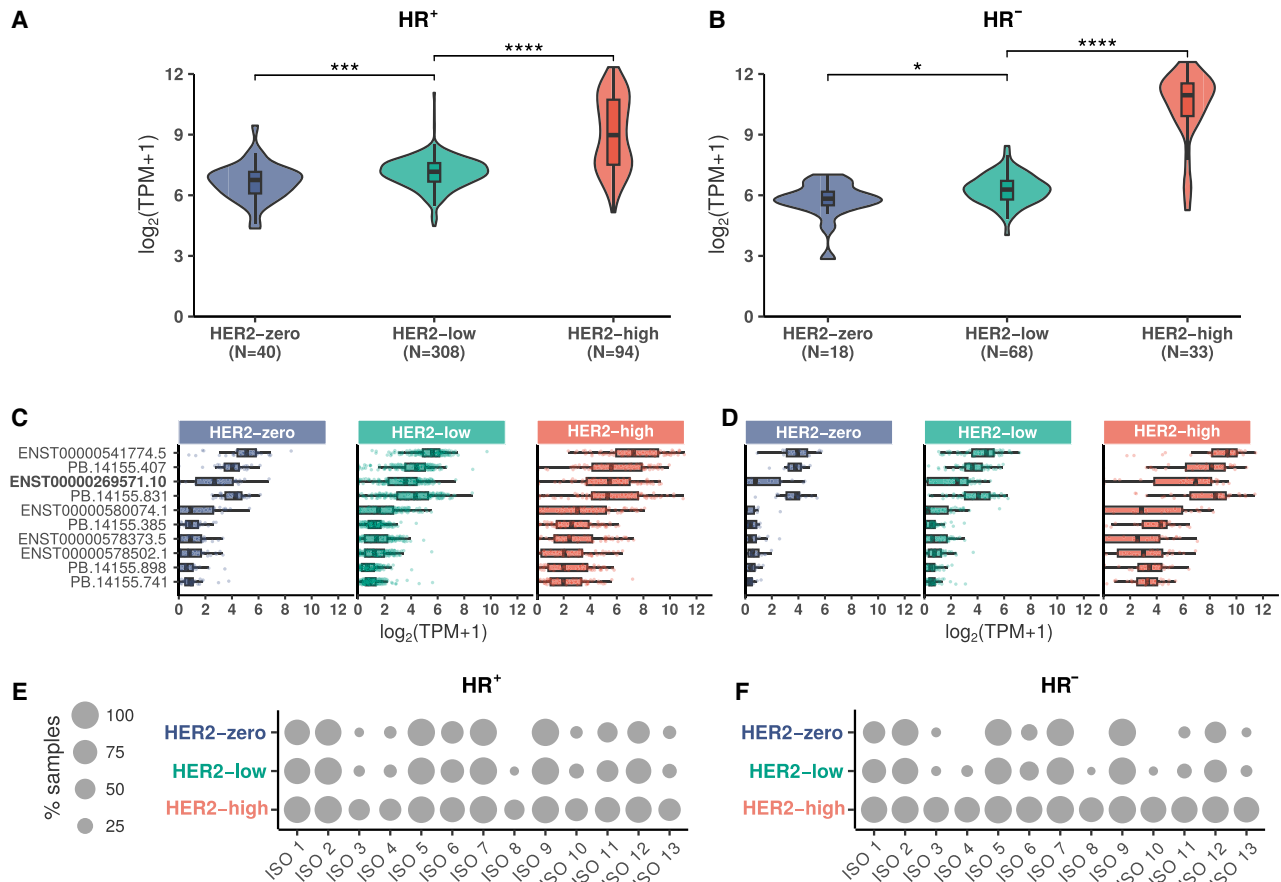


Figure 3. Expression profiles of *HER2* gene and splicing isoforms in breast cancer samples classified by immunohistochemistry status of HER2 and hormone receptor. (A,B) *HER2* gene expression levels in HR⁺ (A) and HR⁻ (B) breast cancer samples, stratified by HER2-high (red), HER2-low (green), and HER2-zero (blue) status. Mann-Whitney *U* test statistical significance: (*) *P*-value < 0.05, (***) *P*-value < 0.001, (****) *P*-value < 0.0001. (C,D) Expression profiles of the top 10 most-expressed *HER2* isoforms in HR⁺ (C) and HR⁻ (D) breast cancer samples, stratified by HER2 status. The canonical *HER2* isoform is highlighted in bold. Expression levels are shown in $\log_2(\text{TPM}+1)$. (E,F) Dot plots showing the percentage of HR⁺ (E) and HR⁻ (F) TCGA breast cancer samples expressing each *HER2* isoform group (ISO 1–13), stratified by HER2 status: zero (blue), low (green), and high (red).

Subsequently, we assessed the prevalence of (% of TCGA patient samples expressing – $\text{TPM} \geq 1$) different *HER2* isoform groups (ISO 1–13) across breast cancer subtypes, as shown in Figure 3E (HR⁺) and 3F (HR⁻). In HR⁺ samples, isoform groups ISO 1, 2, 5, 6, 7, 9, and 12 showed moderate to high prevalence across all HER2 categories (zero, low, and high). A similar overall pattern was observed in HR⁻ samples, although with some differences. HER2-high tumors showed more uniform prevalence across all isoform groups compared with HER2-low and HER2-zero categories. ISO 1–7 were more expressed in HER2-high HR⁻ samples. HER2-zero and HER2-low HR⁻ samples showed more selective isoform expression, with ISO 1, 2, 5, and 6 being the most consistently expressed groups. In both HR⁺ and HR⁻, HER2-high presented the highest number of isoforms expressed in the highest number of samples.

Given the established association between p95 expression and poor clinical outcomes, including drug resistance (Scaltriti et al. 2007; Arribas et al. 2011), we conducted a detailed investigation of isoforms encoding p95-like proteins. Specifically, we analyzed the presence of upstream open reading frames (uORFs), known regulatory elements that modulate protein expression through various mechanisms (Young and Wek 2016). Among the p95 *HER2* isoforms identified, six (PB.14155.66, PB.14155.556, PB.14155.187, PB.14155.988, PB.14155.651, and

PB.14155.309) contained both predicted ORFs and uORFs, whereas two (PB.14155.407 and PB.14155.1411) lacked uORFs (Supplemental Fig. S8A). Expression analysis revealed significant differences between isoforms with and without uORFs across all HER2 status categories (Supplemental Fig. S8B). Consistent with the typical repressive effect of uORFs on downstream protein expression (Lee et al. 2021; Jagannatha et al. 2024), isoforms lacking uORFs showed significantly higher expression levels compared with those containing uORFs (*P*-value < 0.0001). Subsequently, we extended the uORF analysis across all *HER2* isoform groups. We identified an additional 19 isoforms containing uORFs distributed across groups ISO 2, 4, 7, 9, and 11 (Supplemental Fig. S9; Supplemental Table S11). In concordance with our observations in P95-like isoforms, comparative expression analysis revealed that isoforms lacking uORFs exhibited consistently higher expression levels than those containing uORFs across nearly all isoform groups and *HER2* status categories. The sole exception to this pattern occurred in group ISO 4 within HR⁻/HER2-high patients.

These findings regarding uORF regulatory mechanism add another layer of complexity to *HER2* isoform expression and potentially influence their biological impact in breast cancer. Altogether, our results reveal a complex landscape of *HER2* isoform expression across breast cancer subtypes, diverging from total *HER2* levels.

Mass spectrometry validation and structural prediction of HER2 isoform–derived proteins

Because expression evidence was observed for all identified *HER2* splicing isoforms, we proceeded with their characterization at the protein level to strengthen their biological relevance and functionality. Mass spectrometry (MS) provided crucial validation by offering direct evidence of protein isoform expression. Using MS data from 76 breast tumors, we confirmed the presence of proteins from all isoform groups (Fig. 4A; Supplemental Table S12). Consistent with RNA-seq data, HR⁺ tumors showed proportionally more proteins confirmed by MS compared with HR⁻ tumors. Similarly, HER2-high tumors presented more validated proteins compared with HER2-low and HER2-zero tumors, with the latter showing the lowest level of validation (Fig. 4A, bottom panel). Although isoform groups 5, 6, 9, 10, 11, and 12 showed variable detection patterns, groups 1, 2, 3, 7, 8, and 13 were more frequently detected (Fig. 4A, bottom panel).

It is worth noting that peptides detected by MS may be either shared among multiple isoform groups or specific to a single group. For isoforms with highly similar sequences, peptide assignment may favor one group over others based on spectrum quality and alignment confidence. For example, the peptide CPSSGWR is unique to isoform group 4, whereas MALESILR, detected in group 9, is also present in other isoform groups. These findings confirm the presence of HER2 protein isoforms in breast tumors while acknowledging the inherent limitations of peptide-level resolution in distinguishing closely related isoforms.

Next, we investigated the three-dimensional structure of protein isoforms, a fundamental step to understanding their function, interactions, and potential as therapeutic targets. For HER2 isoforms, structural information is particularly crucial as it can reveal how alternative splicing events may alter receptor conformation, ligand binding, dimerization, and downstream signaling capabilities. However, experimental determination of protein structures, especially for multiple isoforms, is time-consuming, expensive, and often challenging. The advent of AlphaFold has revolutionized the field of protein structure prediction (Jumper et al. 2021), allowing the investigation of protein isoforms at an unprecedented scale and speed. Using AlphaFold2, we found striking diversity in the predicted protein structures of HER2 isoforms, which further corroborated our previous *in silico* predictions. The canonical HER2 isoform (ISO 1: ENST00000269571) displays the complete domain structure (transmembrane domain, full extracellular and intracellular domains, and juxtamembrane domain) with high-confidence predictions across most regions (Fig. 4B; Supplemental Fig. S10). In contrast, ISO 5 (PB14155.141) shows significant structural alterations, particularly in the extracellular region (Fig. 4C). Similarly, ISO 9 (PB14155.831) exhibits an altered structure, lacking several key domains (transmembrane and extracellular domains) and likely retaining only a partial kinase domain (Fig. 4D).

