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

Gabriela D. A. Guardia[1,*], Carlos H dos Anjos[1,*], Aline Rangel-Pozzo[2], Filipe F. dos Santos[1,3], Alexander Birbrair[4], Paula F. Asprino[1], Anamaria A. Camargo[1], Pedro A F Galante[1,#]

1 - Centro de Oncologia Molecular, Hospital Sirio Libanes, Sao Paulo, Brazil.

2 - Depart. of Physiology and Pathophysiology, CancerCare Manitoba Research Institute, University of Manitoba, Winnipeg, Canada

3 - Depart. de Bioquímica, Universidade de Sao Paulo, Brazil.

4 - Depart. of Dermatology, University of Wisconsin-Madison, Madison, USA.

* co-first authors. # corresponding author: pgalante@mochsl.org.br

ABSTRACT

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 pose 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 characterized the *HER2* alternative splicing isoforms, assessed their expression in primary BC patients and cell lines, and explored their role in resistance to anti-HER2 therapies. We expanded 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 discovered 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 revealed that drug-resistant cells shifted 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.

INTRODUCTION

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 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 (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 approximately 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 the 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 inhibitors such as lapatinib, neratinib, and tucatinib, and, more recently, HER2-targeted antibody-drug conjugates like T-DM1 and 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, while approximately 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, where the disease progresses shortly after treatment initiation, or acquired resistance, where 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 (Mishra et al. 2018; Nagata et al. 2004; Cizkova et al. 2013). 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 Trastuzumab emtansine (T-DM1) and Trastuzumab deruxtecan (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 this 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, over 90% of multi-exon genes undergo alternative splicing, generating roughly seven mRNA isoforms per gene on average, though only a fraction of these yield functionally distinct proteins (Pan et al. 2008; Uhlén et al. 2015; Reixachs-Solé and Eyras 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 p95HER2 and delta16 HER2 (Δ16HER2) isoforms (Scaltriti et al. 2007; Turpin et al. 2016). The p95HER2 (also known as CTF611 or HER2-CTF; here and after, P95) is an incomplete isoform of the HER2 protein (molecular weight of approximately 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).

The Δ16HER2, results from skipping *HER2* exon 16, which encodes a small portion of the extracellular domain (Kwong and Hung 1998). The Δ16HER2 isoform assembles stable homodimers and is associated with increased transforming activity and metastatic potential (Turpin et al. 2016). Controversially, studies suggest that Δ16HER2 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 a 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 investigate their potential role in mediating resistance to anti-HER2 therapies, particularly antibody-drug conjugates. 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

144 **Assessing the *HER2* splicing isoform diversity in breast cancer**

145

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

167

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 approximately 42.5 kilobases on Chromosome 17 (Chr17: 39,687,914 - 39,730,426; reference genome version GRCh38), comprising 27 canonical exons and 13 distinct protein-coding isoforms, along with 9 non-coding 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 data. By using full-length mRNA transcripts (Veiga et al. 2022) (see Methods), 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 (A5, 26 isoforms), alternative 3' splice sites (A3, 15 isoforms), and other types (12 isoforms) (Figure 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 (Figure 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: i) completeness of HER2 protein domains (Receptor L, Furin-like, Growth factor receptor and Protein kinase); ii) predicted cellular localization (cell membrane, cytoplasm, extracellular, or Golgi apparatus); iii) presence of the trastuzumab-binding domain; and iv) presence of the immunohistochemical (IHC) ligand domain (Figure

2B; Supplemental Table S2; see Methods for details). 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 (Figure 2B). To illustrate how the combination of multiple splicing events (e.g., exon skipping and intron retention) shape HER2 isoform diversity, Supplemental Fig. 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 Δ16HER2. Here, we took advantage of the full-length transcript analysis to examine beyond individual splicing events (e.g., exon skipping), 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 (Δ16HER2 variants): only one reported in GENCODE (ENST00000580074.1) and eight novel isoforms discovered through our long-read sequencing strategy (Supplemental Fig. S2A). These Δ16HER2 isoforms are distinct because they exhibited additional alternative splicing events, including alternative first exon usage, 5'

alternative splice site selection, skipping of exons 19 and 24, and, in one case, three additional alternative 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 patterns, including alternative first exon usage (1 isoform), exon skipping (1 isoform), combinations of alternative first exon with other splicing events (6 isoforms), and upstream small ORFs (see below).

Collectively, this comprehensive characterization reveals a complex and diverse landscape of *HER2* splicing isoforms, extending even to well-studied variants like p95 and Δ16. 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* isoforms 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 and S3). Transcript length analysis revealed two significant sets of isoforms. First, isoforms with a primary peak length near the canonical *HER2* transcript (ENST00000269571; 4,557 nt) and another minor peak around 2,600 nt (Supplemental Fig. S3A). Protein length distribution showed a significant peak corresponding to the canonical HER2 protein (1,255 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 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 Δ16 and p95 HER2 splicing isoforms (Supplemental Fig. S4B and S4B, respectively). While the eight p95 isoforms showed high similarities to each other at both transcript and protein levels, three of the nine Δ16 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 (Cancer Genome Atlas Network 2012) (Figure 1B). To have groups of clinically and biologically similar tumors and patients, we initially stratified tumors into six specific subgroups based on their hormone receptor (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 sub-classified in 442 samples HR+, including 40 with HER2-zero, 308 with HER2-low, and 94 samples with HER2-high status; 119 samples ER-, divided into 18 HER2-zero, 68 HER2-low, and 33 HER2-high samples (Figure 1B, Supplemental Table S5).

To confirm the splicing changes observed in *HER2* full-length isoforms (Figure 2), we examined exon-exon junction read counts from TCGA breast cancer samples. This analysis reproduced the key splicing patterns initially identified, including exon skipping, 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 dataset, 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 (Figure 3A and 3B). 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+ (Figure 3A) and HR- (Figure 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 CCLE (cell lines) samples from distinct HR and HER2 statuses. First, using a threshold of TPM > 0, expression evidence was observed for all isoforms (Supplemental Fig. S5). When applying more

stringent thresholds (TPM ≥ 1 and TPM ≥ 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 (Figures 3C (HR+) and 3D (HR-), the top 10 most expressed isoforms; Supplemental Fig. S6 and Supplemental Table S9, all 90 isoforms). 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) (Figure 3C-D). However, the canonical isoform was not the highest expressed within all groups. In HR+ samples (Figure 3C), the canonical *HER2* isoform exhibited a gradient of expression (median: HER2-high ~6 TPM; HER2-low ~ 4 TPM, HER2-zero ~ 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 (Figure 3D), while overall patterns of *HER2* splicing isoforms were similar to HR+, some differences were highlighted. The median of expression of HER2-high versus HER2-low and -zero was more pronounced for multiple splicing isoforms, including the canonical. An increased expression variability was noted in the HER2-high group for specific

306 isoforms (e.g., ISO 1: PB.14155.385). Although some differences were observed, no
307 distinct expression patterns were evident within patient groups.

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

317 Subsequently, we assessed the prevalence of (% of TCGA patient samples
318 expressing - TPM ≥ 1) different HER2 isoform groups (ISO 1-13) across breast cancer
319 subtypes, as shown in Figures 3E (HR+) and 3F (HR-). In HR+ samples, isoform groups
320 ISO 1, 2, 5, 6, 7, 9, and 12 showed moderate to high prevalence across all HER2
321 categories (zero, low, and high). A similar overall pattern was observed in HR- samples,
322 although with some differences. HER2-high tumors showed more uniform prevalence
323 across all isoform groups than HER2-low and HER2-zero categories. ISO 1-7 were more
324 expressed in HER2-high HR- samples. HER2-zero and HER2-low HR- samples showed
325 more selective isoform expression, with ISO 1, 2, 5, and 6 being the most consistently
326 expressed groups. In both HR+ and HR-, the HER2-high presented the highest
327 number of isoforms expressed in the highest number of samples.

328 Given the established association between p95 expression and poor clinical
329 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, while 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 to 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 influences 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

Since 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 (Figure 4A, Supplemental Table S12). Consistent with RNA-sequencing (RNA-seq) data, HR+ tumors showed proportionally more proteins confirmed by MS than HR- tumors. Similarly, HER2-high tumors presented more validated proteins than HER2-low and HER2-zero tumors, with the latter showing the lowest level of validation (Figure 4A, bottom panel). While 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 (Figure 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 (Figure 4B and Supplemental Fig. S10). In contrast, ISO 5 (PB14155.141) shows significant structural alterations, particularly in the extracellular region (Figure 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 (Figure 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 (Percent 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 (Figure 5A). In HR+ patients, HER2-high samples showed significantly lower dissimilarity than 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 zero (p-value < 0.05; Mann-Whitney U test), Figure 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 (HR+ breast cancer samples: Figure 5B; HR- breast cancer samples: 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) (Figure 5B for HR+ samples; Supplemental Fig. S11A for HR- samples), 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 to all others, but with distinct expression patterns among them. Specifically, in HR+ clusters Z1, L1, and H3 (Figure 5B), as well as HR- clusters H3, L3, and Z2 (Supplemental Fig. S11A), we observed high expression of ISO 9 and ISO 5. Since these isoforms lack the trastuzumab-binding domain, their predominance might predict poor response to HER2-targeted therapies. Conversely, HR+ clusters H1, L2, and Z3 (Figure 5B), along with HR- clusters H1, L2, and Z3 (Supplemental Fig. S11A), showed 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 (Figure 5B; H2, L1, and Z1 for HR-, 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. While the subgroup L3 closely resembled the HER2-high profile, the L1 showed isoform expression patterns more similar to 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 $\log_2$(TMP+1)) compared to the HER2-low (median 6.5-7.5 $\log_2$(TMP+1)) and HER2-zero (median 5.5-7 $\log_2$(TMP+1)) clusters

(HR+, Figure 5C; HR-, Supplemental Fig. S11B). No statistically significant differences were found in total *HER2* expression among clusters within each HER2 group , except for Z3 (HR+).

