

RESEARCH ARTICLE

# Identification of cell proliferation, immune response and cell migration as critical pathways in a prognostic signature for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer

Jeffrey C. Liu<sup>1\*</sup>, Miriam Zacksenhouse<sup>2</sup>, Andrea Eisen<sup>3</sup>, Sharon Nofech-Mozes<sup>3</sup>, Eldad Zacksenhaus<sup>1,4\*</sup>

**1** Division of Advanced Diagnostics, Toronto General Research Institute—University Health Network, Toronto, Ontario, Canada, **2** Brain-computer Interfaces for Rehabilitation Laboratory, Faculty of Mechanical Engineering, Technion—Israel Institute of Technology Haifa, Israel, **3** Sunnybrook Health Sciences Centre, University of Toronto; Toronto, Ontario, Canada, **4** Department of Medicine, University of Toronto, Toronto, Ontario, Canada

\* [jeff.liu@utoronto.ca](mailto:jeff.liu@utoronto.ca) (JCL); [eldad.zacksenhaus@utoronto.ca](mailto:eldad.zacksenhaus@utoronto.ca) (EZ)



**OPEN ACCESS**

**Citation:** Liu JC, Zacksenhouse M, Eisen A, Nofech-Mozes S, Zacksenhaus E (2017) Identification of cell proliferation, immune response and cell migration as critical pathways in a prognostic signature for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer. PLoS ONE 12(6): e0179223. <https://doi.org/10.1371/journal.pone.0179223>

**Editor:** Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

**Received:** November 21, 2016

**Accepted:** May 25, 2017

**Published:** June 20, 2017

**Copyright:** © 2017 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This study was conducted with support from the Canadian Breast Cancer Foundation (CBCF), Ontario, Canada to AE, SNM, and EZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

### Background

Multi-gene prognostic signatures derived from primary tumor biopsies can guide clinicians in designing an appropriate course of treatment. Identifying genes and pathways most essential to a signature performance may facilitate clinical application, provide insights into cancer progression, and uncover potentially new therapeutic targets. We previously developed a 17-gene prognostic signature (HTICS) for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer patients, using genes that are differentially expressed in tumor initiating cells (TICs) versus non-TICs from MMTV-Her2/neu mammary tumors. Here we probed the pathways and genes that underlie the prognostic power of HTICS.

### Methods

We used Leave-One Out, Data Combination Test, Gene Set Enrichment Analysis (GSEA), Correlation and Substitution analyses together with Receiver Operating Characteristic (ROC) and Kaplan-Meier survival analysis to identify critical biological pathways within HTICS. Publically available cohorts with gene expression and clinical outcome were used to assess prognosis. NanoString technology was used to detect gene expression in formalin-fixed paraffin embedded (FFPE) tissues.

### Results

We show that three major biological pathways: cell proliferation, immune response, and cell migration, drive the prognostic power of HTICS, which is further tuned by Homeostatic and Glycan metabolic signalling. A 6-gene minimal Core that retained a significant prognostic power, albeit less than HTICS, also comprised the proliferation/immune/migration

**Competing interests:** A U.S. Patent Application for HTICS has been filed. J. C. Liu & E. Zacksenhaus. Signature for Predicting Clinical Outcome in Human HER2+ Breast Cancer U.S. Patent Application No. 13/829,234, May 2013, for the purposes of transparency we have no other potential competing interests. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

pathways. Finally, we developed NanoString probes that could detect expression of HTICS genes and their substitutions in FFPE samples.

## Conclusion

Our results demonstrate that the prognostic power of a signature is driven by the biological processes it monitors, identify cell proliferation, immune response and cell migration as critical pathways for HER2<sup>+</sup>:ERα<sup>-</sup> cancer progression, and defines substitutes and Core genes that should facilitate clinical application of HTICS.