Collectively, these results add a new layer of functionality to the HER2 isoforms by confirming their translation and presence in breast cancer samples and the structural predictions indicating their distinct functional properties.

HER2 isoform clustering reveals novel subgroups within breast cancer patients

Next, we investigated the *HER2* splicing isoform usage levels (percentage spliced in [PSI] values), which measure the relative abundance (expression) of a particular splicing isoform of a gene, across 561 breast tumors. First, we examined the internal variability

in *HER2* isoform PSI values per cancer subtype. This analysis revealed significant differences within variability in *HER2* isoform expression patterns (Fig. 5A). In HR⁺ patients, HER2-high samples showed significantly lower dissimilarity compared with the HER2-low and HER2-zero samples (P -value < 0.0001; Mann–Whitney U test), indicating more consistent isoform expression patterns in HER2-high tumors. HR⁻ patients exhibited a similar trend, with HER2-high samples showing the lowest dissimilarity between HER2-low and HER2-zero (P -value < 0.05; Mann–Whitney U test) (Fig. 5A).

Given this significant internal variability in isoform expression, we subsequently explored the possibility of identifying patient clusters exhibiting similar expression patterns within each HER2 subgroup. Accordingly, unsupervised hierarchical clustering of PSI values revealed distinct subgroups within clinically defined HER2 status categories (for HR⁺ breast cancer samples, see Fig. 5B; for HR⁻ breast cancer samples, see Supplemental Fig. S11A; Supplemental Table S13). We found a complex landscape of *HER2* isoform usage that extends beyond conventional *HER2* expression levels. The clustering analysis identified three major clusters within each clinically defined tumor group for both HR⁺ and HR⁻ patients—HER2-zero (clusters Z1–Z3), HER2-low (clusters L1–L3), and HER2-high (clusters H1–H3) (for HR⁺ samples, see Fig. 5B; for HR⁻ samples, see Supplemental Fig. S11A)—revealing an intricate pattern of isoform expression that varied within tumor groups with the same HR/HER2 status, yet it was to some extent recapitulated in clusters of distinct groups. Across all patients, isoforms from groups (ISO) 1, 5, 7, and 9 showed proportionally higher expression compared with all others, but with distinct expression patterns among them. Specifically, in HR⁺ clusters Z1, L1, and H3 (Fig. 5B), as well as HR⁻ clusters H3, L3, and Z2 (Supplemental Fig. S11A), we observed high expression of ISO 9 and ISO 5. Because these isoforms lack the trastuzumab-binding domain, their predominance might predict poor response to HER2-targeted therapies. Conversely, HR⁺ clusters H1, L2, and Z3 (Fig. 5B), along with HR⁻ clusters H1, L2, and Z3 (Supplemental Fig. S11A), showed the highest expression of ISO 7 and ISO 1, which retain the trastuzumab-binding domain, suggesting potential favorable response to HER2-targeted therapies. Clusters H2, L3, and Z2 (Fig. 5B; for H2, L1, and Z1 for HR⁻, see Supplemental Fig. S11A) exhibited a gradient of expression from ISO 7 to ISO 5, ISO 9, and ISO 1, suggesting more variable therapeutic responses.

The HER2-low category exhibited the most heterogeneous isoform expression patterns. Although the subgroup L3 closely resembled the HER2-high profile, the L1 showed isoform expression patterns more similar to those of HER2-zero tumors, enriched with expression of isoforms from group 9 (out of cell membrane and without trastuzumab ligand). This heterogeneity within the HER2-low category may explain the variable responses to ADC HER2-targeted therapies (T-DXd) observed in clinical studies of this patient population. To further characterize the *HER2* expression patterns across the identified patient clusters, we analyzed both total *HER2* gene expression and the expression of specific *HER2* isoform groups. Total *HER2* gene expression levels (i.e., the sum of all isoforms), as expected, confirmed that HER2-high clusters (H1, H2, and H3) have an elevated HER2 expression (median: 8–10 $\log_2(\text{TMP} + 1)$) compared with the HER2-low (median: 6.5–7.5 $\log_2(\text{TMP} + 1)$) and HER2-zero (median: 5.5–7 $\log_2(\text{TMP} + 1)$) clusters (for HR⁺, see Fig. 5C; for HR⁻, see Supplemental Fig. S11B). No statistically significant differences were found in total *HER2* expression among clusters within each HER2 group, except for Z3 (HR⁺).

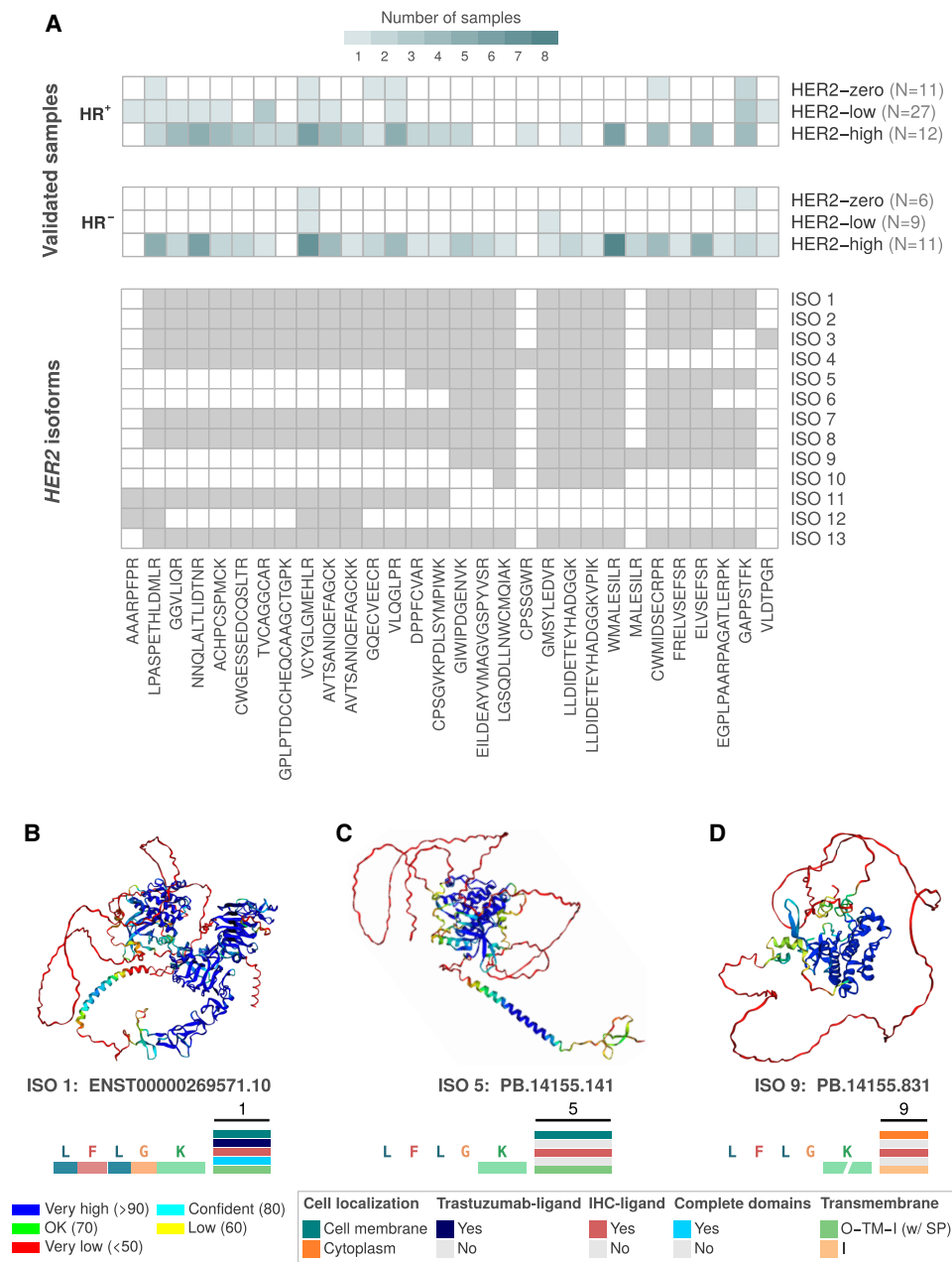


Figure 4. Mass spectrometry (MS) validation of HER2 isoform-derived proteins and their predicted 3D structures. (A) The top panels display the number of samples with MS confirmation of HER2 proteins, stratified by HER2 status (HER2-high, HER2-low, HER2-zero) and hormone receptor status (HR⁺ and HR⁻). Isoform-derived proteins are grouped from 1 to 13, and the validated peptides for each group are indicated in the lower panel. (B–D) AlphaFold2-predicted protein structures for HER2 isoforms. (B) The canonical isoform (ISO 1: ENST00000269571) with well-defined domain regions. (C) ISO 5 (PB.14155.141), a variant from group 5 with specific domain alterations affecting its cellular localization and trastuzumab-binding potential. (D) ISO 9 (PB.14155.831), an isoform with unique structural characteristics lacking complete domains, potentially affecting functional properties. Color codes in each structure represent pLDDT confidence scores for structural predictions and the corresponding HER2 protein domains, as the bottom legend indicates. Transmembrane: (O) outside, (TM) transmembrane region, (I) inside, (SP) signal peptide.

