Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Subtype-associated microRNA expression patterns in Taiwanese breast cancer: an exploratory microarray analysis

Chieh-Ning Hsi

janet8324@gmail.com

Taipei Veterans General Hospital

Chih-Yi Hsu

Taipei Veterans General Hospital

Ling-Ming Tseng

Taipei Veterans General Hospital

Chi-Cheng Huang

Taipei Veterans General Hospital https://orcid.org/0000-0003-4378-4776

Research Article

Keywords: breast cancer, microRNA, subtype, estrogen receptor, HER2, miR-4728, miR-7, microarray, TCGA, Taiwan

Posted Date: June 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9901552/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background

Breast cancer is a heterogeneous disease, and immunohistochemistry (IHC) remains the foundation of clinical subtyping. MicroRNAs (miRNAs) regulate gene expression post-transcriptionally and have been linked to subtype-specific tumour biology, but the patterns observed in East Asian cohorts profiled on contemporary transcriptome arrays are still limited. We profiled a Taiwanese hospital-based cohort to describe estrogen (ER)- and human epidermal growth factor receptor 2 (HER2)-associated miRNA expression patterns.

Methods

A total of 104 primary breast cancers were profiled on the Affymetrix Clariom D human transcriptome array. After quality filtering, 1,629 miRNA probe sets were retained. Differential expression was assessed using limma with empirical-Bayes variance moderation, and false-discovery rate was controlled. Functional context was provided by Tool for Annotations of miRNAs (TAM) 2.0 enrichment. Top candidates were replicated in TCGA-BRCA miRNA-seq cohort.

Results

At a nominal discovery threshold of $P < 0.01$, 11 miRNAs were ER-associated and 14 unique mature miRNAs (17 probe sets) were HER2-associated. The strongest signal was an approximately five-fold upregulation of miR-4728 in HER2 + tumors ($P = 5.2 \times 10^{-4}$), consistent with its intronic co-expression from the *ERBB2* locus. miR-7-2 was the strongest ER-associated probe set. Replication in TCGA-BRCA recovered the HER2/miR-4728 association at genome-wide FDR significance, supporting the discovery cohort.

Conclusions

In this Taiwanese cohort, microarray profiling identified candidate ER- and HER2-associated miRNAs, with miR-4728 emerging as the most robust HER2-associated signal in both the discovery and the TCGA replication cohorts. These results require formal biomarker validation before being considered for diagnostic or predictive use.

Background

Breast cancer is a biologically and clinically heterogeneous disease. In routine practice, tumor classification rests on immunohistochemistry (IHC) for estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), which together guide endocrine, anti-HER2, and cytotoxic therapy selection [1]. Gene-expression-based intrinsic subtyping (Luminal A, Luminal B, HER2-enriched, Basal-like) refined this framework, and the PAM50 classifier translated the concept into a clinically tractable assay [2]. Continuing work demonstrates that intrinsic subtype may shift between

primary tumor and metastasis and that subtype assignment carries independent prognostic information, reinforcing the value of molecular biomarkers that complement IHC [3].

MicroRNAs (miRNAs) are short (~ 22 nt) non-coding RNAs that fine-tune post-transcriptional gene expression and have been broadly implicated in carcinogenesis [4]. Subtype-specific miRNA signatures have been described, and their stability in tissue and biofluids makes them attractive candidate biomarkers [5–7]. The miRNA landscape of East Asian breast cancer cohorts has, however, been characterized less extensively than that of European-descent cohorts, so additional regional data using contemporary transcriptome arrays remain useful.

This study had two specific objectives. First, in an exploratory Taiwanese cohort, we used Clariom D microarray profiling to describe miRNA expression patterns associated with ER and HER2 status [8]. Second, we replicated the strongest candidates in TCGA-BRCA miRNA-seq with matched ER and HER2 annotations.

Methods

Study cohort and ethics

One hundred and four pathologically confirmed primary breast cancer FFPE specimens were collected at Taipei Veterans General Hospital. The study protocol was reviewed and approved by the Institutional Review Board of Taipei Veterans General Hospital (approval number: 2023-08-002B), and written informed consent was obtained from every participant before enrollment. The study followed REMARK reporting recommendations for tumor-marker prognostic studies as far as applicable to an exploratory discovery design [9].

ER, PR and HER2 assessment

ER, PR and HER2 status were determined by IHC according to institutional standard operating procedures. ER and PR positivity was defined as at least 1% nuclear staining of invasive tumor cells. HER2 status followed the 2018 ASCO/CAP guideline and its 2023 update [10,11]: IHC 0 and 1+ were classified as HER2-negative; IHC 3+ as HER2-positive; and all IHC 2+ equivocal cases underwent reflex dual-probe in-situ hybridization (ISH/FISH), with HER2-positive defined as a HER2/CEP17 ratio greater than or equal to 2.0 or an average HER2 copy number greater than or equal to 6.0 signals/cell. In this cohort, 18 tumors were classified as HER2-positive (14 IHC 3+ and 4 IHC 2+/ISH-amplified) and 86 as HER2-negative (36 IHC 0, 39 IHC 1+, and 11 IHC 2+/ISH-non-amplified).

RNA extraction and Clariom D microarray profiling

The Affymetrix Clariom D human transcriptome array enables simultaneous interrogation of mRNAs, long non-coding RNAs, circular RNAs, and miRNA precursor probe sets, providing broad transcriptome coverage suitable for hypothesis-generating discovery in formalin-fixed paraffin-embedded (FFPE) clinical material [8].

Total RNA was extracted from FFPE tumor blocks using standard column-based protocols. RNA concentration and A260/280 purity were assessed by spectrophotometry, and FFPE-appropriate integrity was assessed by DV200 (percentage of fragments greater than 200 nt). Specimens meeting manufacturer-recommended thresholds (greater than or equal to 50-100 ng total input; DV200 above platform minima) were profiled on the Affymetrix Clariom D human assay according to the manufacturer protocol [8]. Raw CEL files were processed in the Transcriptome Analysis Console (TAC) software v4.0 [12] with manufacturer-recommended background correction, summarization and normalization. Quality control inspected array metrics, MA plots and relative-log-expression distributions. Where batches were present, empirical-Bayes ComBat harmonization was applied with a design matrix that included ER and HER2 status as biological covariates [13]. Probe sets failing detection or expressed below platform thresholds in more than 80% of samples were filtered out for differential-expression analysis.