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 (HR+, Figure 5D; HR-, 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 isoforms (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 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 HR+ and HR- samples (Supplemental Fig. S12 and S13, respectively). Overall, HR+ and HR- samples showed a similar expression pattern in terms of the most abundant 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, while 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 maps transcription start sites through detection of 5'-capped RNA, and H3K4me3 ChIP-seq data from ENCODE, which marks 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 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) - which 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).

All together, 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 antibody-drug conjugate 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 (Figure 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, while UACC-812 was resistant. In HR-/HER2+ cells, AU565, HCC1954, MDA-MB-453, and SK-BR-3 demonstrated sensitivity to T-DM1, while JIMT-1 and UACC-893 were resistant. MDA-MB-453 and SK-BR-3 showed sensitivity to both T-DM1 and T-DXd, while 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. While 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 (ISO 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 (Figure 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 (Figure 6B). Since 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 maintain a balanced ratio between ISO 1-4 and ISO 5-13 expression. UACC-812, in spite of 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 to 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, but 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 non-responsive (NR) to antibody-drug conjugates (Supplemental Fig. S17). We found significantly increased expression of isoform groups ISO 1, 6, and 12 in responsive cell lines. While 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, though the functional implications remain to be determined.

Altogether, these results indicated that high HER2 expression and a balanced - though 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 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 (Figure 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 (Figure 7A). Both cell lines in the two states maintained significantly high levels of *HER2* expression (Figure 7A), indicating that resistance to trastuzumab is not primarily mediated by a global downregulation of *HER2* expression (Vernieri et al. 2019). In fact, since trastuzumab targets HER2 function, we hypothesize that its upregulation may serve as a compensatory

591 mechanism to offset its own inhibition by treatment, thus sustaining downstream
592 signaling pathways that promote rapid tumor cell growth and proliferation.

593 Next, to gain deeper insights into potential resistance mechanisms, we
594 analyzed the *HER2* isoform expression profiles and fold change of each isoform group
595 in trastuzumab-resistant versus sensitive cells for both SK-BR-3 and BT-474 lines
596 (Figure 7B; Supplemental Fig. S18). In SK-BR-3 cells, we observed significant
597 upregulation of multiple isoform sets. In contrast, isoforms from group 8 showed
598 downregulation (log_2 fold change ~ -0.4) in resistant conditions (Figure 7B;
599 Supplemental Fig. 16A). Overall, BT-474 cells also exhibited upregulation of sets of
600 isoforms, except for isoforms from groups 3, 5, and 8 (Supplemental Fig. 16B).

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

607 Lastly, we performed a detailed analysis of *HER2* isoform expression levels in T-
608 DM1-resistant and -sensitive SK-BR-3 and BT-474 cells. We observed complex
609 changes in the isoform landscape for SK-BR-3 (Figure 7D) and BT-474 (Supplemental
610 Fig. S19B). First, the fold change (T-DM1 resistant/sensitive) analysis showed that most
611 splicing isoforms are significantly downregulated, including isoforms from sets 1 to 4
612 (isoforms with intact domains; Figure 7D). Splicing isoforms from sets 5 and 9 are
613 upregulated in T-DM1-resistant cells (Figure 7D). In BT-474, we observed a significant
614 downregulation of all splicing isoforms (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 to 4) and upregulating the pro-survival signaling (isoforms from groups ISO 5 and 9, which lack the trastuzumab binding site and seems 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, Δ16 and p95: i) identified 8 distinct isoforms with different splicing events

639 that encode proteins lacking the extracellular domain and have an approximate
640 molecular weight of 95 kDa; ii) for Δ16, we found 9 isoforms, all containing the exon 16
641 skipping (Δ16's hallmark) and exhibiting other alternative splicing events.
642 Understanding not just the event (e.g., exon skipping), but the full set of isoforms that
643 contains such events and others certainly gives us a more complete understanding of
644 the importance and functionality of each alternative splicing isoform.

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

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

659 Identifying HER2 splicing isoforms lacking the antibody (drug) binding
660 domains, but retaining the signaling capabilities (tyrosine kinase domain), may
661 explain a potential resistance mechanism to antibody-based therapies like
662 trastuzumab and ADCs. This aligns with previous studies on p95, which lacks the

663 trastuzumab binding site and has been associated with poor prognosis and resistance
664 in antibody-based therapy treatment (Scaltriti et al. 2007). Therefore, the expanded
665 repertoire of HER2 isoforms presented here suggests that alternative splicing may be
666 a more prevalent and leading mechanism used by cancer cells in acquiring resistance
667 and progression, especially in gene-targeted therapies.

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

681 Furthermore, our observations indicate that among tumor subtypes (HER2-
682 high, -low, and -zero), there are distinct subgroups of tumors expressing different
683 splicing isoforms, some of which encode protein isoforms lacking the binding domain
684 for antibodies used in immunohistochemistry. This may explain why a percentage of
685 HER2-low or even HER2-zero patients (as determined by immunohistochemistry)
686 respond to treatment. It is reasonable to hypothesize that specific isoform

687 compositions may create a false classification of HER2 status and vulnerabilities to
688 ADC treatment, even in contexts with low *HER2* expression (by
689 immunohistochemistry) (Tarantino et al. 2020).

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

709 Despite these limitations, our findings suggest that incorporating HER2
710 isoform profiling into clinical assessment could significantly enhance prediction of

response to trastuzumab and antibody-drug conjugates. 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 antibody-drug conjugates (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 sequencing data

We obtained unprocessed RNA sequencing (RNA-seq) data from 561 primary tumors sourced from female breast cancer patients, publicly accessible via The Cancer Genome Atlas (TCGA) repository (https://portal.gdc.cancer.gov). Additionally, we obtained clinical data detailing immunohistochemical (IHC) staining results for HER2, estrogen receptor (ER), and progesterone receptor (PR), as well as fluorescence *in situ* hybridization (FISH) data for HER2. In addition, RNA-seq data from 50 breast cancer cell lines were acquired from the Cancer Cell Line Encyclopedia (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 -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 -resistant SK-BR-3 and BT-474 cell lines were obtained from (Gedik et al. 2024)

Public long-read RNA sequencing data

We obtained processed long-read RNA sequencing data from 26 tumor and 4 normal breast samples from (Veiga et al. 2022) to create an expanded catalog of *HER2* isoforms. Briefly, in (Veiga et al. 2022), full-length cDNA was synthesized from poly-A+ transcripts, PCR-amplified, and sequenced on 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.gencodegenes.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 pseudo-aligned the RNA sequencing 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 transcripts per million (TPM) were submitted to SUPPA2 (version 2.3; default parameters) (Trincado et al. 2018), which quantifies percent spliced-in (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 (IR) events, additional parameters --boundary V and --threshold 10 were applied. Multiple exon skipping (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 (v4.3.0) (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 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

804 based on the DeepLoc 2.0 tool (default parameters) (Ødum et al. 2024). To evaluate
805 the presence of the immunohistochemical (IHC)-binding region in HER2 isoforms, we
806 considered 3 IHC antibodies: PATHWAY HER2, Herceptest and Oracle HER (Cho et al.
807 2003). Besides, similarities among HER2 proteins were evaluated through pairwise
808 protein alignments using the needle global aligner
809 (https://www.ebi.ac.uk/jdispatcher/psa/emboss_needle).

810 **Validation of HER2 isoforms at the protein level**

811

812 Validation of HER2 isoforms at the protein level was performed using MS/MS data
813 from 76 breast cancer patients in the CPTAC TCGA dataset (study ID: PDC000173). We
814 employed the PepQuery tool (v2.0.2) (Wen and Zhang 2023) with default parameters,
815 using as reference 103,069 proteins from GENCODE release 36. Each HER2 protein
816 derived from expressed splicing isoforms was queried in the MS/MS data to identify
817 supporting peptide spectrum matches (PSMs). Following *in silico* trypsin digestion,
818 PepQuery attempted to validate the resulting peptides against the mass
819 spectrometry data. After multiple filtering steps, only PSMs passing all criteria with an
820 FDR < 0.05 were considered confident. Validated peptides were then assigned to their
821 respective isoform groups (ISO 1 to 13), rather than being restricted to unique peptides
822 for each protein. Given the high sequence similarity among HER2 proteins, this
823 approach allowed for a more comprehensive and reliable assessment of isoform
824 detection.