## Introduction

Breast cancer is a highly heterogeneous disease [1–6]. In the clinic, most (~65–70%) tumors are classified as estrogen receptor alpha positive (ERα<sup>+</sup>) that are treated with endocrine therapy [7]. About 20% is driven by amplification of the receptor tyrosine kinase HER2/ERBB2/NEU and either express ERα (HER2<sup>+</sup>:ERα<sup>+</sup>) or not (HER2<sup>+</sup>:ERα<sup>-</sup>). A third category comprises triple negative tumors (~10–15%), which do not express ERα, progesterone receptor or HER2. HER2<sup>+</sup> patients are treated with chemotherapy plus HER2 antagonists such as trastuzumab, trastuzumab-emtansine, pertuzumab or lapatinib [8–13]. Without HER2-directed therapy, HER2<sup>+</sup>:ERα<sup>-</sup> patients have the worst clinical outcome. Yet, some of these tumors do not progress to develop macro-metastases, and therefore surgical removal alone with local radiation or conventional therapy may suffice, at least as front-line therapy. Indeed, there is growing evidence that patients with small (T1), node negative HER2 positive tumors may profit from less aggressive therapies [14] [15]. In contrast, other tumors disseminate to form distal metastases that are virtually incurable, and should be treated most aggressively. Generating a prognostic signature from primary biopsies of these tumors may therefore help guide clinicians and patients as to the most appropriate treatment (Reviewed in Ref.[16]). Many such multi-gene based signatures have been developed for breast cancer and other malignancies [17–22]. Understanding the biology of these signatures and the basis for their prognostic power may provide further insight into processes that drive metastatic disease and mortality.

We previously reported on the development of a powerful prognostic signature for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer patients [23]. Our strategy was based on the recognition that Tumor Initiating Cells (TICs), which sustain growth following transplantation into recipient mice, may express genes that drive metastatic dissemination, colonization and growth at distal sites [24]. We therefore developed a prognostic signature for HER2<sup>+</sup>:ERα<sup>-</sup> patients using enriched TIC fraction from a mouse model for this subtype, MMTV-Her2/Neu [23, 25, 26]. Using both differentially up- and down- regulated genes between TIC-enriched and non-TIC fractions, we generated a 17-gene Her2-TIC-enriched signature (HTICS) that predicted clinical outcome on publicly available HER2<sup>+</sup>:ERα<sup>-</sup> cohorts, but not HER2<sup>+</sup>:ERα<sup>+</sup> patients [23]. When tested head-to-head, HTICS had superior prognostic power for HER2<sup>+</sup>:ERα<sup>-</sup> patients than a HER2-Derived Prognostic Predictor (HDPP) [27], Stroma-Derived Prognostic Predictor (SDPP) [19], IGS [18], mammaPrint [28] or a proliferation signature [29]. Furthermore, in multivariate analysis, HTICS was independent of multiple clinical variables [23]. It predicted overall survival (OS) for HER2<sup>+</sup>:ERα<sup>-</sup> patients with hazard ratio (HR) of 5.57 (P = 0.002), and metastatic free survival (MFS) with HR of 7.94 (P = 0.00084). Retrospective analysis on a small cohort of patients treated with trastuzumab (Herceptin) revealed that HTICS<sup>+</sup> HER2<sup>+</sup>:ERα<sup>-</sup> patients had a worse prognosis compared with HTICS<sup>-</sup> patients but benefited from trastuzumab therapy, and therefore should be prioritized for anti-HER2 therapy [23]. In contrast, HTICS<sup>-</sup> patients

exhibited good prognosis and may benefit from withholding anti-HER2 therapy as a frontline treatment.

We hypothesized that a prognostic signature monitors the combinatory effect of several biological pathways, and that understanding these pathways may shed light into the biology behind aggressive forms of a particular cancer subtype. The identification of a minimal “core” set of genes may also help define essential pathways. From a clinical perspective, a minimal set of genes may also be useful in generating immunohistochemical (IHC)-based prognosis. In addition, targeting these core genes or pathways may be therapeutically beneficial. Finally, it is important to identify substitutes that can be used to replace specific genes in a signature. If such substitutions are derived from the same functional pathways as the original signature genes, one can further establish the importance of the pathway. Such substitutions are also useful because most signatures are developed using RNA microarray data derived from fresh biopsies, whereas clinical biopsies are routinely processed as formalin-fixed paraffin embedded (FFPE) tissues, necessitating other modes of transcript detection such as NanoString technology [30]. Substituting genes may be required to compensate for limited technical performance of certain probes when a molecular signature acquired from frozen tissues is applied to another platform such as FFPE. This is especially critical for small signatures in which the contribution of each gene is highly important.