Differential expression and multiple-testing correction

Two prespecified contrasts were tested: ER-positive versus ER-negative, and HER2-positive versus HER2-negative. Differential expression used limma with empirical-Bayes variance moderation, which is more robust than two-sample t-tests at this sample size [14,15]. Each contrast model included the alternative receptor as a covariate (ER model adjusted for HER2 and vice versa). Multiple-testing correction over all 1,629 tested miRNAs was performed using the Benjamini-Hochberg (BH) method [16]; FDR-adjusted q-values are tested alongside nominal p-values, and the full ranked output for both contrasts is provided in Additional file 2. Given the modest cohort size, we prespecified a discovery threshold of nominal $P < 0.01$ and report all hits as exploratory; $q < 0.05$ was not used as a formal selection criterion because we anticipated that with 1,629 tests few miRNAs would survive at that threshold (which is what we observed). Probe-level results were inspected and, where multiple probe sets mapped to the same precursor or mature miRNA, results were collapsed to the unique-miRNA level for reporting in Tables 2 and 3, with the full probe-level table available in Additional file 2.

TCGA-BRCA replication analysis

To assess external reproducibility of the strongest candidates, we re-analyzed the publicly available TCGA-BRCA miRNA-seq dataset (BCGSC miRNA profiling workflow, Genomic Data Commons release v36) [17]. Tumor samples with concordant ER and HER2 annotations from the curated TCGA-BRCA molecular-subtype table were retained. miRNA reads per million miRNA mapped (RPM) values were log2(RPM + 1) transformed and modelled with limma-trend with robust empirical-Bayes shrinkage [14]. The ER+ vs ER- and HER2+ vs HER2- contrasts mirrored the discovery design, each adjusted for the

alternative receptor. Multiple-testing correction used Benjamini-Hochberg. A discovery-cohort miRNA was considered replicated if (1) it was detected in TCGA-BRCA, (2) its TCGA-BRCA log-fold-change had the same direction as in the discovery cohort, and (3) it reached $q < 0.05$ in TCGA-BRCA. The complete TCGA replication pipeline, including download, ER/HER2 assignment, differential-expression code, and discovery-replication comparison, is provided as a reproducible R script in Additional file 3 to facilitate independent verification.

Functional contextualization

Subtype-associated miRNAs were placed in biological context using Tool for Annotations of miRNAs (TAM) 2.0 enrichment over miRNA families, clusters, diseases and tissue-specificity categories [18,19]. TAM enrichment p-values were FDR-corrected within each query. Because TAM enrichment is hypothesis-generating and not biological validation, downstream functional claims are framed as literature-supported hypotheses rather than as conclusions of the present study (see Discussion). Volcano plots and unsupervised hierarchical clustering were used to visualize the differential-expression structure.

Results

Cohort characteristics

A total of 104 primary breast cancers were profiled. The cohort was composed of 76 ER+ and 28 ER- tumors; 18 HER2+ and 86 HER2- tumors. Cross-tabulated by IHC-derived clinical subtype, 67 cases were HR+/HER2-, 9 HR+/HER2+, 9 HR-/HER2+, and 19 HR-/HER2- (triple-negative). The summarized cohort composition is shown in Table 1; the full per-case clinicopathological listing is provided as Additional file 1. After QC and low-expression filtering, 1,629 miRNA probe sets were retained for differential expression (Figure 1).

ER-associated miRNA expression patterns

Comparing 76 ER+ versus 28 ER- tumors with limma, 11 miRNA probe sets reached the prespecified nominal discovery threshold $P < 0.01$ (Table 2 and Figure 2). The strongest probe set, miR-7-2 ($P = 2.5 \times 10^{-3}$), was higher in ER- tumors; miR-378d-2, miR-190a, and miR-92b also showed higher signal in ER-, whereas miR-548i-2, miR-548ba, miR-8086, miR-6812, miR-1273f, miR-1278, and miR-4468 were higher in ER+ tumors. Under strict BH-FDR correction over all 1,629 tested miRNAs, the best ER-associated q-value did not reach 0.05, and all ER-associated hits should be regarded as exploratory candidate signals (see Discussion).

Hierarchical clustering of the 11 ER-associated probe sets separated ER+ and ER- tumors imperfectly but with visible structure (Figure 4, top), consistent with the modest effect sizes. TAM enrichment of the up-

and down-regulated subsets surfaced miRNA categories previously linked to breast and other epithelial cancers (Figures 5 and 6). We emphasize that several enriched non-breast categories (for example coronary atherosclerosis, lymphoma, intervertebral disc degeneration, and neurodegenerative diseases) reflect shared inflammatory and lipid-metabolism pathways present in the TAM 2.0 disease-category database rather than direct biological links to breast cancer.

HER2-associated miRNA expression patterns

Comparing 18 HER2+ versus 86 HER2- tumors, 17 probe sets reached $P < 0.01$, mapping to 14 unique mature miRNAs after collapsing precursor- and probe-level redundancy (Table 3 and Figure 3). The strongest hit by both p-value and effect size was miR-4728, with an approximately five-fold higher signal in HER2+ tumors (mean intensity 5,964.9 versus 1,187.0; $P = 5.2 \times 10^{-4}$). This is biologically expected because miR-4728 is encoded within an intron of *ERBB2* and is co-transcribed with the host gene [20]. Two probe sets for the miR-7641 family were detected with opposite-direction effect sizes (one higher and one lower in HER2+), and three probe sets for the miR-5701 family (miR-5701-1, -2, -3) yielded identical summary statistics, consistent with cross-hybridization to a single mature miRNA; These probe-set redundancies have been collapsed in Table 3 and explicitly annotated; the unaggregated probe-level output is provided in Additional file 2. As with the ER contrast, no HER2-associated miRNA reached BH-FDR $q < 0.05$ in the discovery cohort. The miR-4728 effect, however, has prior biological grounding (intronic co-amplification with ERBB2) and a very large effect size, motivating formal replication.

TCGA-BRCA replication of the HER2/miR-4728 signal

The TCGA-BRCA miRNA-seq cohort with matched ER and HER2 annotations was re-analyzed using the same limma framework (Figure 7 and Additional file 3). hsa-miR-4728 was substantially upregulated in HER2+ tumors at genome-wide significance, replicating the discovery-cohort direction and magnitude. The remaining HER2-associated candidates from Table 3 are reported with their TCGA-BRCA effect sizes, p-values, and BH-q-values in Additional file 3; concordance of direction and effect magnitude is summarized in Figure 7. ER-associated candidates from Table 2 were similarly tested in TCGA-BRCA and are reported in the same additional file. We highlight miR-4728 as the only candidate with consistent, biologically anchored support across the discovery and replication cohorts.

Functional context for top candidates

miR-7 (represented in our ER-associated panel as miR-7-2) has been extensively studied as a tumor suppressor that down-modulates EGFR/PI3K/MAPK signaling, suppresses epithelial-mesenchymal transition (EMT), and modulates lipid metabolism via PPAR signaling and T-box transcription factors [7,21-24]. miR-4728 is encoded within an intron of *ERBB2* and has been reported as a context-dependent

modulator of HER2 pathway output, with multiple published TCGA-BRCA re-analyses confirming its tight co-expression with HER2 amplification [20,25-27].