825

826 **Prediction of 3D protein structures of HER2 isoforms**

827 The prediction of the HER2 isoforms' 3D protein structures was made with AlphaFold2
828 (Jumper et al. 2021) through the free and publicly available Google collaborator
829 ColabFold (v1.5.5) platform (Mirdita et al. 2022). We opted to run the predictions this
830 way due to its speed by combining it with a fast homology search with MMseqs2
831 (Steinegger and Söding 2017) and HHsearch (Steinegger et al. 2019), and the usage of
832 the highly accurate PDB100 (Varadi et al. 2024) as its database. All analyses were run
833 in a “High-RAM (system: 51GB; GPU: 15GB) T4 GPU” machine with Python 3 and more
834 than 200GB of disk space.

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

847 **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. (Demircioğlu et al. 2019), we first identified transcription start sites (TSSs), defined as the start of the first exon, for 90 protein-coding HER2 isoforms (annotated in GENCODE (v36) or identified through long-read RNA sequencing). 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. (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 $\log_2$(total splice junction read counts / 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.

DATASETS

All data used in this study are publicly available from the following sources. RNA sequencing (RNA-seq) data from 561 primary breast tumors from female patients were obtained from the database of Genotypes and Phenotypes (dbGaP) under TCGA accession number phs000178.v11.p8. Long-read RNA sequencing data from 26 tumor

and 4 normal breast samples are available at the European Genome-phenome Archive (EGA) 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 -resistant SK-BR-3 and BT-474 cell lines were obtained under accession PRJNA995876 and GSE244537, respectively. RNA-seq data from T-DM1-sensitive and -resistant SK-BR-3 and BT-474 cell lines were obtained under access PRJNA1048320. Mass spectrometry data from 76 breast cancer patients were obtained under accession 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 PAFG; grants 2020/14158-9 (to FFS), São Paulo Research Foundation (FAPESP). GDAG. was supported by a fellowship from the Young Scientist program, Hospital Sírio-Libanês. It was also partially supported by funds from CNPq (PAFG, AAC), Serrapilheira Foundation (PAFG and AB), and Hospital Sírio-Libanês to PAFG and AAC.

AUTHOR CONTRIBUTIONS

G.D.A.G., CH dos 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 CH dos A. The manuscript was written by G.D.A.G., CH dos 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.

Arribas J, Baselga J, Pedersen K, Parra-Palau JL. 2011. p95HER2 and breast cancer. *Cancer Res* **71**: 1515–1519.

Arteaga CL, Engelman JA. 2014. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25**: 282–303.

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.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527.

Cancer Genome Atlas Network. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490**: 61–70.

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.

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.

Cizkova M, Dujaric M-E, Lehmann-Che J, Scott V, Tembo O, Asselain B, Pierga J-Y, Marty M, de Cremoux P, Spyratos F, et al. 2013. Outcome impact of PIK3CA mutations in HER2-positive breast cancer patients treated with trastuzumab. *Br J Cancer* **108**: 1807–1809.

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.

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.

928 Dvinge H, Guenthoer J, Porter PL, Bradley RK. 2019. RNA components of the spliceosome
929 regulate tissue- and cancer-specific alternative splicing. *Genome Res* **29**: 1591–1604.

930 Frankish A, Diekhans M, Ferreira A-M, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C,
931 Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and
932 mouse genomes. *Nucleic Acids Res* **47**: D766–D773.

933 Gedik ME, Saatci O, Oberholtzer N, Uner M, Akbulut Caliskan O, Cetin M, Aras M, Ibis K,
934 Caliskan B, Banoglu E, et al. 2024. Targeting TACC3 Induces Immunogenic Cell Death and
935 Enhances T-DM1 Response in HER2-Positive Breast Cancer. *Cancer Res* **84**: 1475–1490.

936 Gordon SP, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, Kang D, Underwood J, Grigoriev IV,
937 Figueroa M, et al. 2015. Widespread Polycistronic Transcripts in Fungi Revealed by Single-
938 Molecule mRNA Sequencing. *PLoS One* **10**: e0132628.

939 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
940 Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq
941 using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512.

942 Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, Krogh A,
943 Winther O. 2022. DeepTMHMM predicts alpha and beta transmembrane proteins using
944 deep neural networks. *bioRxiv* 2022.04.08.487609.
945 https://www.biorxiv.org/content/10.1101/2022.04.08.487609v1.abstract (Accessed April 30,
946 2024).

947 Jagannatha P, Tankka AT, Lorenz DA, Yu T, Yee BA, Brannan KW, Zhou CJ, Underwood JG, Yeo
948 GW. 2024. Long-read Ribo-STAMP simultaneously measures transcription and translation
949 with isoform resolution. *Genome Res* **34**: 2012–2024.

950 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates
951 R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with
952 AlphaFold. *Nature* **596**: 583–589.

953 Kaur G, Perteghella T, Carbonell-Sala S, Gonzalez-Martinez J, Hunt T, Mądry T, Jungreis I,
954 Arnan C, Lagarde J, Borsari B, et al. 2024. GENCODE: massively expanding the lncRNA
955 catalog through capture long-read RNA sequencing. *bioRxiv*.
956 http://dx.doi.org/10.1101/2024.10.29.620654.

957 Kwong KY, Hung MC. 1998. A novel splice variant of HER2 with increased transformation
958 activity. *Mol Carcinog* **23**: 62–68.

959 Lee DSM, Park J, Kromer A, Baras A, Rader DJ, Ritchie MD, Ghanem LR, Barash Y. 2021.
960 Disrupting upstream translation in mRNAs is associated with human disease. *Nat*
961 *Commun* **12**: 1515.

962 Loi S, Pommey S, Haibe-Kains B, Beavis PA, Darcy PK, Smyth MJ, Stagg J. 2013. CD73
963 promotes anthracycline resistance and poor prognosis in triple negative breast cancer.
964 *Proc Natl Acad Sci U S A* **110**: 11091–11096.

965 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
966 RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

967 Luque-Cabal M, García-Teijido P, Fernández-Pérez Y, Sánchez-Lorenzo L, Palacio-Vázquez I.
968 2016. Mechanisms behind the resistance to trastuzumab in HER2-amplified breast cancer
969 and strategies to overcome it. *Clin Med Insights Oncol* **10**: 21–30.

970 Marcelino Meliso F, Hubert CG, Favoretto Galante PA, Penalva LO. 2017. RNA processing as an
971 alternative route to attack glioblastoma. *Hum Genet* **136**: 1129–1141.

972 Marín A, Mamun AA, Patel H, Akamatsu H, Ye D, Sudhan DR, Eli L, Marcelain K, Brown BP,
973 Meiler J, et al. 2023. Acquired secondary HER2 mutations enhance HER2/MAPK signaling
974 and promote resistance to HER2 kinase inhibition in breast cancer. *Cancer Res* **83**: 3145–
975 3158.

976 Mayr C. 2017. Regulation by 3'-untranslated regions. *Annu Rev Genet* **51**: 171–194.

977 Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold:
978 making protein folding accessible to all. *Nat Methods* **19**: 679–682.

979 Mishra R, Patel H, Alanazi S, Yuan L, Garrett JT. 2018. HER3 signaling and targeted therapy in
980 cancer. *Oncol Rev* **12**: 355.

981 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE,
982 Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein families database in 2021.
983 *Nucleic Acids Res* **49**: D412–D419.

984 Mitra D, Brumlik MJ, Okamgba SU, Zhu Y, Duplessis TT, Parvani JG, Lesko SM, Brogi E, Jones
985 FE. 2009. An oncogenic isoform of HER2 associated with locally disseminated breast
986 cancer and trastuzumab resistance. *Mol Cancer Ther* **8**: 2152–2162.

987 Moasser MM. 2007. The oncogene HER2: its signaling and transforming functions and its role
988 in human cancer pathogenesis. *Oncogene* **26**: 6469–6487.

989 Modi S, Saura C, Yamashita T, Park YH, Kim S-B, Tamura K, Andre F, Iwata H, Ito Y, Tsurutani J,
990 et al. 2020. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. *N*
991 *Engl J Med* **382**: 610–621.

992 Molina MA, Codony-Servat J, Albanell J, Rojo F, Arribas J, Baselga J. 2001. Trastuzumab
993 (herceptin), a humanized anti-Her2 receptor monoclonal antibody, inhibits basal and
994 activated Her2 ectodomain cleavage in breast cancer cells. *Cancer Res* **61**: 4744–4749.

995 Mukund K, Alva-Ornelas JA, Maddox AL, Murali D, Veraksa D, Saftics A, Tomsic J, Frankhouser
996 D, Razo M, Jovanovic-Talisman T, et al. 2024. Molecular Atlas of HER2+ Breast Cancer Cells
997 Treated with Endogenous Ligands: Temporal Insights into Mechanisms of Trastuzumab
998 Resistance. *Cancers (Basel)* **16**. http://dx.doi.org/10.3390/cancers16030553.