In the present study, we describe the identification of biological pathways monitored by HTICS and demonstrate that most genes capable of substituting for this signature share the same pathways. Expression of HTICS genes and their substitutes could be reliably detected in FFPE samples from breast cancer patients using NanoString technology, thus extending the clinical utility of HTICS. Through gene combination studies we have identified a core of 6 genes in HTICS and validated the effectiveness of this core using an independent RNA-Seq platform.

Together, our results demonstrate that a prognostic signature reflects critical biological pathways in a particular breast cancer subtype that dictate patient outcome. Specifically, we found that the prognostic power of HTICS lies in its ability to track cell proliferation, migration and immune-response pathways, which are critical for HER2<sup>+</sup>:ERα<sup>-</sup> tumorigenesis. Our analysis also identifies a core and substituting genes that would facilitate clinical implementation of HTICS. The critical role of the immune response, even within the 6 gene Core, for the prognostic power of HER2<sup>+</sup>:ERα<sup>-</sup> patients is particularly important given the accumulative evidence linking the immune system to the therapeutic effects of trastuzumab [31].

## Materials and methods

### Research ethics board (REB)

Experimental protocols with human breast cancer samples were approved by Sunnybrook Health Sciences Centre REB (PIN 418–2011), and were carried out in accordance with the approved guidelines.

### Amplicon analysis

HER2 IHC data from GSEs 24185, 22358, 25066, 2603, 2034, 5460, 21653, 26639, 19697, 17907, 16446, & 20194 (S1 Table) were pooled together and compared to expression levels of 5 HER2 gene amplicon (Her2/ErbB2, Stard3, Perld1, Grb7, C17orf37). Normalized gene expression values were used to determine fold changes against the median across the samples. Samples with greater than 2-fold elevation in 3 of 5 genes in the amplicon were considered HER2<sup>+</sup> (S1 Table). We note that at this cutoff (2-fold) we achieve the best balance of Positive Predictive Value (PPV) = 80.7% and sensitivity = 69.5%. Higher cutoff further enhance the PPV

(>90%) at the expense of sensitivity (total number of IHC HER2+ samples captured). ROC analysis was performed comparing HER2<sup>+</sup> status as determined by amplicon genes with IHC.

### Score for Signature Match (SSM) Classifier

For the comparison of effectiveness of our signatures, we used the Score for Signature Match (SSM) classifier to determine the status (signature match or no match) of each sample in the cohort [23]. In short, the SSM classifier takes into consideration 2 major aspects of the signature genes for calculation: 1) the rank of each gene among the samples in the cohort ( $X_n$ ). This is measured by median-centering the log2 transformed gene expression value, where a positive value is given to  $X_n$  if the gene is expressed at levels above the median of the cohort and a negative value if below; 2) the expected direction (up or down) and weight of the gene ( $I_n$ ). We gave every gene equal weight and therefore  $I_n = +1$  for Up-regulated signature genes and  $I_n = -1$  for Down-regulated signature genes. The score for each gene is the product of  $I_n$  and  $X_n$  ( $I_n \times X_n / |X_n|$ ). The formula for SSM is:

$$SSM = \frac{\sum \left( \frac{I_n X_n}{|X_n|} \right)}{\sum (|I_n|)}$$

The formula can be simplified as: Sum of Scores for Every Gene in Signature / Total Weight. If a gene is supposed to be up-regulated ( $I_n = 1$ ), a sample with positive expression above median ( $X_n > 0$ ) will be  $I_n X_n / |X_n| = 1$ . If a sample is below median expression ( $X_n < 0$ ) for the same gene,  $I_n X_n / |X_n| = -1$  leading to deduction of SSM score. Samples with  $SSM \geq 0$  were considered a match for the signature and used to compare with all other samples with  $SSM < 0$ .