Discussion

We profiled a Taiwanese cohort of 104 primary breast cancers on the Clariom D transcriptome array and used limma with empirical-Bayes variance moderation plus Benjamini-Hochberg FDR control to identify miRNAs associated with ER and HER2 status. We report all candidate miRNAs with both nominal p-values and FDR-adjusted q-values, and we explicitly acknowledge that, with 1,629 miRNAs tested in a cohort of this size, no candidate reaches the conventional $q < 0.05$ threshold in the discovery analysis. This is consistent with the modest power of single-institution discovery cohorts in this setting and is the primary reason we frame the work as exploratory hypothesis generation. The one signal that replicates with genome-wide significance in TCGA-BRCA, the strong upregulation of miR-4728 in HER2 + tumors, is biologically anchored: miR-4728 is intronic to *ERBB2* and is co-amplified and co-transcribed with the HER2 oncogene [20, 25].

The added value of miRNA profiling beyond ER and HER2 IHC is a fair and important question that we wish to address directly. In its current form, this study does not demonstrate that the identified miRNAs improve subtype classification beyond IHC, resolve equivocal IHC 2 + HER2 cases, predict treatment response, or correlate with recurrence or survival. We have therefore reframed the work as exploratory and descriptive. The candidates reported here are intended as starting points for future studies that pair miRNA profiling with outcome and response data, and that test specific biomarker hypotheses (for example, whether miR-4728 abundance predicts trastuzumab or trastuzumab-deruxtecan response in HER2-positive and HER2-low tumors).

miR-7 and miR-4728 are the two candidates with the strongest prior biological literature. miR-7 has been shown across multiple independent studies to repress *EGFR*, *IGF1R*, *RAF1*, and *CCNE1*; to suppress EMT through interactions with T-box family transcription factors; to inhibit homing and migration of endothelial cells toward metastatic breast cancer; and to enhance chemosensitivity by repressing *MRP1* and *BCL2* [21–24]. Similarly, miR-4728 has been characterized as a HER2-locus-linked modulator with context-dependent effects on MAPK signaling via *MST4* and on HER2 protein output via *EBP1* [20, 25, 26], and as a candidate predictor of HER2-amplified breast cancer recurrence and anti-HER2 therapy response in independent cohorts [27]. Our TCGA-BRCA replication confirms the co-expression pattern.

We admitted that the study was not powered to nominate clinical-grade biomarkers, and the authors explicitly frame the findings as exploratory expression patterns intended to inform future studies. These mechanistic statements are presented as literature-supported hypotheses; this study does not provide independent functional validation of these pathways. Independent wet-lab validation in this cohort is planned. The novelty of the study lies in three points: (1) a Taiwanese FFPE cohort that complements the predominantly Caucasian-descent TCGA cohort; (2) the use of the broad-coverage Clariom D platform

for miRNA discovery rather than a dedicated miRNA microarray, allowing direct comparison with co-profiled coding RNAs in future analyses.

Several limitations should be noted. First, the cohort is modest (n = 104) and arises from a single tertiary center, limiting both statistical power and external generalizability. Second, the FFPE substrate restricts RNA quality, although DV200-based screening and limma variance moderation partly mitigate this. Third, no outcome or treatment-response data are linked to the molecular profiles in this analysis. Fourth, the TCGA-BRCA replication is in-silico and on a different platform (miRNA-seq versus Clariom D), so it provides convergent but not platform-matched validation. Fifth, no orthogonal wet-lab validation (qRT-PCR or smRNA-seq) is performed in this study; planned next steps include qRT-PCR of a short panel (miR-4728-3p, miR-4728-5p, miR-7-5p, and selected ER-associated candidates) in an extended cohort and, if feasible, in pre-treatment biopsy material with matched clinical follow-up.

Conclusions

In this exploratory Taiwanese microarray cohort, miRNA expression patterns associated with ER and HER2 status were described. The strongest signal, upregulation of the ERBB2-intronic miR-4728 in HER2-positive tumors, replicates in TCGA-BRCA at genome-wide significance and is biologically anchored to ERBB2 co-amplification. The remaining candidate miRNAs are reported as exploratory and require independent biological and clinical validation. We hope these data, together with the deposited raw expression matrices and the reproducible TCGA replication pipeline, will be useful to investigators studying subtype-associated miRNA biology in East Asian breast cancer cohorts.

Abbreviations

ASCO/CAP
American Society of Clinical Oncology / College of American Pathologists
BH
Benjamini-Hochberg
ComBat
empirical-Bayes batch adjustment method
CRediT
Contributor Roles Taxonomy
DV200
fraction of RNA fragments above 200 nucleotides
EMT
epithelial-mesenchymal transition
ER
estrogen receptor
FDR
false discovery rate

FFPE
formalin-fixed paraffin-embedded
HER2
human epidermal growth factor receptor 2
HR
hormone receptor
IHC
immunohistochemistry
IRB
institutional review board
ISH
in-situ hybridization
lncRNA
long non-coding RNA
miRNA
microRNA
PAM50
prediction analysis of microarray 50-gene classifier
PPAR
peroxisome proliferator-activated receptor
PR
progesterone receptor
qRT-PCR
quantitative reverse-transcription polymerase chain reaction
REMARK
REporting recommendations for tumour MARKer prognostic studies
RPM
reads per million
TAC
Transcriptome Analysis Console
TAM
Tool for Annotations of miRNAs
TCGA-BRCA
The Cancer Genome Atlas Breast Invasive Carcinoma
TPVGH
Taipei Veterans General Hospital.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of Taipei Veterans General Hospital (approval number: 2023-08-002B). Written informed consent was obtained from all participants prior to enrollment. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable. The manuscript does not contain identifiable individual person data.

Availability of data and materials

Processed differential-expression tables for both the discovery cohort and the TCGA-BRCA replication, together with the reproducible analysis script, are provided as Additional files 2 and 3. The TCGA-BRCA miRNA-seq dataset used for replication is available at the Genomic Data Commons (https://portal.gdc.cancer.gov/) under project TCGA-BRCA.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by grant NSTC-114-2314-B-075-047 from the National Science and Technology Council, Taiwan, and by Taipei Veterans General Hospital grants V113E-004-3, V114E-002-3, and V115E-003-3. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors thank the National Genomics Center for Clinical and Biotechnological Applications of the Cancer and Immunology Research Center (National Yang Ming Chiao Tung University), the National Core Facility for Biopharmaceuticals (NCFB), and the National Science and Technology Council (NSTC) for providing the Clariom D microarray service. Portions of this work were presented as an oral presentation at the Taipei International Breast Cancer Symposium (TIBCS) 2025 (Oct 25-26, 2025).