999 Nagata Y, Lan K-H, Zhou X, Tan M, Esteva FJ, Sahin AA, Klos KS, Li P, Monia BP, Nguyen NT, et
1000 al. 2004. PTEN activation contributes to tumor inhibition by trastuzumab, and loss of
1001 PTEN predicts trastuzumab resistance in patients. *Cancer Cell* **6**: 117–127.

1002 Nahta R, Esteva FJ. 2006. HER2 therapy: molecular mechanisms of trastuzumab resistance.
1003 *Breast Cancer Res* **8**: 215.

1004 Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing.
1005 *Nature* **463**: 457–463.

1006 Ødum MT, Teufel F, Thumuluri V, Almagro Armenteros JJ, Johansen AR, Winther O, Nielsen H.
1007 2024. DeepLoc 2.1: multi-label membrane protein type prediction using protein language
1008 models. *Nucleic Acids Res*. http://dx.doi.org/10.1093/nar/gkae237.

1009 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing

1010 complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**:
1011 1413–1415.

1012 Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen
1013 H, Akslen LA, et al. 2000. Molecular portraits of human breast tumours. *Nature* **406**: 747–
1014 752.

1015 Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, Gianni L,
1016 Baselga J, Bell R, Jackisch C, et al. 2005. Trastuzumab after adjuvant chemotherapy in
1017 HER2-positive breast cancer. *N Engl J Med* **353**: 1659–1672.

1018 Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018
1019 update. *Nucleic Acids Res* **46**: W200–W204.

1020 Prat A, Lluch A, Turnbull AK, Dunbier AK, Calvo L, Albanell J, de la Haba-Rodríguez J, Arcusa A,
1021 Chacón JI, Sánchez-Rovira P, et al. 2017. A PAM50-based chemoendocrine score for
1022 hormone receptor-positive breast cancer with an intermediate risk of relapse. *Clin*
1023 *Cancer Res* **23**: 3035–3044.

1024 Reixachs-Solé M, Eyras E. 2022. Uncovering the impacts of alternative splicing on the
1025 proteome with current omics techniques. *Wiley Interdiscip Rev RNA* **13**: e1707.

1026 Roma-Rodrigues C, Raposo LR, Cabral R, Paradinha F, Baptista PV, Fernandes AR. 2017.
1027 Tumor microenvironment modulation via gold nanoparticles targeting malicious
1028 exosomes: Implications for cancer diagnostics and therapy. *Int J Mol Sci* **18**: 162.

1029 Scaltriti M, Rojo F, Ocaña A, Anido J, Guzman M, Cortes J, Di Cosimo S, Matias-Guiu X, Ramon
1030 y Cajal S, Arribas J, et al. 2007. Expression of p95HER2, a truncated form of the HER2
1031 receptor, and response to anti-HER2 therapies in breast cancer. *J Natl Cancer Inst* **99**:
1032 628–638.

1033 Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. 1987. Human breast cancer:
1034 correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*
1035 **235**: 177–182.

1036 Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann
1037 W, Wolter J, Pegram M, et al. 2001. Use of chemotherapy plus a monoclonal antibody
1038 against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**:
1039 783–792.

1040 Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level
1041 estimates improve gene-level inferences. *F1000Res* **4**: 1521.

1042 Sonnenblick A, Salmon-Divon M, Salgado R, Dvash E, Pondé N, Zahavi T, Salmon A, Loibl S,
1043 Denkert C, Joensuu H, et al. 2020. Reactive stroma and trastuzumab resistance in HER2-
1044 positive early breast cancer. *Int J Cancer* **147**: 266–276.

1045 Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M,
1046 Jeffrey SS, et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor
1047 subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**: 10869–10874.

1048 Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. 2019. HH-suite3 for
1049 fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**:
1050 473.

1051 Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the
1052 analysis of massive data sets. *Nat Biotechnol* **35**: 1026–1028.

1053 Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. 2016. Aberrant RNA splicing in
1054 cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**:
1055 2413–2427.

1056 Swain SM, Baselga J, Kim S-B, Ro J, Semiglazov V, Campone M, Ciruelos E, Ferrero J-M,
1057 Schneeweiss A, Heeson S, et al. 2015. Pertuzumab, trastuzumab, and docetaxel in HER2-
1058 positive metastatic breast cancer. *N Engl J Med* **372**: 724–734.

1059 Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, Marra A, Viale G, Trapani
1060 D, Cardoso F, et al. 2020. HER2-low breast cancer: Pathological and clinical landscape. *J*
1061 *Clin Oncol* **38**: 1951–1962.

1062 Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, Ferrell M,
1063 Mellado M, Macchietto M, Verheggen K, et al. 2018. SQANTI: extensive characterization of
1064 long-read transcript sequences for quality control in full-length transcriptome
1065 identification and quantification. *Genome Res* **28**: 396–411.

1066 Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E. 2018. SUPPA2: fast,
1067 accurate, and uncertainty-aware differential splicing analysis across multiple conditions.
1068 *Genome Biol* **19**: 40.

1069 Turpin J, Ling C, Crosby EJ, Hartman ZC, Simond AM, Chodosh LA, Rennhack JP, Andrechek
1070 ER, Ozcelik J, Hallett M, et al. 2016. The ErbB2ΔEx16 splice variant is a major oncogenic
1071 driver in breast cancer that promotes a pro-metastatic tumor microenvironment.
1072 *Oncogene* **35**: 6053–6064.

1073 Uapinyoying P, Goecks J, Knoblach SM, Panchapakesan K, Bonnemann CG, Partridge TA,
1074 Jaiswal JK, Hoffman EP. 2020. A long-read RNA-seq approach to identify novel transcripts
1075 of very large genes. *Genome Res* **30**: 885–897.

1076 Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å,
1077 Kampf C, Sjöstedt E, Asplund A, et al. 2015. Proteomics. Tissue-based map of the human
1078 proteome. *Science* **347**: 1260419.

1079 Vander Velde R, Yoon N, Marusyk V, Durmaz A, Dhawan A, Miroshnychenko D, Lozano-Peral
1080 D, Desai B, Balynska O, Poleszhuk J, et al. 2020. Resistance to targeted therapies as a
1081 multifactorial, gradual adaptation to inhibitor specific selective pressures. *Nat Commun*
1082 **11**: 2393.

1083 Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M, Nair S,
1084 Mirdita M, Yeo J, et al. 2024. AlphaFold Protein Structure Database in 2024: providing
1085 structure coverage for over 214 million protein sequences. *Nucleic Acids Res* **52**: D368–
1086 D375.

1087 Veiga DFT, Nesta A, Zhao Y, Deslattes Mays A, Huynh R, Rossi R, Wu T-C, Palucka K, Anczukow
1088 O, Beck CR, et al. 2022. A comprehensive long-read isoform analysis platform and
1089 sequencing resource for breast cancer. *Sci Adv* **8**: eabg6711.

1090 Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, Pegram M, Oh D-Y, Diéras V,
1091 Guardino E, et al. 2012. Trastuzumab emtansine for HER2-positive advanced breast
1092 cancer. *N Engl J Med* **367**: 1783–1791.

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.

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.

Wang E, Aifantis I. 2020. RNA splicing and cancer. *Trends Cancer* **6**: 631–644.

Wang Y, Xie Z, Kutschera E, Adams JI, 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.

Wen B, Zhang B. 2023. PepQuery2 democratizes public MS proteomics data for rapid peptide searching. *Nat Commun* **14**: 1–13.

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.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.

Yang Q, Zhao J, Zhang W, Chen D, Wang Y. 2019. Aberrant alternative splicing in breast cancer. *J Mol Cell Biol* **11**: 920–929.

Yarden Y, Sliwkowski MX. 2001. Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* **2**: 127–137.

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.

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.

FIGURE LEGENDS

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 samples selection and stratification based on their 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).

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 Δ16 are represented by "p" and "Δ", respectively. The heatmap shows the presence (dark gray squares) or absence (light gray squares) of specific protein domain configurations (rows) for each isoform-encoded protein (column). 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.

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 **A)** HR+ and **B)** HR- 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 **C)** HR+ and **D)** HR- breast cancer samples, stratified by HER2 status. The canonical *HER2* isoform is highlighted in bold. Expression levels are shown in $\log_2$(TPM+1). **E-F)** Dot plots showing the percentage of **E)** HR+ and **F)** HR- TCGA breast cancer samples expressing each *HER2* isoform group (ISO 1-13), stratified by HER2 status - zero (blue), low (green), high(red).

Figure 4. Mass spectrometry validation of HER2 isoform-derived proteins and their predicted 3D structures. A) The top panels display the number of samples with mass spectrometry (mass spec) 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)** shows the canonical isoform (ISO 1: ENST00000269571), with well-defined domain regions. **C)** displays ISO 5 (PB.14155.141), a variant from Group 5 with specific domain alterations affecting its cellular localization and trastuzumab binding potential. **D)** represents 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.

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 hormone receptor-positive (HR+) and hormone receptor-negative (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, Z3 for HER2-zero; L1, L2, L3 for HER2-low; H1, H2, H3 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.