### Receiver Operating Characteristic (ROC) analysis

The ROC curve represents the tradeoff between True-Positive (correct match) and False-Positive (incorrect match) rates as the threshold level of the classifier varies. The Area Under the ROC Curve (AUC) provides a standard measure of the discrimination power of the classifier, with 1/2 indicating no discrimination power and 1 indicating perfect discrimination. We used ROC and AUC to assess and compare the discrimination/prognostic power of the different signatures/combinations. ROC analysis was performed using algorithms developed by John Eng from the Russell H. Morgan Department of Radiology and Radiological Science at Johns Hopkins University School of Medicine (<http://www.rad.jhmi.edu/jeng/javarad/roc/JROCFITi.html>). In brief, the score of each sample for a given signature/combination was calculated by Score for Signature Match (SSM) algorithm and compared to a threshold to determine the match/no-match groups. The match/no-match class was compared with the survival events in the datasets. The survival events of patients were divided into 4 categories to reflect the degree of cancer aggressiveness: 1) disease-free patients after 36 months, the least aggressive; 2) disease-free patients within 36 months, not aggressive; 3) patients' disease progressed after 36 months, aggressive; 4) patients' disease progressed within 36 months, most aggressive.

### Human breast cancer patient samples and FFPE RNA extraction

Formalin-fixed paraffin-embedded breast cancer samples from the Department of Anatomic Pathology at Sunnybrook Hospital, Toronto, were used to generate 5- $\mu$ m thick tissue section slides. Tumors were classified to HER2<sup>-</sup>:ER $\alpha$ <sup>+</sup>, HER2<sup>+</sup>:ER $\alpha$ <sup>-</sup>, HER2<sup>+</sup>:ER $\alpha$ <sup>+</sup> and triple-negative using immunohistochemistry and FISH according to standard methodology. Sections were macrodissected to enrich the sample for invasive component, and de-waxed through sequential



10-minutes washes in xylene (3x), 100%, 90%, 75%, 50%, 30% Ethanol, and PBS. The de-waxed samples were then scrapped off by surgical scalpels and placed in 1.5 ml eppendorf tubes. RNA was extracted using High Pure FFPE RNA Micro Kit (Roche, cat# 04823125001).

## Human peripheral blood mononuclear cell, breast cancer cell lines and RNA extraction

Frozen human peripheral blood mononuclear cell (PBMC) were gifts from Dr. Shannon Dunn in the Department of Immunology, University of Toronto. MDA157, MDA231, MDA361, MDA436, MDA468, BT549, CAMA1, JIMT1. HCC3153 and HCC38 cells were cultured as described [22, 32]. The cells were plated in 10 cm plates and harvested at 80% confluence. RNA was extracted from cells using RNeasy Mini Kit (Qiagen, cat#74104).