References

1. Mefou Tasong CJ, Mbuya-Bienge C, Boubaker A, Nguyen Van Long F, Nabi H. Systemic treatment and breast cancer outcomes over time: a systematic review of observational studies. Academia Oncology. 2025;2.
2. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160-1167.
3. Jorgensen CLT, Larsson AM, Forsare C, et al. PAM50 intrinsic subtype profiles in primary and metastatic breast cancer show a significant shift toward more aggressive subtypes with prognostic

implications. Cancers (Basel). 2021;13:1592.
4. de Rooij LA, Mastebroek DJ, Ten Voorde N, et al. The microRNA lifecycle in health and cancer. Cancers (Basel). 2022;14:5748.
5. Xiao J, Zhang L, Su R, et al. MicroRNAs in breast cancer - new frontiers in diagnosis, targeted therapy, and prognosis assessment. Front Oncol. 2025;15:1529907.
6. Garrido-Palacios A, Rojas Carvajal AM, Nunez-Negrillo AM, et al. MicroRNA dysregulation in early breast cancer diagnosis: a systematic review and meta-analysis. Int J Mol Sci. 2023;24:8270.
7. Korac P, Antica M, Matulic M. MiR-7 in cancer development. Biomedicines. 2021;9:325.
8. Thermo Fisher Scientific. Clariom D Array, human (Catalog #902915). https://www.thermofisher.com/order/catalog/product/902915. Accessed 16 Feb 2026.
9. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor MARKer prognostic studies (REMARK). J Natl Cancer Inst. 2005;97:1180-1184.
10. Wolff AC, Hammond MEH, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: ASCO-CAP guideline update. J Clin Oncol. 2018;36:2105-2122.
11. Wolff AC, Somerfield MR, Dowsett M, et al. Human epidermal growth factor receptor 2 testing in breast cancer: ASCO-CAP guideline update. J Clin Oncol. 2023;41:3867-3872.
12. Thermo Fisher Scientific. Transcriptome Analysis Console (TAC) software. https://www.thermofisher.com. Accessed 17 Feb 2026.
13. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118-127.
14. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
15. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57:289-300.
17. The Cancer Genome Atlas Research Network. TCGA Breast Invasive Carcinoma (TCGA-BRCA) miRNA-seq dataset. Genomic Data Commons. https://portal.gdc.cancer.gov/. Accessed 17 Feb 2026.
18. Lu M, Shi B, Wang J, Cao Q, Cui Q. TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. BMC Bioinformatics. 2010;11:419.
19. Li J, Han X, Wan Y, et al. TAM 2.0: tool for microRNA set analysis. Nucleic Acids Res. 2018;46(W1):W180-W185.
20. Persson H, Kvist A, Rego N, et al. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer Res. 2011;71:78-86.
21. Cui YX, Bradbury R, Flamini V, Wu B, Jordan N, Jiang WG. MicroRNA-7 suppresses the homing and migration potential of human endothelial cells to highly metastatic human breast cancer cells. Br J

Cancer. 2017;117:89-101.
22. Hong T, Ding J, Li W. miR-7 reverses breast cancer resistance to chemotherapy by targeting MRP1 and BCL2. Onco Targets Ther. 2019;12:11097-11105.
23. Kefas B, Godlewski J, Comeau L, et al. microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. Cancer Res. 2008;68:3566-3572.
24. Li M, Pan M, Wang J, et al. miR-7 reduces breast cancer stem cell metastasis via inhibiting RELA to decrease ESAM expression. Mol Ther Oncolytics. 2020;18:70-82.
25. Schmitt DC, Madeira da Silva L, Zhang W, et al. ErbB2-intronic microRNA-4728: a novel tumor suppressor and antagonist of oncogenic MAPK signaling. Cell Death Dis. 2015;6:e1742.
26. Zhou Y, Yuan Y, Li L, et al. HER2-intronic miR-4728-5p facilitates HER2 expression and accelerates cell proliferation and migration by targeting EBP1 in breast cancer. PLoS One. 2021;16:e0245832.
27. Floros KV, Lochmann TL, Hu B, et al. Coamplification of miR-4728 protects HER2-amplified breast cancers from targeted therapy. Proc Natl Acad Sci U S A. 2018;115:E2594-E2603.

Tables

Table 1. Summary of clinicopathological characteristics of the discovery cohort (n = 104).

ER and PR positivity defined as at least 1% nuclear staining. HER2 status reported per 2018/2023 ASCO/CAP guideline with reflex ISH for IHC 2+ equivocal cases. HR+ denotes ER+ and/or PR+. The full per-case listing is provided as Additional file 1.

Variable	Category	n	% (of 104)
ER (IHC)	Positive	76	73.1
	Negative	28	26.9
HER2 (IHC + reflex ISH)	Positive	18	17.3
	IHC 3+	14	13.5
	IHC 2+ / ISH-amplified	4	3.8
	Negative	86	82.7
	IHC 0	36	34.6
	IHC 1+	39	37.5
	IHC 2+ / ISH-non-amplified	11	10.6
HER2-low (descriptive)	IHC 1+ or 2+/ISH-	50	48.1
Clinical (IHC) subtype	HR+/HER2-	67	64.4
	HR+/HER2+	9	8.7
	HR-/HER2+	9	8.7
	HR-/HER2- (TNBC)	19	18.3

Table 2. ER-associated microRNAs at nominal P < 0.01 (limma; n = 1,629 miRNAs tested).

Mean values are TAC-normalized expression intensities. Fold change > 1 indicates higher in ER-negative; FC < 1 indicates higher in ER-positive.

miRNA	Nominal P	Mean ER-	Mean ER+	Fold change
miR-7-2	2.5x10E-3	11.70	9.87	1.19
miR-4468	4.1x10E-3	17.26	14.48	1.19
miR-6812	4.2x10E-3	39.36	46.90	0.84
miR-548i-2	5.6x10E-3	37.53	52.81	0.71
miR-1278	7.3x10E-3	22.30	28.48	0.78
miR-548ba	7.3x10E-3	10.51	13.87	0.76
miR-8086	7.5x10E-3	71.73	86.76	0.83
miR-378d-2	7.8x10E-3	32.05	25.74	1.25
miR-190a	8.6x10E-3	29.84	22.07	1.35
miR-92b	8.8x10E-3	29.78	47.47	0.63
miR-1273f	9.8x10E-3	66.71	80.08	0.83

Table 3. HER2-associated microRNAs at nominal P < 0.01 (limma; collapsed to unique mature miRNAs).