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 "check symbol." Unresponsive cell lines are marked with a cross symbol. If drug treatment is not available: "NA". The square block's color represents the total *HER2* expression level ($\log_2$(TPM+1)). Blue semi-circular plots indicate the expression of isoforms with intact trastuzumab-binding domain (ISO 1-4). Red semi-circular plots half circus 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 ($\log_2$(TPM+1)). Cell lines named in green indicate

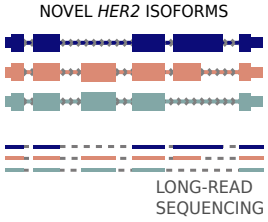
1206 those responsive to antibody-drug conjugates (ADCs: T-DM1 or T-DXd); Cell lines named in gray
1207 indicate those ADC-resistant. **C)** *HER2* isoform expression patterns across breast cancer cell
1208 lines (without ADC treatment) stratified by HR/HER2 status. Semi-circular plots represent
1209 expression levels of isoforms from groups 1-4 (blue) and 5-13 (red). The gray squares below
1210 indicate total *HER2* expression levels.

1211

1212 **Figure 7. *HER2* gene and isoform expression profiles in breast cancer cell lines before and**
1213 **after acquiring resistance to HER2-targeted therapies. A)** Overall *HER2* expression in
1214 trastuzumab-sensitive (S) and -resistant (R) SK-BR-3 and BT-474 cells. **B)** $\log_2$ fold change of
1215 *HER2* isoform expression in trastuzumab-resistant vs. sensitive SK-BR-3 and BT-474 cells. **C)**
1216 Overall *HER2* expression in T-DM1-sensitive and -resistant SK-BR-3 cells. **D)** *HER2* isoform
1217 expression levels and characteristics in T-DM1-sensitive and -resistant SK-BR-3 cells, with $\log_2$
1218 fold change shown above. Isoforms are categorized based on their cell localization,
1219 trastuzumab-binding ligand presence (T-ligand), structural completeness and
1220 transmembrane topology: O = outside, TM = transmembrane region, I = inside, SP = signal
1221 peptide, as indicated by the bottom color legend. Mann-Whitney *U* test statistical significance:
1222 * *p*-value < 0.05, ** *p*-value < 0.01. Comparisons without statistical significance are not depicted
1223 in the figure.

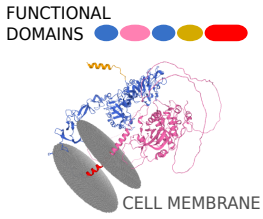
A

1 Identification of *HER2* splicing isoforms



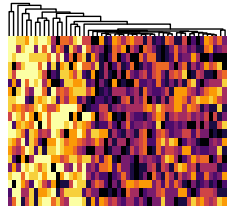
90 *HER2* coding isoforms (77 novel)

2 *In silico* characterization of HER2 proteins



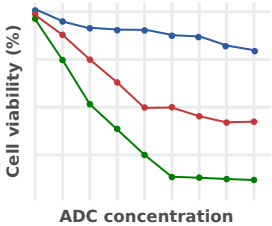
Distinct functional domains & cellular locations

3 *HER2* isoform expression profile by BC subtype



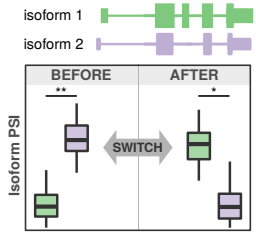
Clusters of *HER2* profiles by BC subtype

4 ADC sensitivity accross *HER2* isoform profiles



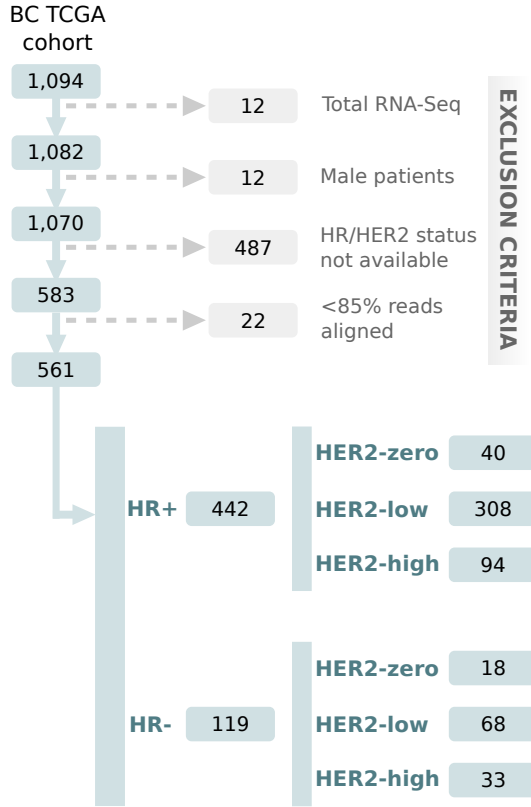
ADC sensitivity varies with *HER2* isoform expression profiles

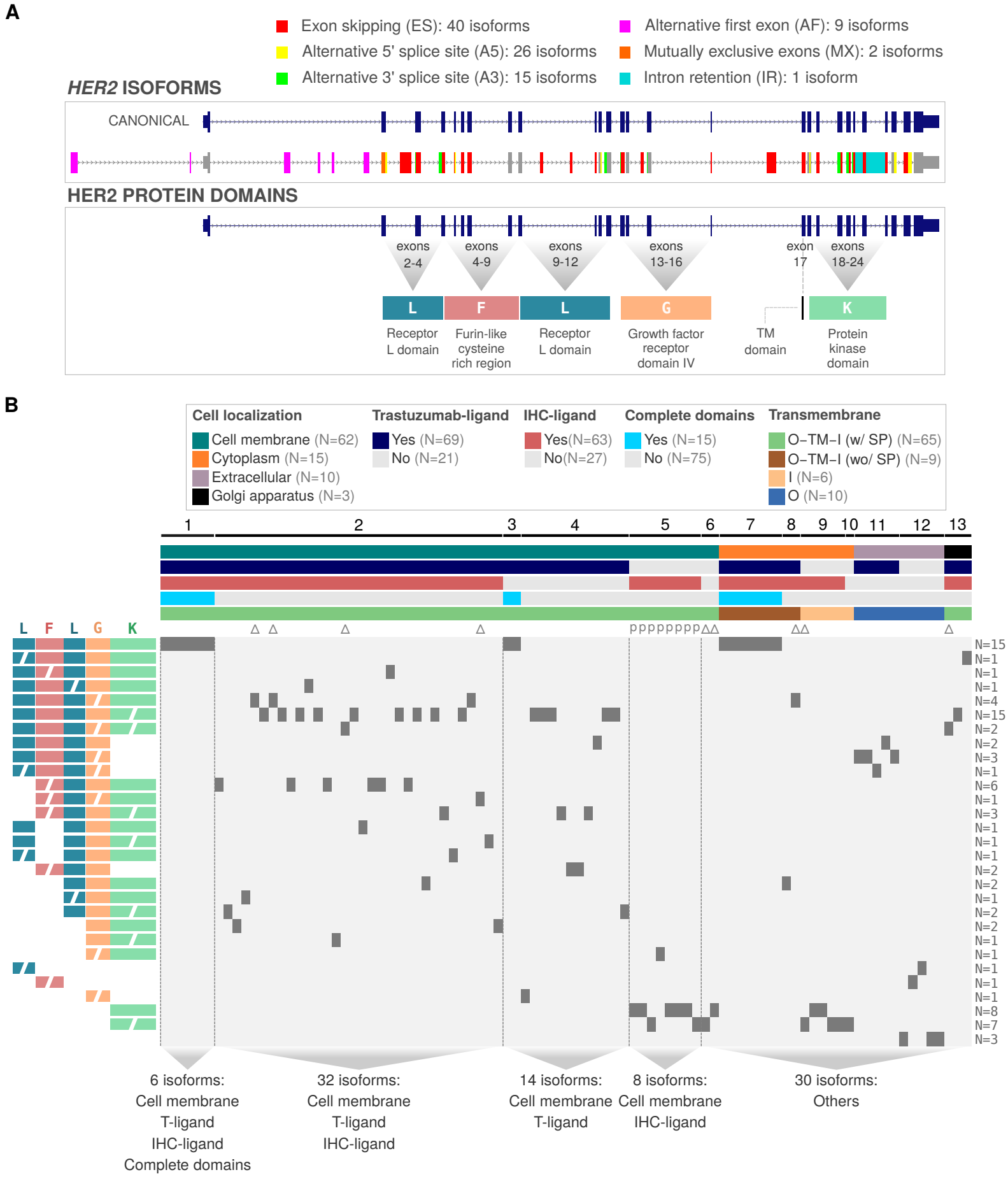
5 Culture cells treated with ADC inducing resistance