## Datasets with clinical outcomes

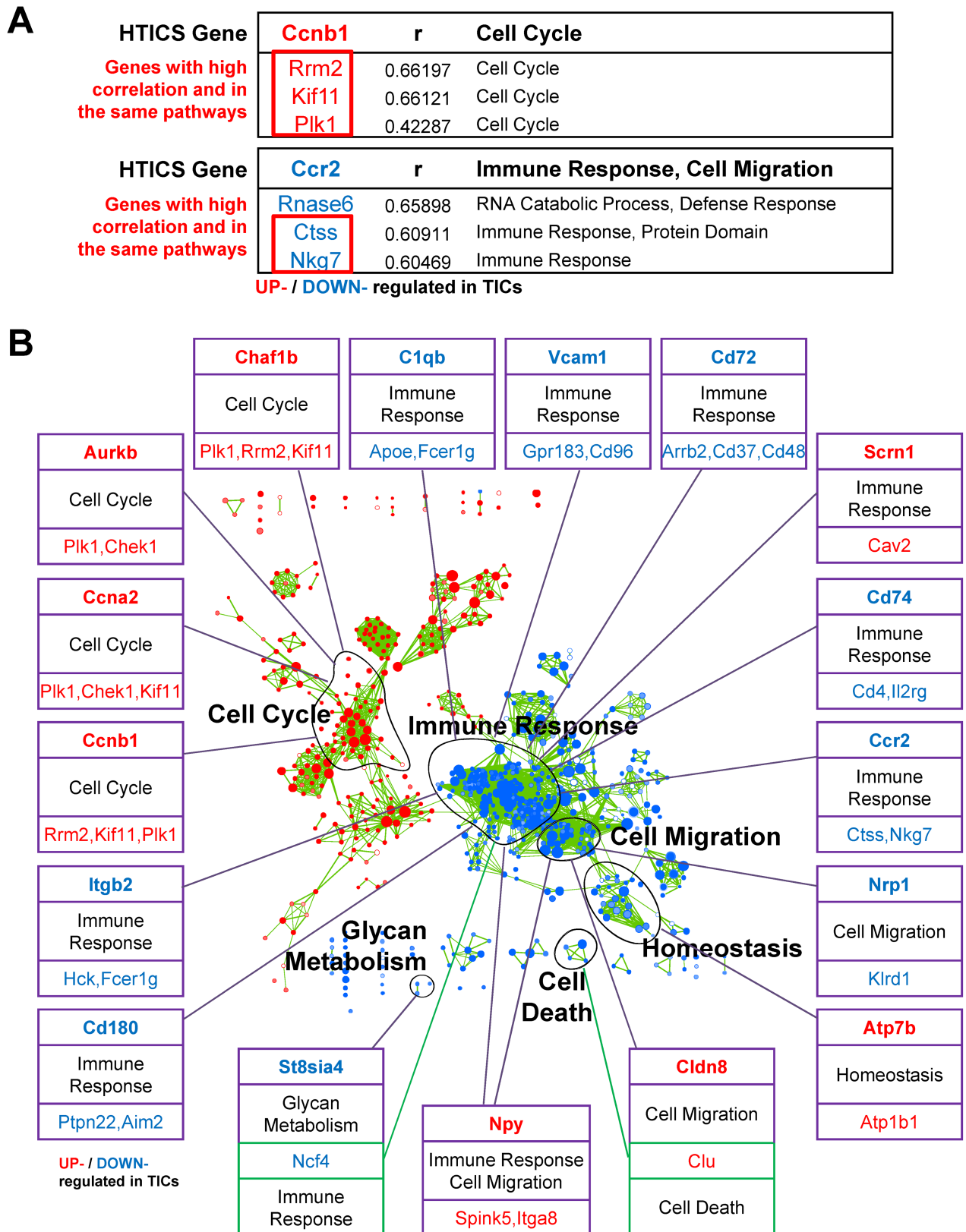
Published datasets were downloaded from GEO datasets at NCBI, NIH (<http://www.ncbi.nlm.nih.gov/gds/>). Datasets analyzed by Affymetrix microarray without Herceptin treatment (a potential confounding factor) but with clinical outcomes, included all breast cancer subtypes. They were used for Kaplan-Meier, ROC and random signature analyses (S2 Table). Datasets with Overall survival (OS) outcomes were: GSEs 16446 and 20685. Datasets with Metastasis-Free Survival (MFS) outcomes were: GSEs 2034, 2603, 5327, 6532, 11121, and 25066 (S3 Table).

## Pathway analysis

Microarray analysis with mouse tumor models was carried out using Affymetrix Mouse Gene 1.0 ST with 500 ng of total RNA isolated by double Trizol extractions (Centre for Applied Genomics, Hospital for Sick Children, Toronto). Microarray data were normalized using RMA method via Partek software and log<sub>2</sub>-transformed gene expression values were obtained. The data were analyzed by GSEA using paired t-test comparing gene expression values in the TIC and CD24- fractions, and parameters set to 2,000 gene-set permutations, gene-sets size between 15 and 500 [23]. The data is deposited in GEO Datasets with identifiers GSE29590 and GSE29616. An enrichment map (version 1.1 of Enrichment Map software) was generated using enriched gene-sets with a nominal p-value < 0.005, FDR < 1% and the overlap coefficient set to 0.5 [33, 34]. The databases included in the GSEA analyses were the Gene Ontology (GO), KEGG, PFAM, BIOCARTA and NCI databases. GO, PFAM and KEGG annotations were downloaded from Bioconductor (org.Mm.eg.db version 2.4.6, GO.db version 2.4.5, KEGG.db version 2.4.5). NCI annotations were downloaded from NCI website (<http://pid.nci.nih.gov/>, 2010-11-08) and BioCarta annotations downloaded from WhichGenes (2010-03-26). Only highly significant gene-sets (p < 0.005; FDR < 1%) are shown as nodes in the enrichment map. When two nodes have high degree of overlap in genes (overlap coefficient > 0.5), they are connected by a green line (edge). The major pathways labeled in Fig 1B are clusters of these significant gene-sets (nodes) by the Markov Cluster Algorithm.