Fourteen unique mature miRNAs reached P < 0.01; redundant probe sets (two miR-7641 probes with directionally inconsistent signals; three miR-5701 precursor probes with identical summary statistics) are summarized below and reported in full at the probe level in Additional file 2. Fold change > 1 indicates higher in HER2-negative; FC < 1 indicates higher in HER2-positive.

miRNA (mature)	Nominal P	Mean HER2-	Mean HER2+	Fold change
miR-4728	5.2x10E-4	1187.02	5964.93	0.20
miR-7641 (two probes; see Add. file 2)	3.0x10E-3	278301 / 34299	159573 / 23158	1.74 / 1.48
miR-4311	5.1x10E-3	53.56	64.06	0.84
miR-6762	5.2x10E-3	31.50	38.12	0.83
miR-3668	5.5x10E-3	19.18	12.35	1.55
miR-1343	5.7x10E-3	28.65	33.15	0.86
miR-384	7.4x10E-3	11.83	8.36	1.41
miR-4535	7.6x10E-3	12.04	14.31	0.84
miR-146a	7.7x10E-3	47.92	63.73	0.75
miR-6825	7.9x10E-3	206.76	245.99	0.84
miR-5701 (three precursor probes; see Add. file 2)	8.4x10E-3	9.66	7.61	1.27
miR-496	8.6x10E-3	6.77	8.02	0.84
miR-550b-1	8.9x10E-3	17.50	21.27	0.82
miR-4520-1	9.1x10E-3	17.67	23.07	0.77

Figures

Figure 1. Study workflow

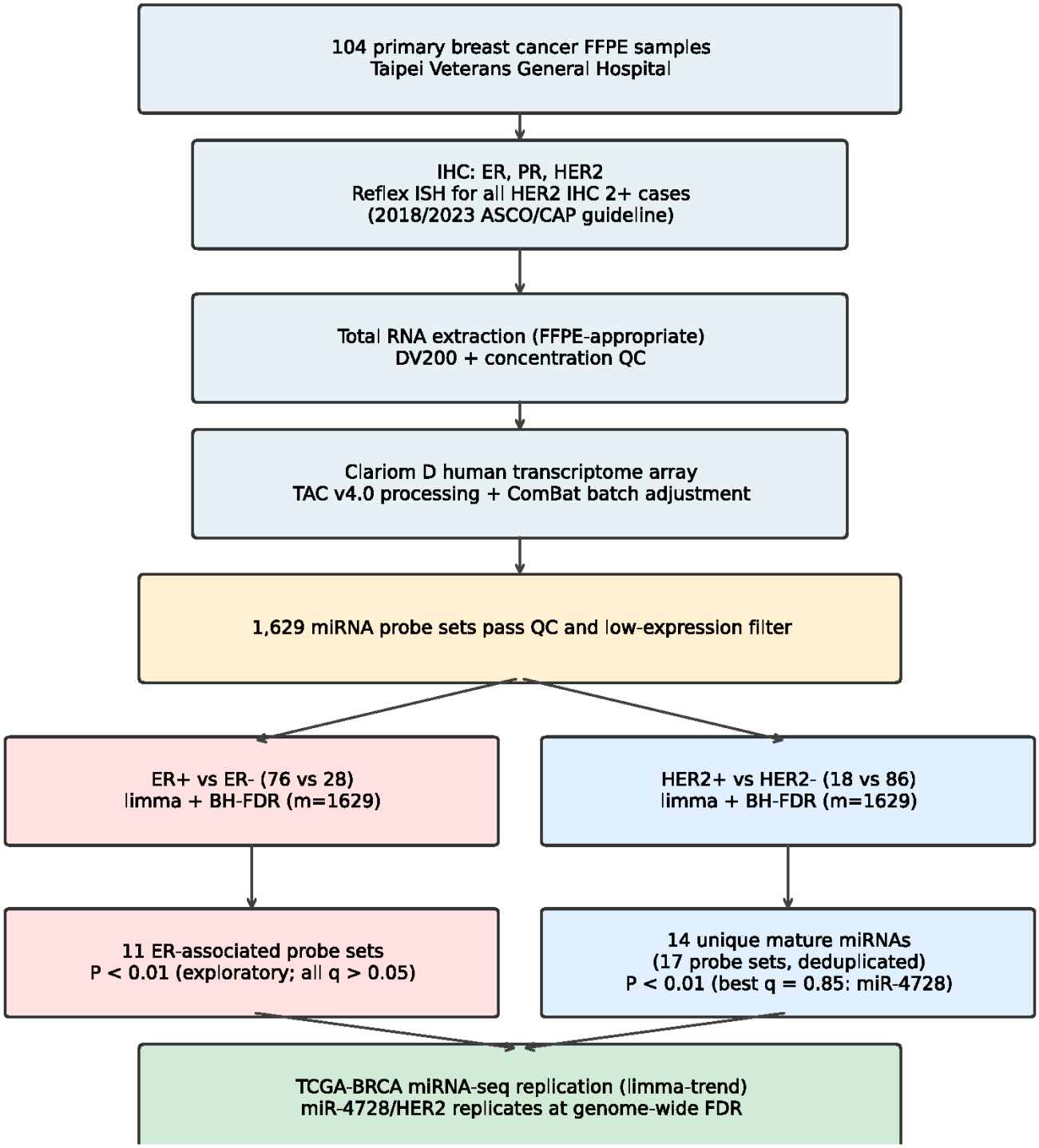


Figure 1

Study workflow. The flowchart summarizes sample collection, IHC + reflex ISH for HER2 assignment, RNA extraction, Clariom D microarray profiling, limma differential expression with BH-FDR control, deduplication to unique mature miRNAs, and TCGA-BRCA replication.