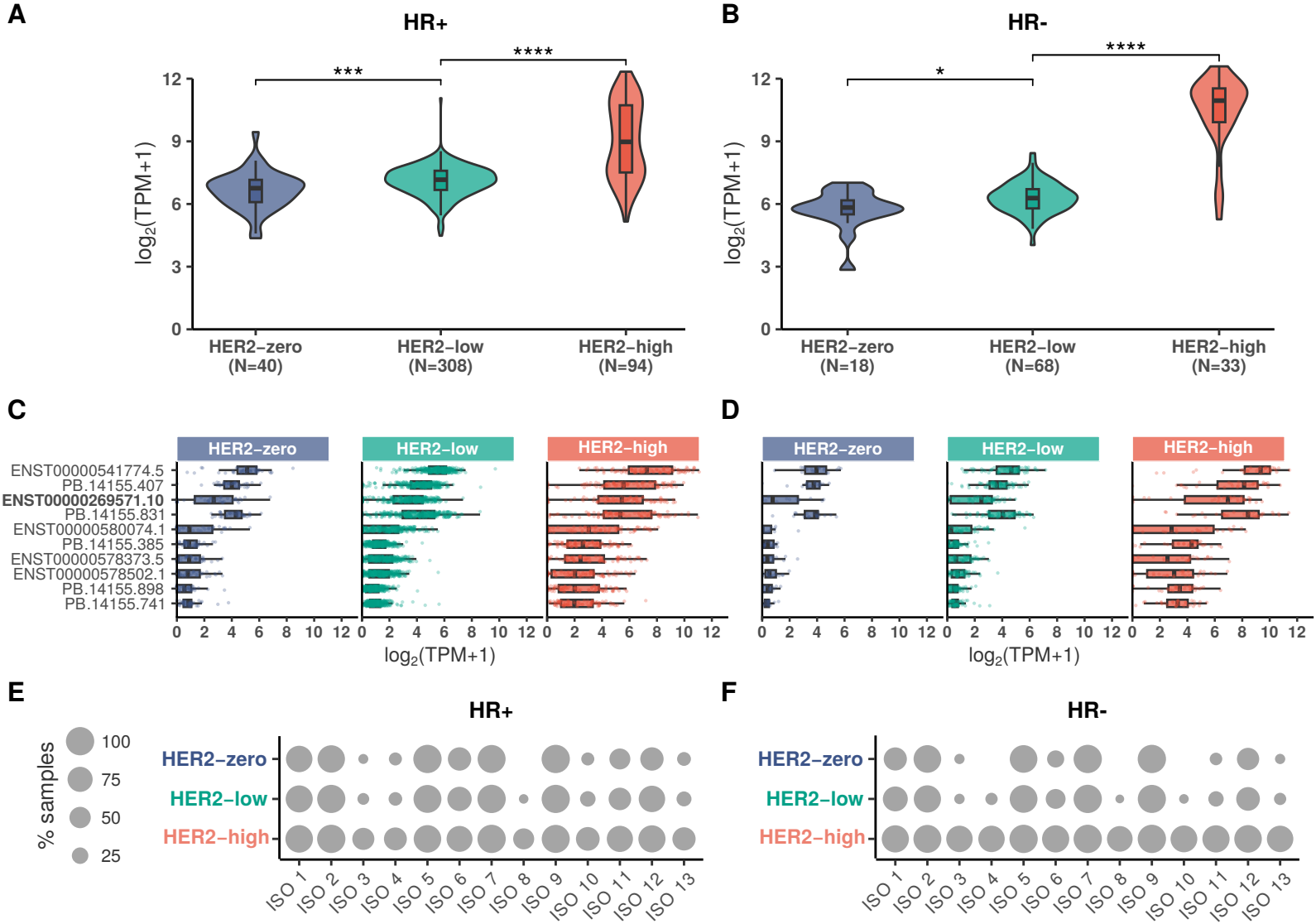


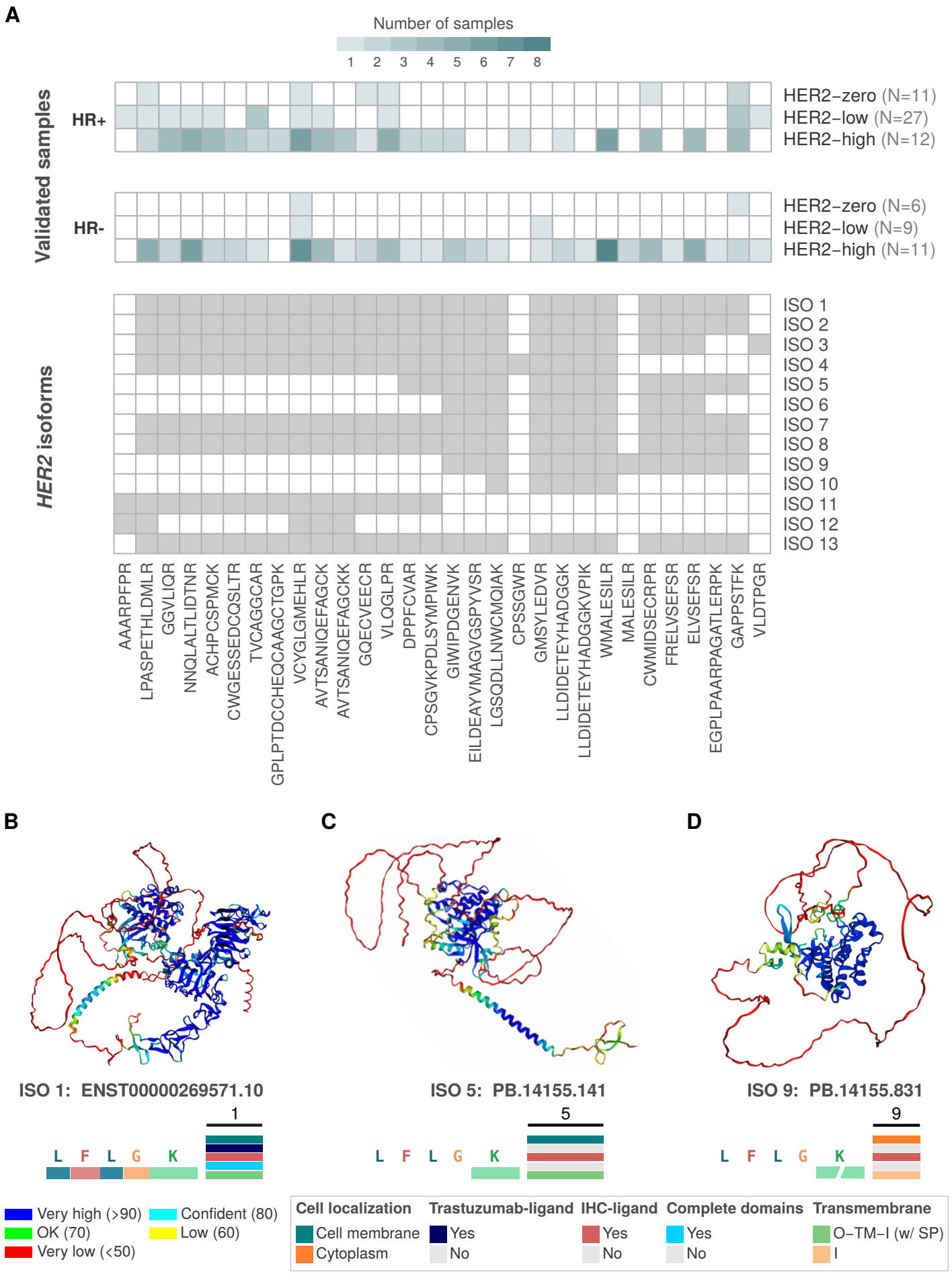
HER2 isoform switches underlie ADC resistance