We next focused on the expression of *HER2* isoforms encoding proteins localized to the cell membrane and containing the trastuzumab-binding domain (ISO 1–4 groups), given their potential clinical relevance for trastuzumab-based therapies (for HR⁺, see Fig. 5D; for HR⁻, see Supplemental Fig. S11C). The H2 and H1 clusters showed the highest expression of these clinically relevant isoforms. The H3 cluster (~26% of HER2-high patients), despite being classified as HER2-high, exhibited lower expression of these iso-

forms (ISO 1–4), more closely resembling the levels seen in some HER2-low clusters. This result highlights the heterogeneity even within the HER2-high category and may suggest why HER2-high tumors have a low or no responsiveness to trastuzumab-based treatments. Among the HER2-low, L2 (46% of HER2-low patients) and L3 (30% of HER2-low patients) show a moderated expression of ISO 1–4 isoforms, with levels between H1 and H3 clusters, which may support the fact that some HER2-low tumors are

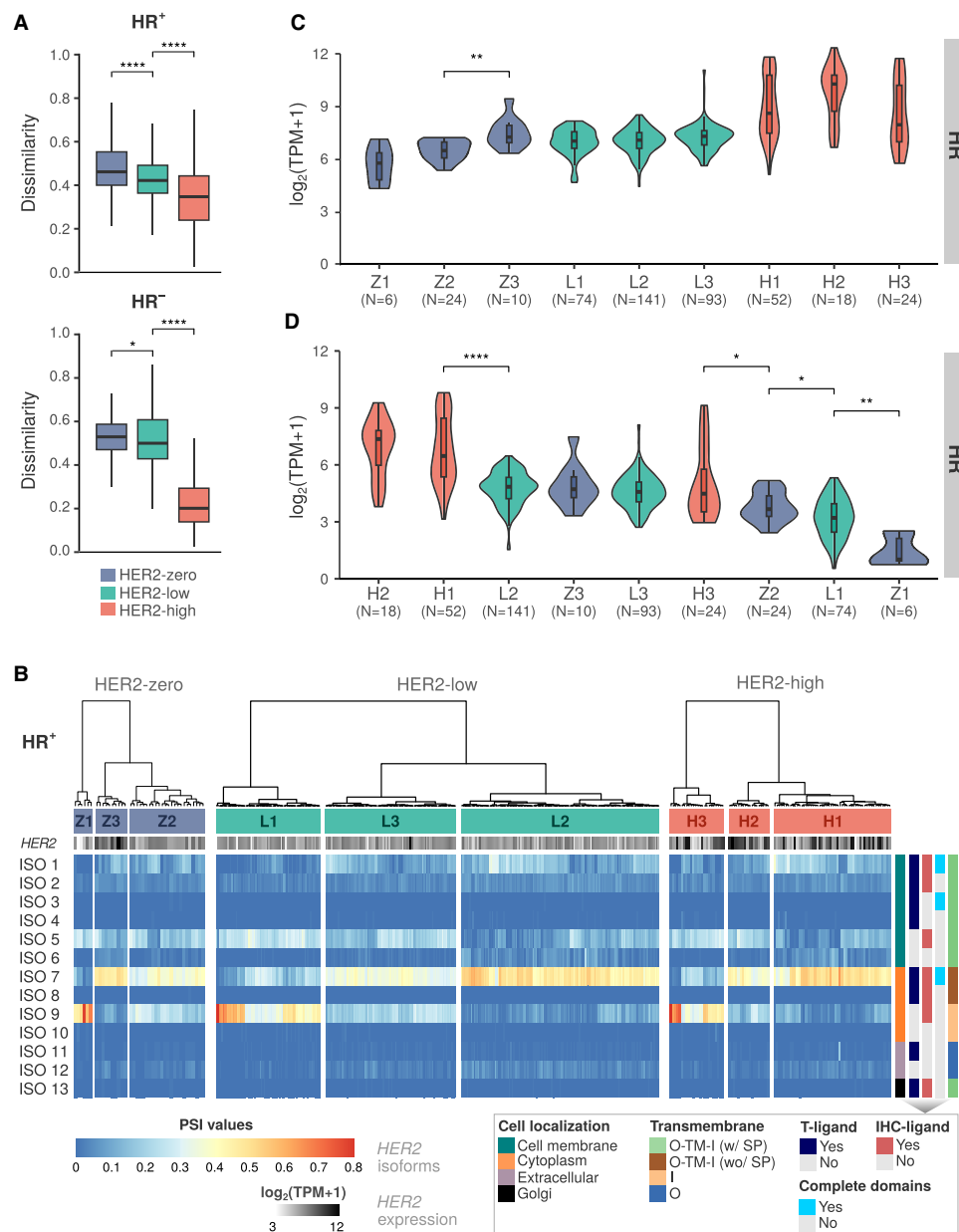


Figure 5. *HER2* isoform expression patterns and clustering analysis in HR^- and HR^+ breast cancer patients stratified by *HER2* status. (A) Jaccard index comparing the dissimilarity of *HER2* isoform expression between *HER2*-zero (blue), *HER2*-low (green), and *HER2*-high (red) breast cancer samples in HR^+ and HR^- patients. Mann–Whitney *U* test significant differences in dissimilarity are noted between *HER2* status categories: (*) *P*-value < 0.05, (****) *P*-value < 0.0001. (B) Heatmap of *HER2* isoform percent spliced in (PSI) levels in HR^+ patients, clustered by expression similarity. Z1, Z2, and Z3 are for *HER2*-zero; L1, L2, and L3 are for *HER2*-low; H1, H2, and H3 are for *HER2*-high. Isoform features (cell localization, antibody-binding sites, domain completeness, and transmembrane topology: [O] outside, [TM] transmembrane region, [I] inside, [SP] signal peptide) are indicated on the right. (C) *HER2* gene expression levels among HR^+ patient clusters. (D) Expression levels of *HER2* isoforms whose encoded proteins are located in the cell membrane and contain the trastuzumab-ligand (groups ISO 1–4) among HR^+ patient clusters. Mann–Whitney *U* test statistical significance: (*) *P*-value < 0.05, (**) *P*-value < 0.01, (****) *P*-value < 0.0001. Comparisons without statistical significance are not depicted in the figure.

responsive to *HER2*-targeted ADCs, as T-DXd. In contrast, group L1 (~24% of *HER2*-low patients) has low expression of ISO 1–4 and high expression of isoforms from groups 5 and 9 (which lack the trastuzumab-binding domain), suggesting a potential group of lower/no response to T-DXd or other ADCs targeting *HER2*. The same pattern was observed for HR^- samples (Supplemental Fig. S11C). Finally, we also identified the most abundantly expressed isoforms from each isoform group (ISO 1–13) in both the HR^+

and HR^- samples (Supplemental Figs. S12, S13, respectively). Overall, HR^+ and HR^- samples showed a similar expression pattern in terms of the most abundantly expressed isoforms per group.

Although alternative splicing is a major source of transcript diversity, alternative promoter usage also contributes by generating isoforms with distinct 5'-UTR regions that may influence mRNA translation efficiency and the coding region start site. Therefore, analyzing promoter activity alongside alternative

splicing provides insight into how transcriptional and post-transcriptional mechanisms shape the complex HER2 isoform landscape in breast cancer. By performing this analysis, we identified 22 potential promoter regions for the HER2 gene, designated P1 to P22 (Supplemental Fig. S14A; Supplemental Table S14). When mapped against our functionally defined isoform groups (ISO 1–13), no clear association emerged between promoter usage and functional properties. Isoforms within the same group were often regulated by different promoters, whereas individual promoters, particularly P4, regulated isoforms across multiple groups (ISO 1–5, 8, 11–13). To validate these predictions, we integrated CAGE data from the FANTOM project, which map transcription start sites (TSSs) through detection of 5'-capped RNA, and H3K4me3 ChIP-seq data from ENCODE, which mark active promoter chromatin states (Supplemental Fig. S14B). A substantial fraction of predicted promoters overlapped with CAGE peaks and/or H3K4me3 signals, confirming their functional relevance. Analysis of promoter activity across breast cancer patient clusters previously stratified by HER2 isoform expression (Supplemental Fig. S15; Supplemental Table S15) revealed that the promoter P4 consistently showed the highest relative activity across all HR/HER2 subgroups. However, we observed no significant differences in promoter usage patterns between patient clusters, indicating that promoter regulation remains conserved across breast cancer subtypes.

To expand our understanding of *HER2* isoform expression beyond breast cancer, we analyzed normal breast tissue and three gynecologic tumor types—cervical (CESC), ovarian (OV), and endometrial (UCEC)—that are clinically relevant in the context of HER2-targeted therapies. Using RNA-seq data from TCGA and focusing on samples with high *HER2* expression, we observed widespread expression of *HER2* isoforms across all tumor types, with CESC, OV, and UCEC tumors expressing a broader diversity of isoforms. Unsupervised clustering of PSI values revealed distinct patterns of isoform usage within each tumor type, including prominent expression of isoforms from groups ISO 1, 5, 7, and 9, mirroring some of the patterns observed in breast cancer. Gynecologic tumors displayed considerable heterogeneity in *HER2* isoform usage, suggesting that similar isoform-driven mechanisms of resistance or responsiveness to HER2-targeted therapies may extend beyond breast tumors (Supplemental Fig. S16).

Altogether, this detailed isoform-level analysis reveals a complex picture of *HER2* expression in breast (and other) cancers, identifying subgroups with distinct isoform utilization patterns that transcend traditional HER2 status classifications. These findings hold significant potential for refining patient stratification and improving response predictions to HER2-targeted therapies.

HER2 isoform profiles concerning ADC sensitivity

Next, we investigated the relationship between *HER2* isoform expression patterns in breast cancer cell lines and their response to T-DM1 or T-DXd (Fig. 6A; Supplemental Table S16). Among HR⁺/HER2⁺ cell lines, BT-474, EFM-192A, MDA-MB-361, and ZR-75-30 showed sensitivity to T-DM1, whereas UACC-812 was resistant. In HR⁻/HER2⁺ cells, AU565, HCC1954, MDA-MB-453, and SK-BR-3 demonstrated sensitivity to T-DM1, whereas JIMT-1 and UACC-893 were resistant. MDA-MB-453 and SK-BR-3 showed sensitivity to both T-DM1 and T-DXd, whereas BT-474 was resistant only to T-DXd. The BT-474 cell line contains a *SLX4* mutation (c.1181G>C, p.R394T), previously identified as a potential mechanism of resistance to T-DXd, because *SLX4* encodes a DNA repair

protein that regulates structure-specific endonucleases and seems to play a role in resistance to TOP1 inhibition. Among HER2- cell lines, MCF-7 and ZR-75-1 (HR⁺) and MDA-MB-231 (HR⁻) showed resistance to both ADCs, consistent with their low *HER2* expression levels. These results indicate a complex relationship between *HER2* (isoform) expression and ADC response. Although most HER2⁺ cell lines expressing high levels of HER2 and both isoform groups (ISO 1–4 and ISO 5–13) showed sensitivity to ADCs, cell lines expressing lower levels of the isoforms from groups 1–4 (UACC-812, ZR-75-1, MDA-MB-231, MDA-MB-468, JIMT-1, BT-474), which have the trastuzumab-binding domains and are predicted to be located in cell membrane, are resistant to T-DM1 or T-DXd (Fig. 6A).