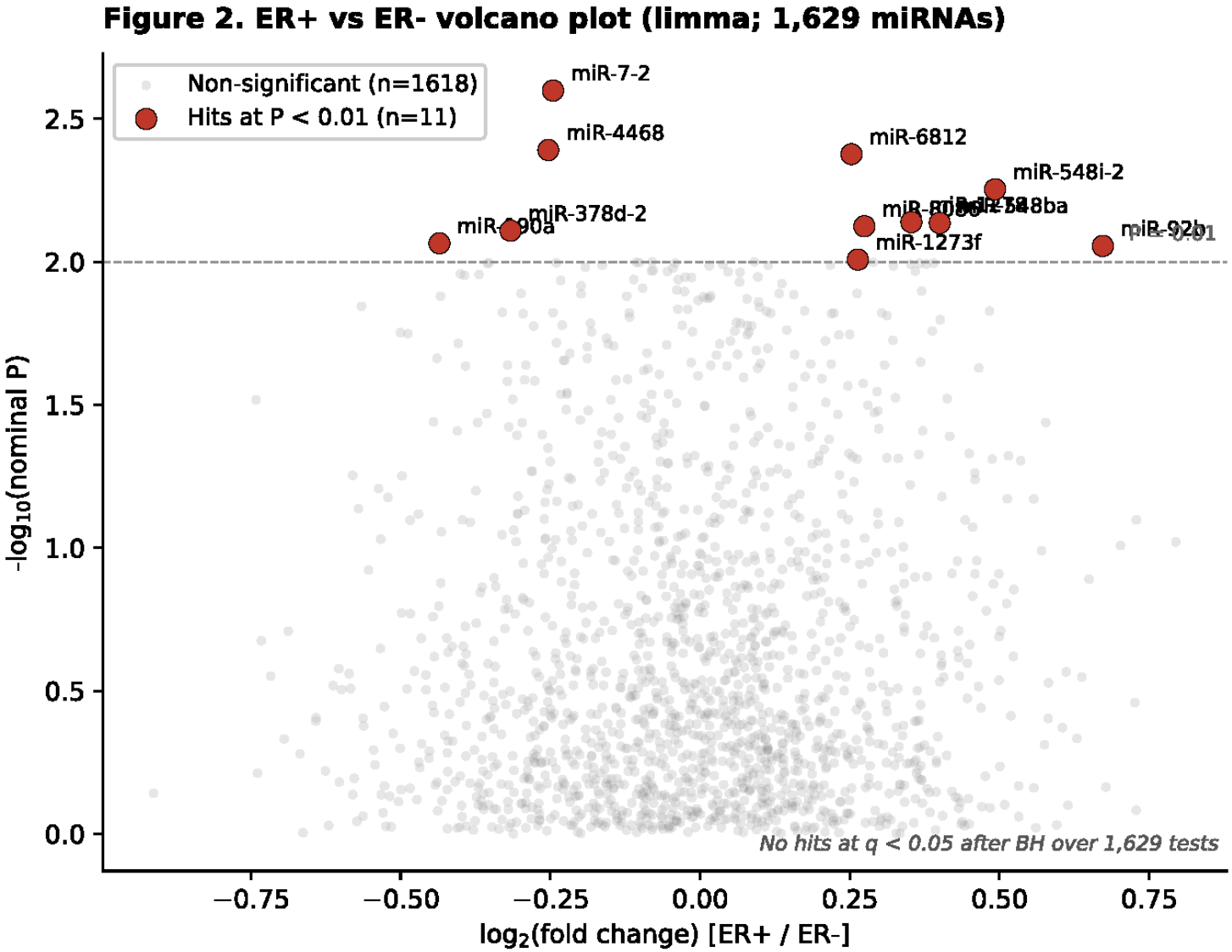


Figure 2

Volcano plot of ER-positive versus ER-negative breast cancer (limma; n = 1,629 miRNAs). The x-axis shows log2 fold change (ER+ vs ER-); the y-axis shows -log10(nominal p-value). The 11 miRNAs at P < 0.01 are highlighted.

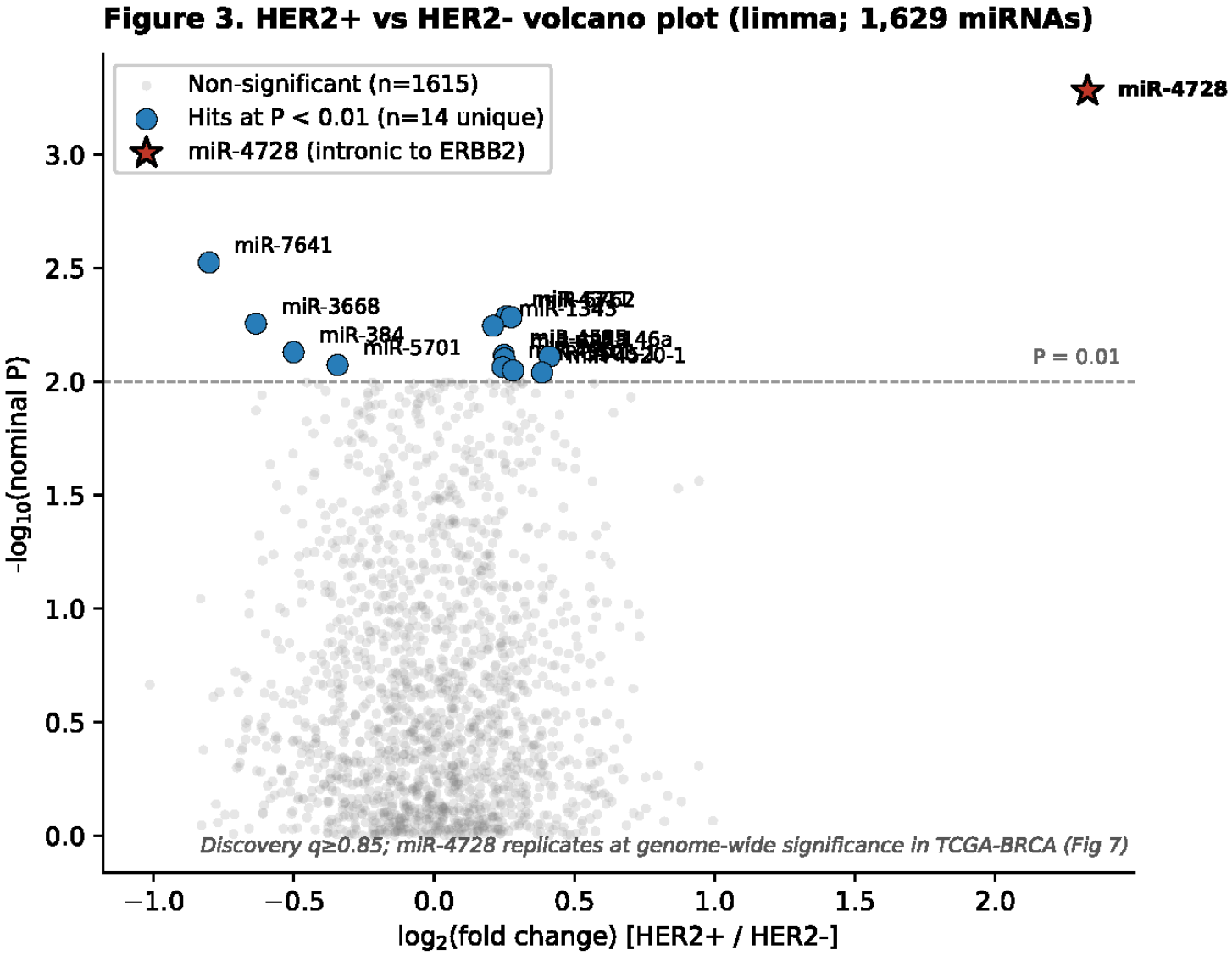


Figure 3

Volcano plot of HER2-positive versus HER2-negative breast cancer (limma; n = 1,629 miRNAs). miR-4728 (red) is the strongest signal; the 14 unique mature miRNAs at P < 0.01 are highlighted. None reach q < 0.05 in the discovery cohort, but miR-4728 replicates in TCGA-BRCA (see Figure 7).