## NanoString analysis

Total RNA from samples were diluted to 20 ng/μl concentration and at least 50 ng of RNA was sent for NanoString analysis (S4 Table). NanoString probes were designed according to Ref sequences with the following ID: AURKB (NM\_004217.2); CCNA2 (NM\_001237.2); SCRNI1 (NM\_001145513.1); NPY (NM\_000905.2); ATP7B (NM\_000053.2); CHAF1B (NM\_005441.2); CCNB1 (NM\_031966.2); CLDN8 (NM\_199328.2); NRPI (NM\_001024629.2); CCR2 (NM\_001123041.2); C1QB (NM\_000491.3); CD74 (NM\_001025159.1); VCAM1 (NM\_001078.3);



**Fig 1. Identification of Substitutes for HTICS Genes.** (A) Examples of replacement genes and corresponding functional pathways for Ccnb1 and Ccr2. (B) Replacement genes and corresponding pathways in relation to Enrichment maps for all HTICS genes. Note that 15/17 substitutes are from the same pathways as the original HTICS genes.

<https://doi.org/10.1371/journal.pone.0179223.g001>

CD180 (NM\_005582.2); ITGB2 (NM\_000211.2); CD72 (NM\_001782.2); ST8SIA4 (NM\_005668.4); Spink5 (NM\_006846); Ctss (NM\_004079); Aim2 (NM\_004833); CD48 (NM\_001778). The genes were normalized using the following controls: C4ORF8 (NM\_003704.3); CAPZB (NM\_001206540.1); EIF3A (NM\_003750.2); RTN3 (NM\_201429.1). The samples were processed and read by nCounter Analysis System from NanoString through UHN facilities (<http://www.uhnresearch.ca/facilities/microarray.htm>).

## Algorithms for random sets and combination analysis

To determine the probability that signatures/combinations could be the result of random chance, we generated 1000 random signatures with the same composition (i.e. same number of up- and down-regulated genes) and estimated their HRs using a formula provided by Kleinbaum and Klein (Chapter 3, Survival Analysis, 2005) with the same patients as indicated [35]. Algorithms for random sets as well as for combination analysis both written in R are available upon request.

## Additional statistical analysis

Parametric paired samples were analyzed by student t-test. Kaplan-Meier Survival analysis was performed using the PAST program (P.D. Ryan and Ø. Hammer, University of Oslo) and p-value was calculated using the Wilcoxon method. Hazard ratios were calculated by Cox Proportional Hazards Survival Regression. Pearson correlation studies were also performed using the PAST program. Differences between values were considered statistically significant at  $p < 0.05$ .

## Results

### ROC analysis of HER2 amplicon versus IHC for HER2<sup>+</sup> breast cancer classification

As a prelude to the forthcoming analysis, we assessed the assay we employ to call HER2<sup>+</sup> breast cancer samples. This assay, developed by Staaf *et al.*, measures expression levels of 5 genes around the HER2 amplicon (Her2/ErbB2, Stard3, Perld1, Grb7, C17orf37) directly from microarray data to predict HER2 status [23, 27]. The advantage of this approach is that the status of HER2, determined by IHC or FISH, is only available for a fraction of cohorts with mRNA profiling and clinical outcome. Moreover, methods and standardization for HER2 expression vary across clinical labs. To assess whether the amplicon strategy is as efficient as IHC in identifying HER2<sup>+</sup> breast cancer samples, we performed Receiver Operating Characteristic (ROC) analysis. This analysis calculates the discriminatory power of a classifier by describing the trade-off between True-Positives (correctly predicting HER2 status from amplicon expression as positive) and False-Positives (incorrectly predicting HER2 status as positive). For this analysis, we combined 11 cohorts containing a total of 1833 breast cancer samples with microarray and IHC data (although most have no clinical outcome, [S1 Table](#)). HER2 amplicon status (HER2 positive or negative for 5 amplicon gene expression) was then compared to HER2 IHC classification (positive or negative IHC; details in Method and Materials, [S1 Table](#)). Fisher's exact test demonstrated a significant detection of IHC HER2<sup>+</sup> samples by HER2 amplicon genes with 96% specificity ([S1 Table](#)). The HER2 amplicon accurately predicted HER2 IHC status of breast cancer patients with an Area Under Curve (AUC) value of 0.90 ([S1A Fig](#)). Together, the two methods have a Concordance Rate of 90.9% and Cohens Kappa coefficient of 0.69. Thus, in agreement with a recent publication [36], microarray