Based on the previous results, we decided to better investigate the relationship between different HER2 isoform groups and ADC response. We evaluated the expression of ISO 5–13 (isoforms lacking trastuzumab-binding domain and/or located outside the cellular membrane) against ISO 1–4 (isoforms with intact trastuzumab-binding domain) across breast cancer cell lines (Fig. 6B). Because the diagonal dashed line represents an equal expression of both isoform groups and the circle size indicates the levels of *HER2* gene expression, most ADC-responsive cell lines (labeled in green, such as HCC1954, AU565, and BT-474) clustered in the high-expression region and maintained a balanced ratio between ISO 1–4 and ISO 5–13 expression. UACC-812, despite showing high overall *HER2* expression, demonstrated resistance to ADCs and exhibited higher expression of ISO 5–13 relative to ISO 1–4 (positioning it above the diagonal). Similarly, JIMT-1, ZR-75-1, MDA-MB-231, and MDA-MB-468 showed resistance to T-DM1 and/or T-DXd and a higher expression of ISO 5–13 relative to ISO 1–4. On the other hand, MCF-7 and UACC-893 showed ADC resistance regardless of their ISO 1–4/ISO 5–13 ratio. We have several different cell lines under and above the diagonal (each circle represents a cell line). Still, for most of them, we have no ADC treatment available (circles without names). To gain further insight into the expression profile of *HER2* and its isoforms in these additional cell lines, without ADC treatment information, we created Figure 6C. HR⁺/HER2⁻ cell lines showed consistently low-to-moderate expression of both isoform groups, with relatively balanced ratios between ISO 1–4 and ISO 5–13. HR⁻/HER2⁺ cell lines exhibited the highest total levels of *HER2* expression and maintained substantial expression of both isoform groups. The largest group, HR⁻/HER2⁻ cell lines, demonstrated consistently low expression of both isoform groups, albeit with some variability in the relative ratios of ISO 1–4 and ISO 5–13. Finally, we conducted an analysis examining expression patterns across all HER2 isoform groups in cell lines responsive (R) and nonresponsive (NR) to ADCs (Supplemental Fig. S17). We found significantly increased expression of isoform groups ISO 1, 6, and 12 in responsive cell lines. Although ISO 1 and 6 encode membrane-localized proteins, only ISO 1 retains the trastuzumab-binding domain, providing a clear mechanistic rationale for ADC efficacy. Group ISO 12, comprising extracellular isoform proteins without direct involvement in HER2 recognition, also showed elevated expression in responsive cell lines, although the functional implications remain to be determined.

Altogether, these results indicated that high HER2 expression and a balanced, although slightly variable, expression of ISO 1–4 and ISO 5–13 in most cell lines appear necessary for effective cellular responses to HER2-targeted therapies. In contrast, disruptions in isoform expression, as observed in some drug-resistant lines (e.g., UACC-812, which shifted toward ISO 5–13), may contribute to resistance against HER2-targeted therapies. However, other

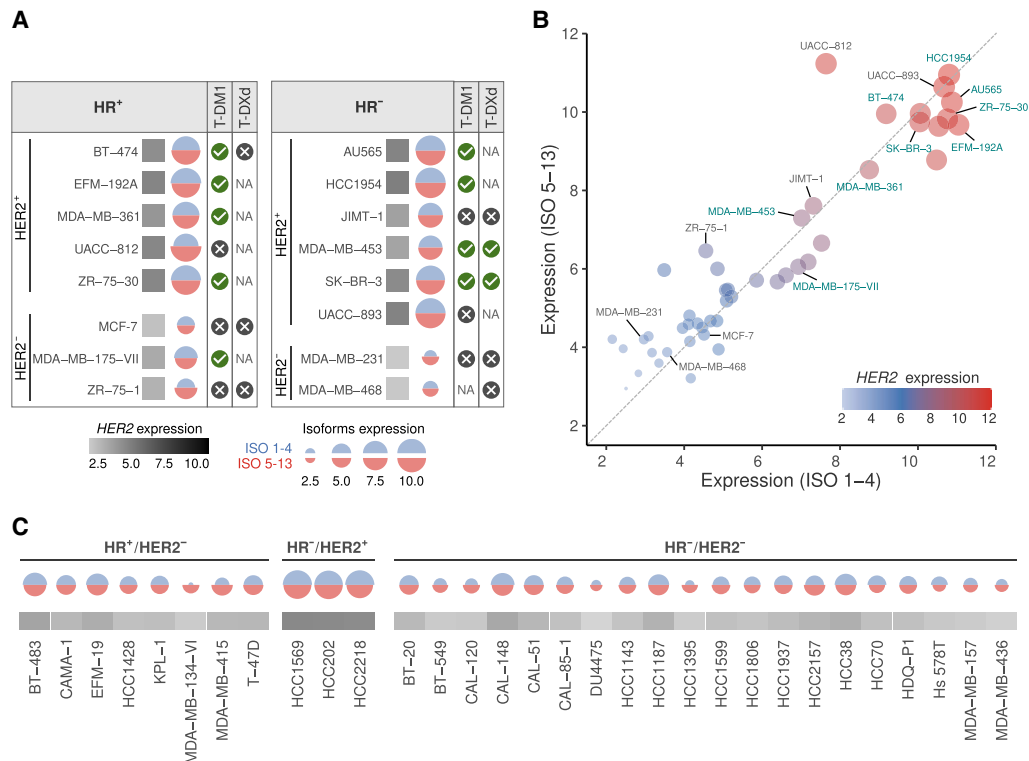


Figure 6. *HER2* splicing isoform profiles in breast cancer cell lines and their response to T-DM1 and T-DXd. (A) *HER2* isoform expression in HR⁺ (left) and HR⁻ (right) breast cancer cell lines treated with T-DXd or T-DM1. Cell lines responsive to treatment are marked with a checkmark. Unresponsive cell lines are marked with a cross. (NA) Drug treatment is not available. The square block's color represents the total *HER2* expression level ($\log_2(\text{TPM} + 1)$). Blue semi-circular plots indicate the expression of isoforms with intact trastuzumab-binding domain (ISO 1–4). Red semi-circular plots represent the expression of isoforms lacking trastuzumab-binding domain and/or cell membrane localization (ISO 5–13). (B) Scatter plot showing the relationship between expression levels of ISO 1–4 and ISO 5–13 groups across breast cancer cell lines. Dot size and color intensity correspond to total *HER2* expression level ($\log_2(\text{TPM} + 1)$). Cell lines named in green indicate those responsive to ADCs (T-DM1 or T-DXd). Cell lines named in gray indicate those ADC-resistant. (C) *HER2* isoform expression patterns across breast cancer cell lines (without ADC treatment) stratified by HR/HER2 status. Semicircular plots represent expression levels of isoforms from groups 1–4 (blue) and 5–13 (red). The gray squares below indicate total *HER2* expression levels.

mechanisms (e.g., the SLX4 mutation in BT-474 and additional isoform-related factors) can confer ADC resistance independent of isoform expression patterns.

Dynamic changes in *HER2* isoform expression associated with acquired resistance to trastuzumab and T-DM1

Finally, to illuminate the molecular mechanisms underlying acquired resistance to *HER2*-targeted therapies, we investigated the *HER2* gene and isoform expression profiles in breast cancer cell lines before and after developing resistance to trastuzumab and T-DM1 (Fig. 7; Supplemental Table S17).

We first examined the overall *HER2* expression in SK-BR-3 and BT-474 cell lines in sensitive and resistant states to trastuzumab (Fig. 7A). Both cell lines in the two states maintained significantly high levels of *HER2* expression (Fig. 7A), indicating that resistance to trastuzumab is not primarily mediated by a global downregulation of *HER2* expression (Vernieri et al. 2019). In fact, because trastuzumab targets *HER2* function, we hypothesize that its upregulation may serve as a compensatory mechanism to offset its own inhibition by treatment, thus sustaining downstream signaling pathways that promote rapid tumor cell growth and proliferation.

Next, to gain deeper insights into potential resistance mechanisms, we analyzed the *HER2* isoform expression profiles and fold

change of each isoform group in trastuzumab-resistant versus trastuzumab-sensitive cells for both the SK-BR-3 and BT-474 lines (Fig. 7B; Supplemental Fig. S18). In SK-BR-3 cells, we observed significant upregulation of multiple isoform sets. In contrast, isoforms from group 8 showed downregulation (\log_2 fold change ~ 0.4) in resistant conditions (Fig. 7B; Supplemental Fig. S16A). Overall, BT-474 cells also exhibited upregulation of sets of isoforms, except for isoforms from groups 3, 5, and 8 (Supplemental Fig. S16B).

Next, we investigated the impact of acquired resistance to T-DM1 on *HER2* expression in SK-BR-3 and BT-474 cells (Fig. 7C; Supplemental Fig. S19A). Unlike trastuzumab-resistant cells (Fig. 7A), both T-DM1-resistant SK-BR-3 and BT-474 cells showed an important decrease in overall *HER2* expression compared with sensitive cells, from 11.3 to 10.6 and 10.9 to 10.1 $\log_2(\text{TPM} + 1)$, respectively (Fig. 7C; Supplemental Fig. S19A).