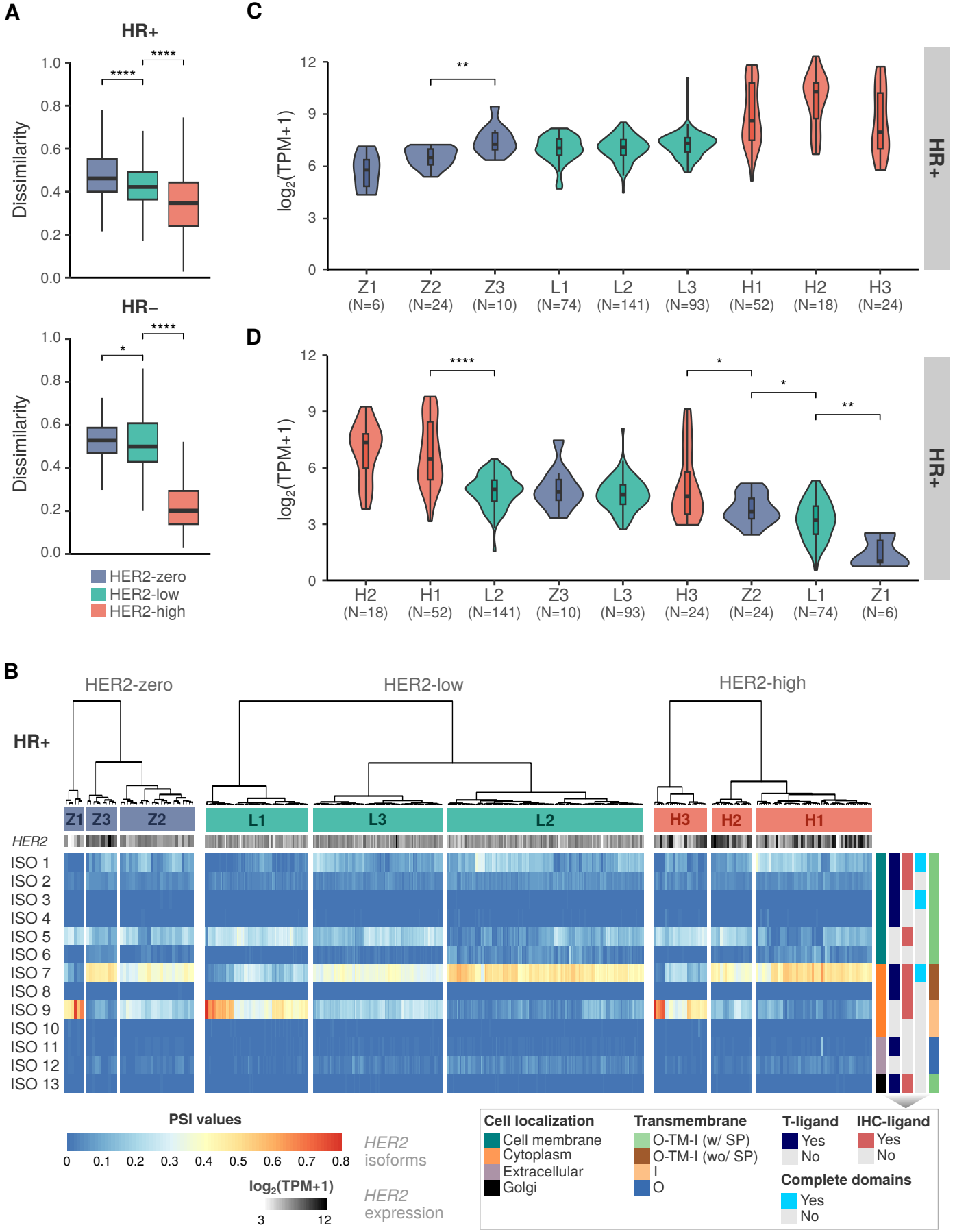
B

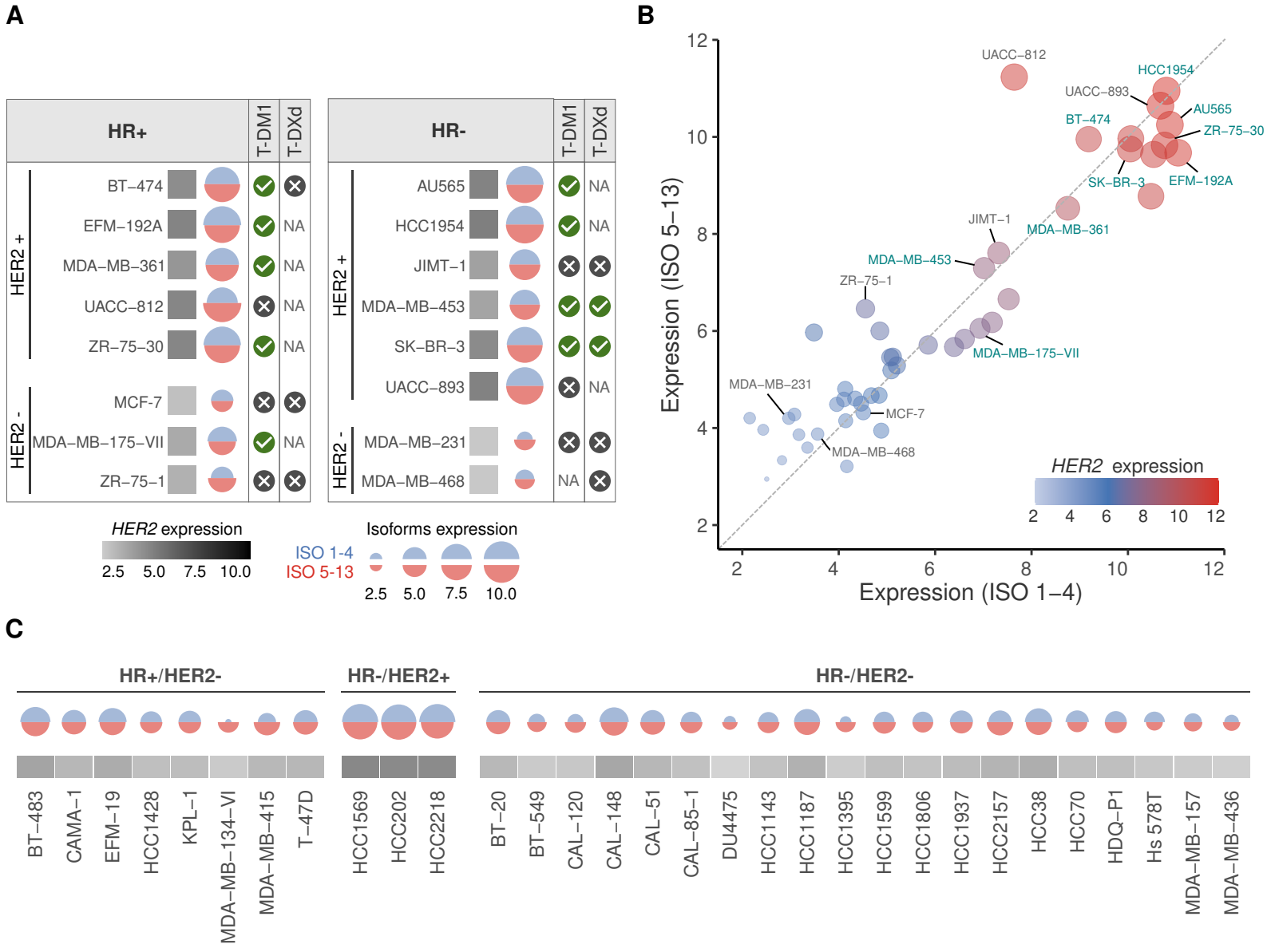


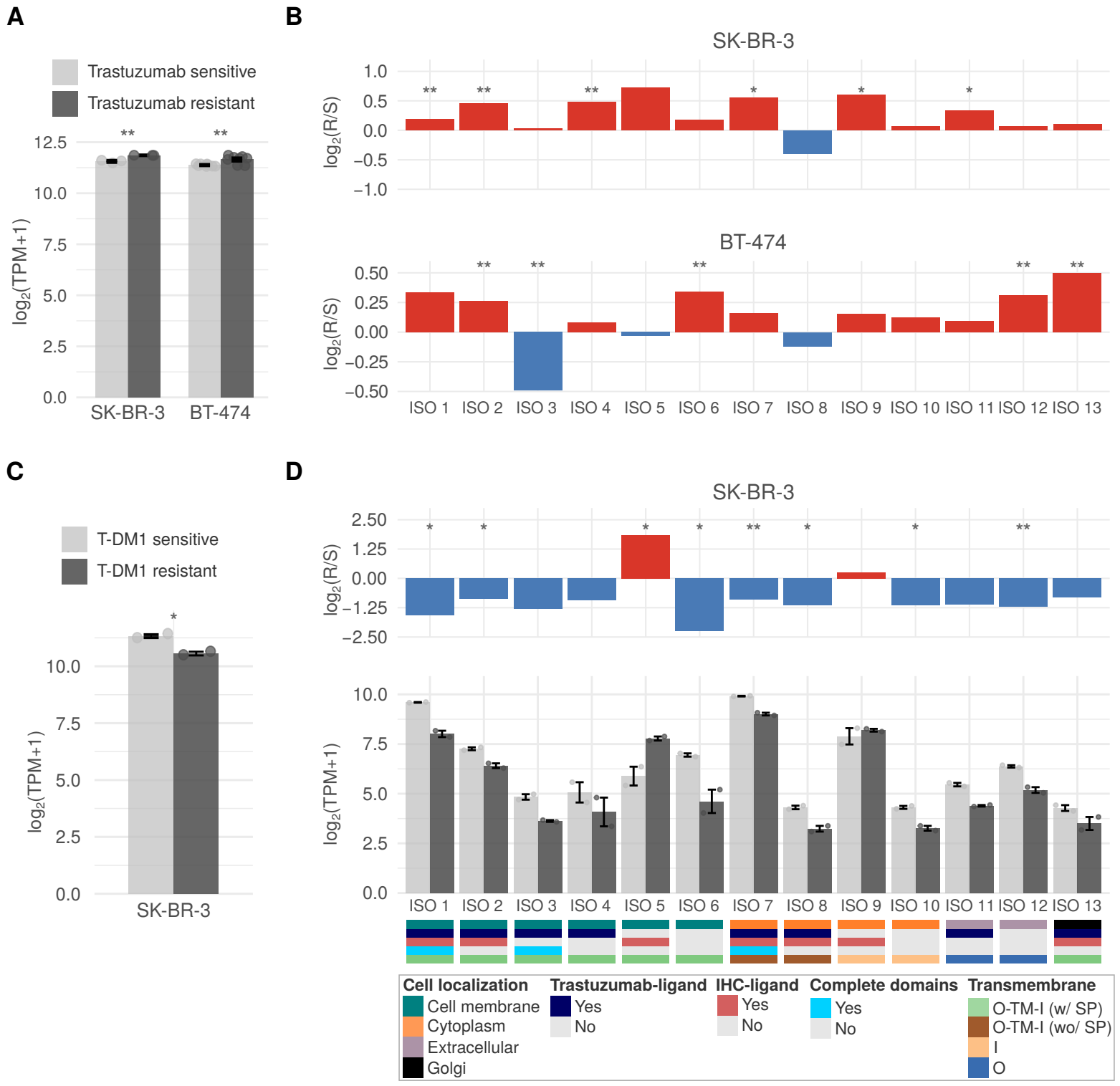

A
Exon skipping (ES): 40 isoforms
Alternative 5' splice site (A5): 26 isoforms
Alternative 3' splice site (A3): 15 isoforms
Alternative first exon (AF): 9 isoforms
Mutually exclusive exons (MX): 2 isoforms
Intron retention (IR): 1 isoform
HER2 ISOFORMS
CANONICAL
HER2 PROTEIN DOMAINS
exons 2-4
exons 4-9
exons 9-12
exons 13-16
exon 17
exons 18-24
L
F
L
G
K
Receptor L domain
Furin-like cysteine rich region
Receptor L domain
Growth factor receptor domain IV
TM domain
Protein kinase domain
B
Cell localization
Cell membrane (N=62)
Cytoplasm (N=15)
Extracellular (N=10)
Golgi apparatus (N=3)
Trastuzumab-ligand
Yes (N=69)
No (N=21)
IHC-ligand
Yes(N=63)
No(N=27)
Complete domains
Yes (N=15)
No (N=75)
Transmembrane
O−TM−I (w/ SP) (N=65)
O−TM−I (wo/ SP) (N=9)
I (N=6)
O (N=10)
1 2 3 4 5 6 7 8 9 10 11 12 13
L F L G K
N=15 N=1 N=1 N=1 N=4 N=15 N=2 N=2 N=3 N=1 N=6 N=1 N=3 N=1 N=1 N=1 N=2 N=2 N=1 N=2 N=2 N=1 N=1 N=1 N=1 N=1 N=8 N=7 N=3
6 isoforms: Cell membrane T-ligand IHC-ligand Complete domains
32 isoforms: Cell membrane T-ligand IHC-ligand
14 isoforms: Cell membrane T-ligand
8 isoforms: Cell membrane IHC-ligand
30 isoforms: Others


A
HR+
***
****
log2(TPM+1)
HER2−zero (N=40)
HER2−low (N=308)
HER2−high (N=94)
B
HR-
*
****
log2(TPM+1)
HER2−zero (N=18)
HER2−low (N=68)
HER2−high (N=33)
C
HER2−zero
HER2−low
HER2−high
ENST00000541774.5
PB.14155.407
ENST00000269571.10
PB.14155.831
ENST00000580074.1
PB.14155.385
ENST00000578373.5
ENST00000578502.1
PB.14155.898
PB.14155.741
log2(TPM+1)
D
HER2−zero
HER2−low
HER2−high
log2(TPM+1)
E
HR+
% samples
100
75
50
25
HER2−zero
HER2−low
HER2−high
ISO 1
ISO 2
ISO 3
ISO 4
ISO 5
ISO 6
ISO 7
ISO 8
ISO 9
ISO 10
ISO 11
ISO 12
ISO 13
F
HR-
HER2−zero
HER2−low
HER2−high
ISO 1
ISO 2
ISO 3
ISO 4
ISO 5
ISO 6
ISO 7
ISO 8
ISO 9
ISO 10
ISO 11
ISO 12
ISO 13


A
Number of samples
1 2 3 4 5 6 7 8
Validated samples
HR+
HER2−zero (N=11)
HER2−low (N=27)
HER2−high (N=12)
HR-
HER2−zero (N=6)
HER2−low (N=9)
HER2−high (N=11)
HER2 isoforms
ISO 1
ISO 2
ISO 3
ISO 4
ISO 5
ISO 6
ISO 7
ISO 8
ISO 9
ISO 10
ISO 11
ISO 12
ISO 13
AAARPFPR
LPASPETHLDMLR
GGVLIQR
NNQLALTLIDTNR
ACHPCSPMCK
CWGESSEDCQSLTR
TVCAGGCAR
GPLPTDCCHEQCAAGCTGPK
VCYGLGMEHLR
AVTSANIQEFAGCK
AVTSANIQEFAGCKK
GQECVEECR
VLQGLPR
DPPFCVAR
CPSGVKPDLSYMPIWK
GIWIPDGENVK
EILDEAYVMAGVGSPYVSR
LGSQDLLNWCMQIAK
CPSSGWR
GMSYLEDVR
LLDIDETEYHADGGK
LLDIDETEYHADGGKVPIK