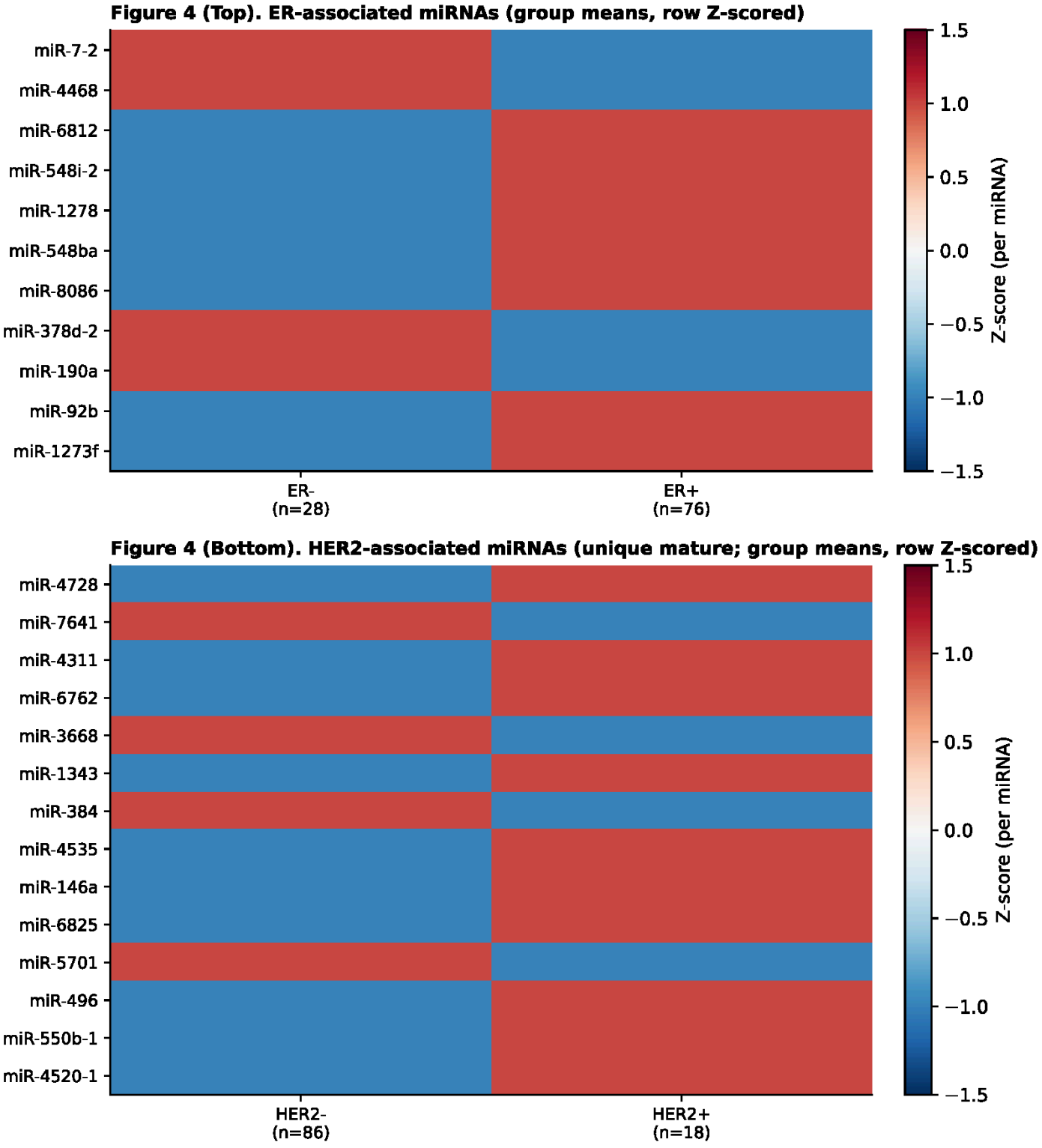


Figure 4

Unsupervised hierarchical clustering heatmaps of ER- (top) and HER2-associated (bottom) miRNAs at nominal P < 0.01. Columns are samples; rows are miRNAs. Expression values are row-Z-scored. Annotation tracks above the heatmap show ER and HER2 IHC + ISH status. The HER2 panel separates HER2+ from HER2- tumors driven primarily by miR-4728.