Lastly, we performed a detailed analysis of *HER2* isoform expression levels in T-DM1-resistant and T-DM1-sensitive SK-BR-3 and BT-474 cells. We observed complex changes in the isoform landscape for SK-BR-3 (Fig. 7D) and BT-474 (Supplemental Fig. S19B). First, the fold change (T-DM1-resistant/sensitive) analysis showed that most splicing isoforms are significantly downregulated, including isoforms from sets 1–4 (isoforms with intact domains) (Fig. 7D). Splicing isoforms from sets 5 and 9 are upregulated in T-DM1-resistant cells (Fig. 7D). In BT-474, we observed a significant downregulation of all splicing isoforms

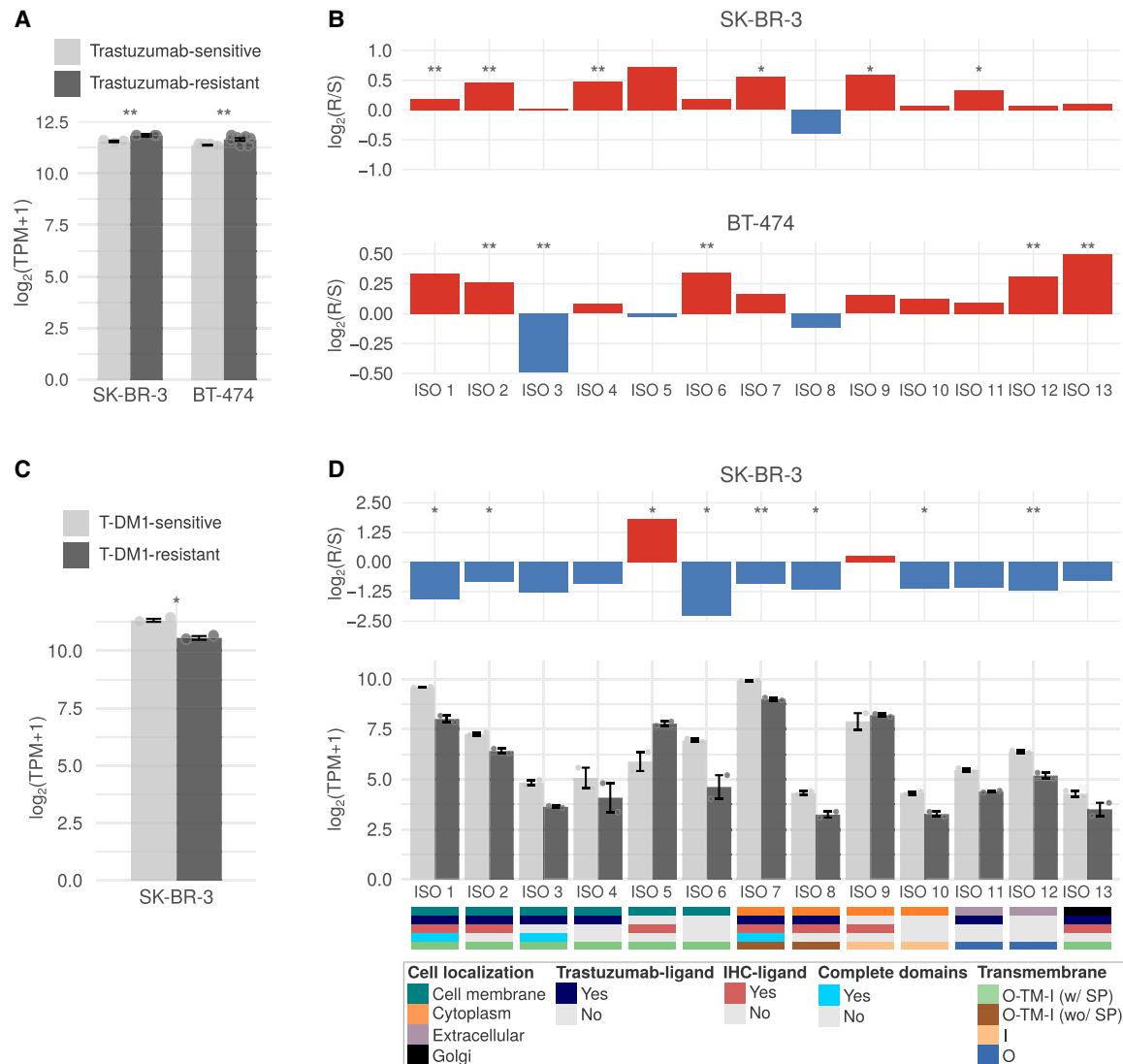


Figure 7. *HER2* gene and isoform expression profiles in breast cancer cell lines before and after acquiring resistance to HER2-targeted therapies. (A) Overall *HER2* expression in trastuzumab-sensitive (S) and trastuzumab-resistant (R) SK-BR-3 and BT-474 cells. (B) \log_2 fold change of *HER2* isoform expression in trastuzumab-resistant versus sensitive SK-BR-3 and BT-474 cells. (C) Overall *HER2* expression in T-DM1-sensitive and T-DM1-resistant SK-BR-3 cells. (D) *HER2* isoform expression levels and characteristics in T-DM1-sensitive and T-DM1-resistant SK-BR-3 cells, with \log_2 fold change shown above. Isoforms are categorized based on their cell localization, trastuzumab-binding ligand presence (T-ligand), structural completeness and transmembrane topology as indicated by the bottom color legend: (O) outside, (TM) transmembrane region, (I) inside, (SP) signal peptide. Mann–Whitney *U* test statistical significance: (*) *P*-value < 0.05, (**) *P*-value < 0.01. Comparisons without statistical significance are not depicted in the figure.

(Supplemental Fig. S19B). Altogether, these findings suggest that SK-BR-3 has adapted to T-DM1 treatment pressures by altering the balance of *HER2* isoforms to downregulate the drug target isoforms (groups ISO 1–4) and upregulating the prosurvival signaling (isoforms from groups ISO 5 and 9, which lack the trastuzumab-binding site and seem to retain the signaling capabilities through the tyrosine kinase domain), a putative mechanism to evade this ADC's effects.

Discussion

In this investigation, we uncover a complex landscape of *HER2* splicing isoform diversity in breast cancer that goes well beyond the conventional understanding of *HER2* biology (Arteaga and

Engelman 2014) and its role in targeted therapies (Modi et al. 2020; Tarantino et al. 2020). The full characterization of the set of 90 *HER2* coding isoforms, including 77 novel variants, significantly expands our knowledge of *HER2* expression and variations and sheds light on the role of *HER2* splicing isoforms in antibody-conjugated targeted therapy resistance.

First, our strategy emphasizes the importance of identifying alternative splicing isoforms through the use of full-length transcripts obtained via long-read sequencing, an approach that has been shown to provide superior accuracy and sensitivity in genomic analysis in breast cancer (Aganezov et al. 2020). This approach has allowed us to achieve a comprehensive characterization at the isoform level, rather than solely focusing on the splicing events themselves. This reveals the composition of all exons and the

open reading frame (and subsequent protein) encoded by each isoform. The breadth of our strategy becomes clear when we look at the two most studied *HER2* isoforms, $\Delta 16$ and p95: (1) we identified eight distinct isoforms with different splicing events that encode proteins lacking the extracellular domain and have an approximate molecular weight of 95 kDa, and (2) for $\Delta 16$, we found nine isoforms, all containing the exon 16 skipping ($\Delta 16$'s hallmark) and exhibiting other alternative splicing events. Understanding not just the event (e.g., ES), but the full set of isoforms that contains such events and others certainly gives us a more complete understanding of the importance and functionality of each alternative splicing isoform.

The structural and functional diversity revealed in the *HER2* isoforms, including alterations in the *HER2* protein domains and cell localization, as well as the presence or lack of the trastuzumab-binding sites, provide new insights into the heterogeneity of response in targeted therapy using antibodies and ADCs. These splicing isoform diversity profiles may explain, in part, the complex mechanisms of resistance to *HER2*-targeted therapies observed in clinical practice (Nahta and Esteva 2006; Luque-Cabal et al. 2016).

We observed variability in the expression of *HER2* splicing isoforms across the intrinsic subtypes of breast cancer, with the *HER2*-high group displaying the most uniform expression profile. This is the same patient group that shows the most consistent and profound response to anti-*HER2* therapies, whether in early-stage or metastatic disease (Verma et al. 2012). This correlation suggests that the homogeneity in *HER2* isoform expression may play a role in influencing therapeutic response.

Identifying *HER2* splicing isoforms lacking the antibody (drug) binding domains but retaining the signaling capabilities (tyrosine kinase domain) may explain a potential resistance mechanism to antibody-based therapies like trastuzumab and ADCs. This aligns with previous studies on p95, which lacks the trastuzumab-binding site and has been associated with poor prognosis and resistance in antibody-based therapy treatment (Scaltriti et al. 2007). Therefore, the expanded repertoire of *HER2* isoforms presented here suggests that alternative splicing may be a more prevalent and leading mechanism used by cancer cells in acquiring resistance and progression, especially in gene-targeted therapies.

The dynamic changes in *HER2* isoform expression observed in cell lines acquiring resistance to trastuzumab or T-DM1 highlight the adaptive nature of cancer cells. Specifically, our findings indicate that the SK-BR3 cancer cell line adapted to T-DM1 treatment pressures by altering the balance of *HER2* isoforms, downregulating those containing the drug target epitopes (isoforms in sets 1–4) and upregulating the *HER2* splicing isoforms (sets 5 and 9) lacking the antibody (trastuzumab) binding site domains. This shift was not observed under trastuzumab treatment, suggesting a mechanism of acquired resistance specific to ADC therapy. Broadly, these findings open new avenues for understanding and potentially mitigating therapy resistance through various strategies, including the development of *HER2* isoform-specific inhibitors, combination approaches targeting multiple isoform-encoded proteins, and the use of tyrosine kinase inhibitors such as lapatinib that target the kinase domain.

Furthermore, our observations indicate that among tumor subtypes (*HER2*-high, *HER2*-low, and *HER2*-zero), there are distinct subgroups of tumors expressing different splicing isoforms, some of which encode protein isoforms lacking the binding domain for antibodies used in immunohistochemistry. This may

explain why a percentage of *HER2*-low or even *HER2*-zero patients (as determined by immunohistochemistry) respond to treatment. It is reasonable to hypothesize that specific isoform compositions may create a false classification of *HER2* status and vulnerabilities to ADC treatment, even in contexts with low *HER2* expression (by immunohistochemistry) (Tarantino et al. 2020).