WMALESILR
MALESILR
CWMIDSECRPR
FRELVSEFSR
ELVSEFSR
EGPLPAARPAGATLERPK
GAPPSTFK
VLDTPGR
B
C
D
ISO 1: ENST00000269571.10
ISO 5: PB.14155.141
ISO 9: PB.14155.831
1
5
9
L F L G K
Very high (>90)
OK (70)
Very low (<50)
Confident (80)
Low (60)
Cell localization
Cell membrane
Cytoplasm
Trastuzumab-ligand
Yes
No
IHC-ligand
Yes
No
Complete domains
Yes
No
Transmembrane
O−TM−I (w/ SP)
I


A
HR+
Dissimilarity
1.0
0.8
0.6
0.4
0.2
0.0
****
****
HR−
*
****
HER2-zero
HER2-low
HER2-high
C
log2(TPM+1)
12
9
6
3
0
**
Z1 (N=6)
Z2 (N=24)
Z3 (N=10)
L1 (N=74)
L2 (N=141)
L3 (N=93)
H1 (N=52)
H2 (N=18)
H3 (N=24)
HR+
D
****
*
*
**
H2 (N=18)
H1 (N=52)
L2 (N=141)
Z3 (N=10)
L3 (N=93)
H3 (N=24)
Z2 (N=24)
L1 (N=74)
Z1 (N=6)
B
HER2-zero
HER2-low
HER2-high
HR+
Z1
Z3
Z2
L1
L3
L2
H3
H2
H1
HER2
ISO 1
ISO 2
ISO 3
ISO 4
ISO 5
ISO 6
ISO 7
ISO 8
ISO 9
ISO 10
ISO 11
ISO 12
ISO 13
PSI values
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
HER2 isoforms
log2(TPM+1)
3
12
HER2 expression
Cell localization
Cell membrane
Cytoplasm
Extracellular
Golgi
Transmembrane
O-TM-I (w/ SP)
O-TM-I (wo/ SP)
I
O
T-ligand
Yes
No
IHC-ligand
Yes
No
Complete domains
Yes
No


A
HR+
T-DM1
T-DXd
HER2 +
BT−474
EFM−192A
MDA−MB−361
UACC−812
ZR−75−30
HER2 -
MCF-7
MDA−MB−175−VII
ZR−75−1
HR-
AU565
HCC1954
JIMT−1
MDA−MB−453
SK−BR−3
UACC−893
MDA−MB−231
MDA−MB−468
NA
HER2 expression
2.5 5.0 7.5 10.0
Isoforms expression
ISO 1-4
ISO 5-13
2.5 5.0 7.5 10.0
B
Expression (ISO 5−13)
Expression (ISO 1−4)
2 4 6 8 10 12
UACC−812
HCC1954
UACC−893
AU565
BT−474
ZR−75−30
SK−BR−3
EFM−192A
JIMT−1
MDA−MB−361
MDA−MB−453
ZR−75−1
MDA−MB−175−VII
MDA−MB−231
MCF-7
MDA−MB−468
HER2 expression
2 4 6 8 10 12
C
HR+/HER2-
HR-/HER2+
HR-/HER2-
BT−483
CAMA−1
EFM−19
HCC1428
KPL−1
MDA−MB−134−VI
MDA−MB−415
T−47D
HCC1569
HCC202
HCC2218
BT−20
BT−549
CAL−120
CAL−148
CAL−51
CAL−85−1
DU4475
HCC1143
HCC1187
HCC1395
HCC1599
HCC1806
HCC1937
HCC2157
HCC38
HCC70
HDQ−P1
Hs 578T
MDA−MB−157
MDA−MB−436


A
Trastuzumab sensitive
Trastuzumab resistant
log2(TPM+1)
SK-BR-3
BT-474
B
SK-BR-3
BT-474
log2(R/S)
ISO 1
ISO 2
ISO 3
ISO 4
ISO 5
ISO 6
ISO 7
ISO 8
ISO 9
ISO 10
ISO 11
ISO 12
ISO 13
C
T-DM1 sensitive
T-DM1 resistant
D
Cell localization
Cell membrane
Cytoplasm
Extracellular
Golgi
Trastuzumab-ligand
Yes
No
IHC-ligand
Yes
No
Complete domains
Yes
No
Transmembrane
O-TM-I (w/ SP)
O-TM-I (wo/ SP)
I
O

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. published online July 15, 2025
Access the most recent version at doi:10.1101/gr.280304.124