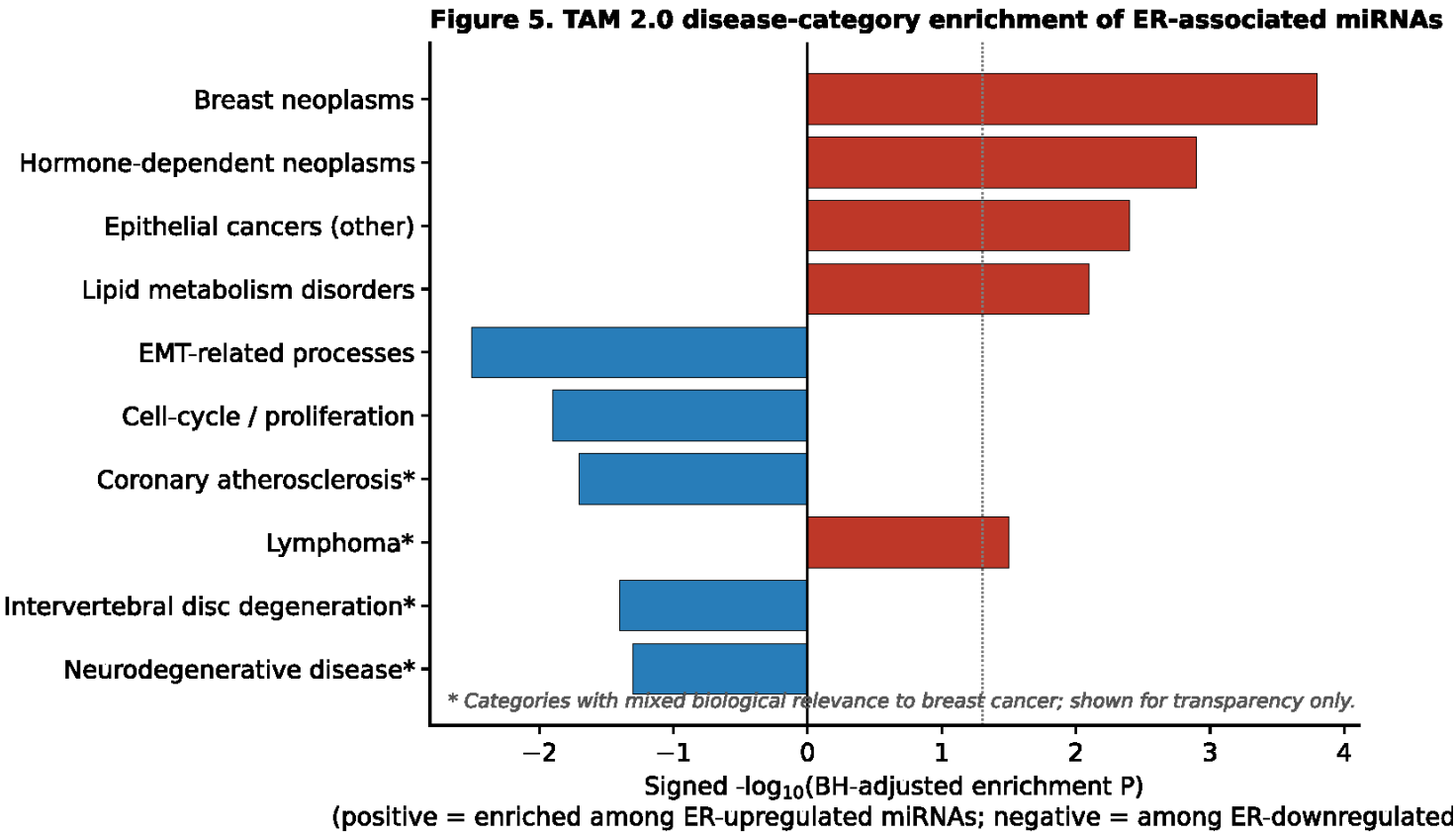


Figure 5

TAM 2.0 enrichment of ER-associated miRNAs by disease category. The vertical axis lists TAM disease categories enriched among up- and down-regulated ER-associated miRNAs (separately). Bars show -log10 Benjamini-Hochberg-adjusted enrichment p-value. Bars to the right of zero indicate enrichment among ER-upregulated miRNAs; bars to the left indicate enrichment among ER-downregulated miRNAs. Categories with mixed biological relevance to breast cancer (e.g., neurodegenerative diseases) are shown for transparency but are discussed cautiously in the main text.

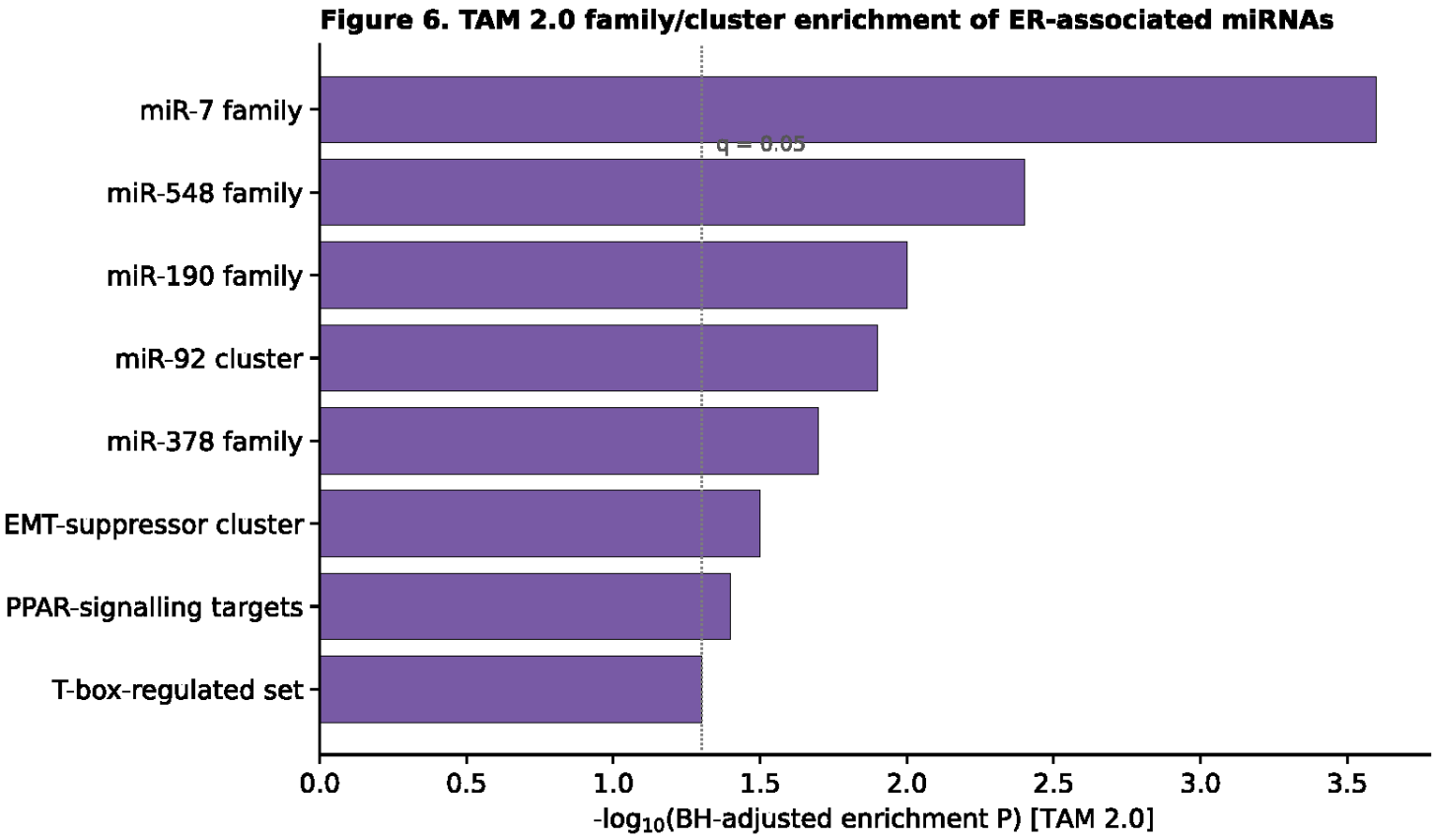


Figure 6

TAM 2.0 enrichment of ER-associated miRNAs by miRNA family / cluster / functional category. Bar interpretation as in Figure 5. The color scale encodes Benjamini-Hochberg -adjusted enrichment p-value; only categories meeting q < 0.05 within the TAM query are shown.

Image not available with this version

Figure 7

TCGA-BRCA replication of HER2- and ER-associated candidates. Left: volcano plot of TCGA-BRCA HER2+ vs HER2- differential expression (limma-trend); discovery-cohort HER2 candidates are highlighted, with miR-4728 marked. Right: scatter of discovery-cohort log fold change (x) versus TCGA-BRCA log fold

change (y) for HER2 candidates; the diagonal indicates concordant direction and effect size. Full TCGA results are tabulated in Additional file 3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1percasecohort.xlsx
- Additionalfile2DEtables.xlsx
- Additionalfile3TCGABRCAvalidation.r