Although our study provides comprehensive insights into *HER2* isoform diversity and its implications for targeted therapy, several limitations should be acknowledged. First, our cell line-based resistance models, although informative, may not fully recapitulate the complexity of resistance mechanisms in breast cancer patients, in which tumor heterogeneity and microenvironment factors play crucial and yet incompletely understood roles (Roma-Rodrigues et al. 2017; Vander Velde et al. 2020). Second, although we validated the existence of *HER2* isoforms through RNA-seq expression in a large patient cohort and MS confirmation, functional validation of individual isoforms' biological roles and their specific contributions to drug resistance mechanisms requires further investigation. Third, although long-read sequencing enabled comprehensive isoform identification, technical limitations in detecting low-abundance transcripts might have led to underestimation of rare isoforms (Uapinyoying et al. 2020). Fourth, our study focused primarily on the role of *HER2* isoforms in antibody-based therapy resistance, and their potential impact on other treatment modalities, such as tyrosine kinase inhibitors, needs to be fully explored. Finally, although our findings suggest the importance of isoform-specific testing in clinical settings, the development and validation of practical diagnostic tools for *HER2* splicing isoform profiling will require additional technical and clinical validation studies (Wang and Aifantis 2020).

Despite these limitations, our findings suggest that incorporating *HER2* isoform profiling into clinical assessment could significantly enhance prediction of response to trastuzumab and ADCs. Future prospective studies with larger *HER2*⁺ patient cohorts are needed to validate whether patients expressing isoforms lacking the trastuzumab-binding domain experience poorer clinical outcomes, which could guide more personalized therapeutic approaches. Additionally, mechanistic studies should investigate the direct functional impact of specific *HER2* isoforms on trastuzumab response through ectopic expression in *HER2*-negative cell backgrounds. Such controlled experiments would eliminate potential confounding factors from diverse genetic backgrounds and provide direct evidence for how specific structural variations in *HER2* isoforms influence therapeutic response. This experimental approach would be particularly valuable for validating the clinical relevance of isoforms lacking the trastuzumab-binding domain and could inform more precise patient selection strategies for *HER2*-targeted therapies.

In conclusion, our comprehensive investigation into the diversity of *HER2* isoforms uncovers a complex landscape that may have significant implications for breast cancer biology and treatment approaches utilizing ADCs. Our findings indicate that integrating *HER2* isoform profiling into clinical practice, despite its current limited implementation in many centers, may greatly improve patient stratification and treatment selection, potentially leading to more effective targeted therapies. This research establishes a solid foundation for a more refined approach to *HER2*-positive breast cancer. We propose that optimal ADC treatment strategies should be tailored not only to *HER2* expression levels but also to the specific isoform profiles present in each tumor.

Methods

Public short-read RNA-seq data

We obtained unprocessed RNA-seq data from 561 primary tumors sourced from female breast cancer patients, publicly accessible via the TCGA repository (<https://portal.gdc.cancer.gov>). Additionally, we obtained clinical data detailing IHC staining results for HER2, estrogen receptor (ESR1 [also known as ER]), and progesterone receptor (PR), as well as FISH data for HER2. In addition, RNA-seq data from 50 breast cancer cell lines were acquired from the CCLE (<https://sites.broadinstitute.org/ccle/>), and data on cell line sensitivity to T-DXd and T-DM1 were obtained from previous studies (Supplemental Table S16). RNA-seq data from trastuzumab-sensitive and trastuzumab-resistant SK-BR-3 and BT-474 cell lines were obtained from Duan et al. (2024) and Mukund et al. (2024), respectively. RNA-seq data from T-DM1-sensitive and T-DM1-resistant SK-BR-3 and BT-474 cell lines were obtained from Gedik et al. (2024).

Public long-read RNA-seq data

We obtained processed long-read RNA-seq data from 26 tumor and four normal breast samples from Veiga et al. (2022) to create an expanded catalog of *HER2* isoforms. Briefly, in work by Veiga et al. (2022), full-length cDNA was synthesized from poly(A)⁺ transcripts, PCR-amplified, and sequenced on Pacific Biosciences (PacBio) RSII or Sequel platforms to capture full-length transcripts. Reads were processed with the ToFU pipeline (Gordon et al. 2015) to generate error-corrected consensus transcripts, which were then aligned to the human genome (GRCh38/hg38) using GMAP (Wu and Watanabe 2005). SQANTI (Tardaguila et al. 2018) was used for transcript annotation and quality control, filtering out artifacts based on criteria such as indel correction, 3'-end validation, non-canonical splice site detection, and no splice junction support. All steps were performed to ensure a high-confidence catalog of novel full-length transcripts, minimizing technical artifacts while capturing transcript diversity in breast cancer samples. Based on this catalog of novel full-length transcripts, we created an expanded version of the human reference transcriptome including 13 known protein-coding *HER2* isoforms from GENCODE (version 36; <https://www.encodegenes.org/human/>) along with 77 novel full-length protein-coding *HER2* isoforms identified in their previous study.

Quantification of isoform and gene expression levels

Using the kallisto tool (version 0.48.0; default parameters with option --bootstrap-samples 100) (Bray et al. 2016), we pseudoaligned the RNA-seq short reads from all patients and cell lines to the expanded version of the human reference transcriptome created using GENCODE and long-read RNA-seq data of breast cancer samples from Veiga et al. (2022). Next, isoform expression levels normalized in TPM were submitted to SUPPA2 (version 2.3; default parameters) (Trincado et al. 2018), which quantifies PSI values, indicating the proportion of expression that each isoform of a gene corresponds to. Gene-level expression profiles were also obtained using the tximport R package (version 1.26.1) (Soneson et al. 2015).

Characterization of *HER2* splicing isoforms

To characterize the *HER2* isoforms in terms of alternative splicing local events, coding potential, functional domains, transmembrane topology, and subcellular localization, we used several strategies. First, coding sequences (ORFs) from 77 novel *HER2* isoforms previously determined (Veiga et al. 2022) using Transdecoder

(Haas et al. 2013) were extracted, and coding sequences from 13 known *HER2* isoforms were directly retrieved from GENCODE (version 36). Local alternative splicing events were identified using the “generateEvents” function in SUPPA2 (Trincado et al. 2018) with default parameters; for intron retention events, the additional parameters --boundary V and --threshold 10 were applied. Multiple ES events, which are not reported by SUPPA2, were manually extracted.

To confirm the presence of alternative splicing events identified with SUPPA2, we additionally employed rMATS turbo (version 4.3.0) (Wang et al. 2024) with default parameters, a statistical tool that detects alternative splicing events from RNA-seq data by examining exon–exon junction read counts. For this analysis, we used RNA-seq data from 561 TCGA breast cancer samples stratified by HR/HER2 status. We extracted read counts supporting both inclusion and exclusion of alternative splicing events using rMATS. Events were considered confirmed when the mean number of both the inclusion and exclusion junction counts exceeded zero in at least one HR/HER2 group.

Protein domains from all *HER2* coding sequences were predicted using the hmmsearch tool from HMMER (version 3.3.1; default parameters) (Potter et al. 2018) based on the Pfam database (Mistry et al. 2021). Predictions of transmembrane topology were performed using the DeepTMHMM web tool (default parameters) (Hallgren et al. 2022), which uses a deep learning algorithm to predict the topology of alpha-helical and beta barrels. Protein subcellular localizations were determined based on the DeepLoc 2.0 tool (default parameters) (Ødum et al. 2024). To evaluate the presence of the IHC-binding region in *HER2* isoforms, we considered three IHC antibodies: PATHWAY HER2, Herceptest, and Oracle HER (Cho et al. 2003). In addition, similarities among *HER2* proteins were evaluated through pairwise protein alignments using the needle global aligner (https://www.ebi.ac.uk/jdispatcher/psa/emboss_needle).

Validation of *HER2* isoforms at the protein level

Validation of *HER2* isoforms at the protein level was performed using MS/MS data from 76 breast cancer patients in the CPTAC TCGA data set (study ID: PDC000173). We employed the PepQuery tool (version 2.0.2) (Wen and Zhang 2023) with default parameters, using as reference 103,069 proteins from GENCODE release 36. Each *HER2* protein derived from expressed splicing isoforms was queried in the MS/MS data to identify supporting peptide spectrum matches (PSMs). Following in silico trypsin digestion, PepQuery attempted to validate the resulting peptides against the MS data. After multiple filtering steps, only PSMs passing all criteria with an FDR < 0.05 were considered confident. Validated peptides were then assigned to their respective isoform groups (ISO 1–13) rather than being restricted to unique peptides for each protein. Given the high sequence similarity among *HER2* proteins, this approach allowed for a more comprehensive and reliable assessment of isoform detection.

Prediction of 3D protein structures of *HER2* isoforms

The prediction of the *HER2* isoforms' 3D protein structures was made with AlphaFold2 (Jumper et al. 2021) through the free and publicly available Google collaborator ColabFold (version 1.5.5) platform (Mirdita et al. 2022). We opted to run the predictions this way owing to its speed by combining it with a fast homology search with MMseqs2 (Steinegger and Söding 2017) and HHsearch (Steinegger et al. 2019), as well as the usage of the highly accurate PDB100 (Varadi et al. 2024) as its database. All analyses were run in

a “high-RAM (system: 51 GB; GPU: 15 GB) T4 GPU” machine with Python 3 and >200 GB of disk space.

All parameters were left default, except “num_recycles = 24” because membrane proteins require a higher number of recycles for better results. Several outputs are made available, including not only the predicted protein structure itself but also alignments for reference, PDB files per ranked model for editions, and other plots to support the results. The quality of the predictions was assessed by analyzing two metrics: (1) the multiple sequence alignment (MSA) coverage outputs, in which at least 30, ideally 100, sequences per position are ideal for better performance, and (2) the pLDDT scores, both for each amino acid and also for the entire structure, in which higher scores (out of 100), ideally >70% (“ok”), especially >80% (“confident”), mean more confidence and, as a consequence, better models (Supplemental Fig. S10). The best model (rank 1, among five runs in total), that is, the one with the highest pLDDT score, was always chosen.

Alternative promoter usage analysis

To investigate alternative promoter usage of HER2 isoforms, we analyzed splice junction data from 561 TCGA breast cancer RNA-seq samples. Following the methodology of Demircioğlu et al. (2019), we first identified TSSs, defined as the start of the first exon, for 90 protein-coding HER2 isoforms (annotated in GENCODE [version 36] or identified through long-read RNA-seq). Overlapping first exons were grouped to define promoter-regulated transcript sets, with the 5'-most TSS selected to represent each promoter. Internal promoters, those whose first splice junctions match internal junctions of other isoforms, were identified based on splice junction coordinates. Promoter activity was quantified using the junction read counts method (Demircioğlu et al. 2019), which aggregates read counts from the first splice junctions of transcripts regulated by each promoter. For internal promoters, we applied the normalization strategy of Zhang et al. (2024), which corrects for ambiguity in read assignment by adjusting donor site read counts based on corresponding acceptor site usage. Absolute promoter activity was calculated using DESeq2 (Love et al. 2014) as $\log_2(\text{total splice junction read counts}/\text{DESeq2 normalization factor})$, and relative promoter activity was derived by dividing each promoter's absolute activity by the total HER2 promoter activity within each sample.

Data sets

All data used in this study are publicly available from the following sources. RNA-seq data from 561 primary breast tumors from female patients were obtained from the NCBI database of Genotypes and Phenotypes (dbGaP; <https://dbgap.ncbi.nlm.nih.gov/>) under TCGA accession number phs000178.v11.p8. Long-read RNA-seq data from 26 tumor and four normal breast samples are available at the European Genome-phenome Archive (EGA; <https://ega-archive.org>) under accession number EGAS00001004819. RNA-seq data from 50 breast cancer cell lines were obtained from the Cancer Cell Line Encyclopedia (CCLE; <https://sites.broadinstitute.org/ccle/>). RNA-seq data from trastuzumab-sensitive and trastuzumab-resistant SK-BR-3 and BT-474 cell lines were obtained from the NCBI BioProject (<https://www.ncbi.nlm.nih.gov>) under accession number PRJNA995876 and the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE244537, respectively. RNA-seq data from T-DM1-sensitive and T-DM1-resistant SK-BR-3 and BT-474 cell lines were obtained under BioProject accession number PRJNA1048320. MS data from 76 breast cancer patients were obtained from the NCI

Proteomic Data Commons (<https://proteomic.datacommons.cancer.gov/pdc/>) under accession number PDC000173.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported by grant 2018/15579–8, São Paulo Research Foundation (FAPESP) to P.A.F.G.; grant 2020/14158–9 (to F.F.d.S.), São Paulo Research Foundation (FAPESP). G.D.A.G. was supported by a fellowship from the Young Scientist program, Hospital Sírio-Libanês. It was also partially supported by funds from CNPq (P.A.F.G., A.A.C.), Serrapilheira Foundation (P.A.F.G. and A.B.), and Hospital Sírio-Libanês to P.A.F.G. and A.A.C.

Author contributions: G.D.A.G., C.H.d.A., and P.A.F.G. developed the concepts in this study. P.A.F.G. supervised the study. Analyses were performed by G.D.A.G. and C.H.d.A. The manuscript was written by G.D.A.G., C.H.d.A., and P.A.F.G. with contributions and revisions from all authors.

References

- Aganezov S, Goodwin S, Sherman RM, Sedlazeck FJ, Arun G, Bhatia S, Lee I, Kirsche M, Wappel R, Kramer M, et al. 2020. Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing. *Genome Res* **30**: 1258–1273. doi:10.1101/gr.260497.119
- Arribas J, Baselga J, Pedersen K, Parra-Palau JL. 2011. p95HER2 and breast cancer. *Cancer Res* **71**: 1515–1519. doi:10.1158/0008-5472.CAN-10-3795
- Arteaga CL, Engelman JA. 2014. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25**: 282–303. doi:10.1016/j.ccr.2014.02.025
- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527. doi:10.1038/nbt.3519
- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. 2024. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **74**: 229–263. doi:10.3322/caac.21834
- The Cancer Genome Atlas Network. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490**: 61–70. doi:10.1038/nature11412
- Castagnoli L, Iezzi M, Ghedini GC, Ciravolo V, Marzano G, Lamolinara A, Zappasodi R, Gasparini P, Campiglio M, Amici A, et al. 2014. Activated d16HER2 homodimers and SRC kinase mediate optimal efficacy for trastuzumab. *Cancer Res* **74**: 6248–6259. doi:10.1158/0008-5472.CAN-14-0983
- Cho H-S, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW Jr, Leahy DJ. 2003. Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* **421**: 756–760. doi:10.1038/nature01392
- Cizkova M, Dujaric M-E, Lehmann-Che J, Scott V, Tembo O, Asselain B, Pierga J-Y, Marty M, de Cremoux P, Spyrtatos F, et al. 2013. Outcome impact of PIK3CA mutations in HER2-positive breast cancer patients treated with trastuzumab. *Br J Cancer* **108**: 1807–1809. doi:10.1038/bjc.2013.164
- Demircioğlu D, Cukuroglu E, Kindermans M, Nandi T, Calabrese C, Fonseca NA, Kahles A, Lehmann K-V, Stegle O, Brazma A, et al. 2019. A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* **178**: 1465–1477.e17. doi:10.1016/j.cell.2019.08.018
- Duan N, Hua Y, Yan X, He Y, Zeng T, Gong J, Fu Z, Li W, Yin Y. 2024. Unveiling alterations of epigenetic modifications and chromatin architecture leading to lipid metabolic reprogramming during the evolutionary trastuzumab adaptation of HER2-positive breast cancer. *Adv Sci (Weinh)* **11**: e2309424. doi:10.1002/adv.202309424
- Dvinge H, Guenthoer J, Porter PL, Bradley RK. 2019. RNA components of the spliceosome regulate tissue- and cancer-specific alternative splicing. *Genome Res* **29**: 1591–1604. doi:10.1101/gr.246678.118
- Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE

- reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766–D773. doi:10.1093/nar/gky955
- Gedik ME, Saatci O, Oberholtzer N, Uner M, Akbulut Caliskan O, Cetin M, Aras M, Ibis K, Caliskan B, Banoglu E, et al. 2024. Targeting TACC3 induces immunogenic cell death and enhances T-DM1 response in HER2-positive breast cancer. *Cancer Res* **84**: 1475–1490. doi:10.1158/0008-5472.CAN-23-2812
- Gordon SP, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, Kang D, Underwood J, Grigoriev IV, Figueroa M, et al. 2015. Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLoS One* **10**: e0132628. doi:10.1371/journal.pone.0132628
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512. doi:10.1038/nprot.2013.084
- Hallgren J, Tsigos KD, Pedersen MD, Armenteros JJA, Marcitili P, Nielsen H, Krogh A, Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv doi:10.1101/2022.04.08.487609
- Jagannatha P, Tankka AT, Lorenz DA, Yu T, Yee BA, Brannan KW, Zhou CJ, Underwood JG, Yeo GW. 2024. Long-read Ribo-STAMP simultaneously measures transcription and translation with isoform resolution. *Genome Res* **34**: 2012–2024. doi:10.1101/gr.279176.124
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583–589. doi:10.1038/s41586-021-03819-2
- Kaur G, Perteghella T, Carbonell-Sala S, Gonzalez-Martinez J, Hunt T, Mądry T, Jungreis I, Aman C, Lagarde J, Borsari B, et al. 2024. GENCODE: massively expanding the lncRNA catalog through capture long-read RNA sequencing. bioRxiv doi:10.1101/2024.10.29.620654
- Kwong KY, Hung MC. 1998. A novel splice variant of HER2 with increased transformation activity. *Mol Carcinog* **23**: 62–68. doi:10.1002/(SICI)1098-2744(199810)23:2<62::AID-MC2>3.0.CO;2-O
- Lee DSM, Park J, Kromer A, Baras A, Rader DJ, Ritchie MD, Ghanem LR, Barash Y. 2021. Disrupting upstream translation in mRNAs is associated with human disease. *Nat Commun* **12**: 1515. doi:10.1038/s41467-021-21812-1
- Loi S, Pommey S, Haibe-Kains B, Beavis PA, Darcy PK, Smyth MJ, Stagg J. 2013. CD73 promotes anthracycline resistance and poor prognosis in triple negative breast cancer. *Proc Natl Acad Sci* **110**: 11091–11096. doi:10.1073/pnas.1222511110
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Luque-Cabal M, García-Tejido P, Fernández-Pérez Y, Sánchez-Lorenzo L, Palacio-Vázquez I. 2016. Mechanisms behind the resistance to trastuzumab in HER2-amplified breast cancer and strategies to overcome it. *Clin Med Insights Oncol* **10s1**: 21–30. doi:10.4137/CMO.S34537
- Marcelino Meliso F, Hubert CG, Favoretto Galante PA, Penalva LO. 2017. RNA processing as an alternative route to attack glioblastoma. *Hum Genet* **136**: 1129–1141. doi:10.1007/s00439-017-1819-2
- Marín A, Mamun AA, Patel H, Akamatsu H, Ye D, Sudhan DR, Eli L, Marcelain K, Brown BP, Meiler J, et al. 2023. Acquired secondary HER2 mutations enhance HER2/MAPK signaling and promote resistance to HER2 kinase inhibition in breast cancer. *Cancer Res* **83**: 3145–3158. doi:10.1158/0008-5472.CAN-22-3617
- Mayr C. 2017. Regulation by 3'-untranslated regions. *Annu Rev Genet* **51**: 171–194. doi:10.1146/annurev-genet-120116-024704
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods* **19**: 679–682. doi:10.1038/s41592-022-01488-1
- Mishra R, Patel H, Alanazi S, Yuan L, Garrett JT. 2018. HER3 signaling and targeted therapy in cancer. *Oncol Rev* **12**: 355. doi:10.4081/oncol.2018.355
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladini L, Raj S, Richardson LJ, et al. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* **49**: D412–D419. doi:10.1093/nar/gkaa913
- Mitra D, Brumlik MJ, Okamgba SU, Zhu Y, Duplessis TT, Parvani JG, Lesko SM, Brogi E, Jones FE. 2009. An oncogenic isoform of HER2 associated with locally disseminated breast cancer and trastuzumab resistance. *Mol Cancer Ther* **8**: 2152–2162. doi:10.1158/1535-7163.MCT-09-0295
- Moasser MM. 2007. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene* **26**: 6469–6487. doi:10.1038/sj.onc.1210477
- Modi S, Saura C, Yamashita T, Park YH, Kim S-B, Tamura K, Andre F, Iwata H, Ito Y, Tsurutani J, et al. 2020. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. *N Engl J Med* **382**: 610–621. doi:10.1056/NEJMoa1914510
- Molina MA, Codony-Servat J, Albanell J, Rojo F, Arribas J, Baselga J. 2001. Trastuzumab (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res* **61**: 4744–4749.
- Mukund K, Alva-Ornelas JA, Maddox AL, Murali D, Veraksa D, Saftics A, Tomsic J, Frankhouser D, Razo M, Jovanovic-Talisman T, et al. 2024. Molecular atlas of HER2+ breast cancer cells treated with endogenous ligands: temporal insights into mechanisms of trastuzumab resistance. *Cancers (Basel)* **16**: 553. doi:10.3390/cancers16030553
- Nagata Y, Lan K-H, Zhou X, Tan M, Esteva FJ, Sahin AA, Klos KS, Li P, Monia BP, Nguyen NT, et al. 2004. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of PTEN predicts trastuzumab resistance in patients. *Cancer Cell* **6**: 117–127. doi:10.1016/j.ccr.2004.06.022
- Nahta R, Esteva FJ. 2006. HER2 therapy: molecular mechanisms of trastuzumab resistance. *Breast Cancer Res* **8**: 215. doi:10.1186/bcr1612
- Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457–463. doi:10.1038/nature08909
- Ødum MT, Teufel F, Thumuluri V, Almagro Armenteros JJ, Johansen AR, Winther O, Nielsen H. 2024. DeepLoc 2.1: multi-label membrane protein type prediction using protein language models. *Nucleic Acids Res* **52**: W215–W220. doi:10.1093/nar/gkae237
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413–1415. doi:10.1038/ng.259
- Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. 2000. Molecular portraits of human breast tumours. *Nature* **406**: 747–752. doi:10.1038/35021093
- Piccatt-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L, Baselga J, Bell R, Jackisch C, et al. 2005. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* **353**: 1659–1672. doi:10.1056/NEJMoa052306
- Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. *Nucleic Acids Res* **46**: W200–W204. doi:10.1093/nar/gky448
- Prat A, Lluich A, Turnbull AK, Dunbier AK, Calvo L, Albanell J, de la Haba-Rodríguez J, Arcusa A, Chacón JI, Sánchez-Rovira P, et al. 2017. A PAM50-based chemoendocrine score for hormone receptor-positive breast cancer with an intermediate risk of relapse. *Clin Cancer Res* **23**: 3035–3044. doi:10.1158/1078-0432.CCR-16-2092
- Reixachs-Solé M, Eyra E. 2022. Uncovering the impacts of alternative splicing on the proteome with current omics techniques. *Wiley Interdiscip Rev RNA* **13**: e1707. doi:10.1002/wrna.1707
- Roma-Rodríguez C, Raposo LR, Cabral R, Paradinha F, Baptista PV, Fernandes AR. 2017. Tumor microenvironment modulation via gold nanoparticles targeting malicious exosomes: implications for cancer diagnostics and therapy. *Int J Mol Sci* **18**: 162. doi:10.3390/ijms18010162
- Scaltriti M, Rojo F, Ocaña A, Anido J, Guzman M, Cortes J, Di Cosimo S, Matias-Guiu X, Ramon y Cajal S, Arribas J, et al. 2007. Expression of p95HER2, a truncated form of the HER2 receptor, and response to anti-HER2 therapies in breast cancer. *J Natl Cancer Inst* **99**: 628–638. doi:10.1093/jnci/djk134
- Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science* **235**: 177–182. doi:10.1126/science.3798106
- Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming J, Eiermann W, Wolter J, Pegram M, et al. 2001. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**: 783–792. doi:10.1056/NEJM200103153441101
- Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**: 1521. doi:10.12688/f1000research.7563.1
- Sonnenblick A, Salmon-Divon M, Salgado R, Dvash E, Pondé N, Zahavi T, Salmon A, Loibl S, Denkert C, Joensuu H, et al. 2020. Reactive stroma and trastuzumab resistance in HER2-positive early breast cancer. *Int J Cancer* **147**: 266–276. doi:10.1002/ijc.32859
- Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* **98**: 10869–10874. doi:10.1073/pnas.191367098
- Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**: 1026–1028. doi:10.1038/nbt.3988
- Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger J, Söding J. 2019. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**: 473. doi:10.1186/s12859-019-3019-7
- Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. 2016. Aberrant RNA splicing in cancer; expression changes and driver mutations of

- splicing factor genes. *Oncogene* **35**: 2413–2427. doi:10.1038/onc.2015.318
- Swain SM, Baselga J, Kim S-B, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero J-M, Schneeweiss A, Heeson S, et al. 2015. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med* **372**: 724–734. doi:10.1056/NEJMoa1413513
- Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, Marra A, Viale G, Trapani D, Cardoso F, et al. 2020. HER2-low breast cancer: pathological and clinical landscape. *J Clin Oncol* **38**: 1951–1962. doi:10.1200/JCO.19.02488
- Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, Ferrell M, Mellado M, Macchietto M, Verheggen K, et al. 2018. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res* **28**: 396–411. doi:10.1101/gr.222976.117
- Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyraes E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**: 40. doi:10.1186/s13059-018-1417-1
- Turpin J, Ling C, Crosby EJ, Hartman ZC, Simond AM, Chodosh LA, Rennhack JP, Andrechek ER, Ozcelik J, Hallett M, et al. 2016. The ErbB2ΔEx16 splice variant is a major oncogenic driver in breast cancer that promotes a pro-metastatic tumor microenvironment. *Oncogene* **35**: 6053–6064. doi:10.1038/onc.2016.129
- Uapinyoying P, Goecks J, Knoblauch SM, Panchapakesan K, Bonnemann CG, Partridge TA, Jaiswal JK, Hoffman EP. 2020. A long-read RNA-seq approach to identify novel transcripts of very large genes. *Genome Res* **30**: 885–897. doi:10.1101/gr.259903.119
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, et al. 2015. Proteomics: tissue-based map of the human proteome. *Science* **347**: 1260419. doi:10.1126/science.1260419
- Vander Velde R, Yoon N, Marusyk V, Durmaz A, Dhawan A, Miroshnychenko D, Lozano-Peral D, Desai B, Balyńska O, Poleszchuk J, et al. 2020. Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. *Nat Commun* **11**: 2393. doi:10.1038/s41467-020-16212-w
- Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S, Mirdita M, Ye J, et al. 2024. AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**: D368–D375. doi:10.1093/nar/gkad1011
- Veiga DFT, Nesta A, Zhao Y, Deslattes Mays A, Huynh R, Rossi R, Wu T-C, Palucka K, Anczukow O, Beck CR, et al. 2022. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. *Sci Adv* **8**: eabg6711. doi:10.1126/sciadv.abg6711
- Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, Pegram M, Oh D-Y, Diéras V, Guardino E, et al. 2012. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* **367**: 1783–1791. doi:10.1056/NEJMoa1209124
- Vernieri C, Milano M, Brambilla M, Mennitto A, Maggi C, Cona MS, Prisciandaro M, Fabbroni C, Celio L, Mariani G, et al. 2019. Resistance mechanisms to anti-HER2 therapies in HER2-positive breast cancer: current knowledge, new research directions and therapeutic perspectives. *Crit Rev Oncol Hematol* **139**: 53–66. doi:10.1016/j.critrevonc.2019.05.001
- von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, Suter T, Arahmani A, Rouchet N, Clark E, et al. 2017. Adjuvant pertuzumab and trastuzumab in early HER2-positive breast cancer. *N Engl J Med* **377**: 122–131. doi:10.1056/NEJMoa1703643
- Wang E, Aifantis I. 2020. RNA splicing and cancer. *Trends Cancer* **6**: 631–644. doi:10.1016/j.trecan.2020.04.011
- Wang Y, Xie Z, Kutschera E, Adams JJ, Kadash-Edmondson KE, Xing Y. 2024. rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nat Protoc* **19**: 1083–1104. doi:10.1038/s41596-023-00944-2
- Wen B, Zhang B. 2023. PepQuery2 democratizes public MS proteomics data for rapid peptide searching. *Nat Commun* **14**: 2213. doi:10.1038/s41467-023-37462-4
- Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JMS, Bilous M, Fitzgibbons P, et al. 2013. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of American pathologists clinical practice guideline update. *J Clin Oncol* **31**: 3997–4013. doi:10.1200/JCO.2013.50.9984
- Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875. doi:10.1093/bioinformatics/bti310
- Yang Q, Zhao J, Zhang W, Chen D, Wang Y. 2019. Aberrant alternative splicing in breast cancer. *J Mol Cell Biol* **11**: 920–929. doi:10.1093/jmcb/mjz033
- Yarden Y, Sliwkowski MX. 2001. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2**: 127–137. doi:10.1038/35052073
- Young SK, Wek RC. 2016. Upstream open reading frames differentially regulate gene-specific translation in the integrated stress response. *J Biol Chem* **291**: 16927–16935. doi:10.1074/jbc.R116.733899
- Zhang M, Sjöström M, Cui X, Foye A, Farh K, Shrestha R, Lundberg A, Dang HX, Li H, Febbo PG, et al. 2024. Integrative analysis of ultra-deep RNA-seq reveals alternative promoter usage as a mechanism of activating oncogenic programmes during prostate cancer progression. *Nat Cell Biol* **26**: 1176–1186. doi:10.1038/s41556-024-01438-3

Received December 9, 2024; accepted in revised form June 30, 2025.