amplicon analysis is a reliable surrogate for HER2 status as determined by IHC, and can be used to improve consistency across datasets.

We previously demonstrated that HTICS is particularly effective in prognostication of HER2<sup>+</sup>:ER $\alpha$ <sup>-</sup> patients. This specificity is also seen by ROC analysis of metastasis-free (MFS) and overall (OS) survival data, having AUCs of 0.80 and 0.69, respectively, for HER2<sup>+</sup>:ER $\alpha$ <sup>-</sup> patients. In contrast, AUCs for all HER2<sup>+</sup> patients (MFS: 0.75; OS: 0.62) or HER2<sup>+</sup>:ER $\alpha$ <sup>+</sup> patients (MFS: 0.72; OS: 0.53) were much reduced (S1B Fig).

### Pathway analysis of HTICS

We generated HTICS by comparing gene expression profiles of enriched fractions of tumor initiating cells (TICs) versus non-TICs from mammary tumors isolated from a mouse model of HER2 breast cancer (MMTV-Her2/Neu) [23]. Originally 283 significantly and differentially expressed genes were identified and through training, described in reference [23], were narrowed down to 17 critical genes with strong prognostic power (as shown with MFS data, SIC Fig). We identified enriched biological pathways in TIC relative to non-TIC fractions using Gene-set-enrichment analysis (GSEA) and Enrichment Map (EM) via Cytoscape [23] (Table 1). This analysis suggests that HTICS monitors several pathways: Cell Cycle (Aurkb, Ccna2, Chaf1b, and Ccnb1, all up-regulated in TICs), Cell Migration (Npy, Cldn8, up-regulated; Nrp1, down-regulated), and Immune Response (Scrn1, up-regulated; C1qb, CD74, Ccr2, Itgb2, Vcam1, CD180, CD72, down-regulated). In addition, HTICS includes one Glycan Metabolism gene St8sia4 (down-regulated), involved in polysialic acid synthesis, and one homeostasis gene, Atp7b (up-regulated), required for efflux of copper out of the cells.

### Leave-one-out analysis of HTICS pathways

To assess the contribution of each pathway as a whole, we deleted each at a time and assessed the prognostic power (as measured by Hazard Ratio and AUC) of the remaining genes, using

**Table 1. HTICS genes and corresponding pathways as determined by GSEA.**

HTICS	Entrez ID	Name	Major Pathway
Aurkb	20877	Aurora Kinase B	<b>Cell Cycle</b>
Ccna2	12428	Cyclin A2	
Chaf1b	110749	Chromatin assembly factor 1, subunit B	
Ccnb1	268697	Cyclin B1	
Scrn1	69938	Secernin 1	<b>Immune Response</b>
C1qb	12260	Complement component 1, q subcomponent binding protein	
CD74	16149	CD74 molecule	
Ccr2	12772	Chemokine (C-C motif) receptor 2	
Itgb2	16414	Integrin, beta 2	
Vcam1	22329	Vascular cell adhesion molecule 1	
CD180	17079	CD180 molecule	
CD72	12517	CD72 molecule	<b>Cell Migration</b>
Npy	109648	Neuropeptide Y	
Cldn8	54420	Claudin 8	
Nrp1	18186	Neuropilin 1	<b>Homeostasis</b>
Atp7b	11979	ATPase, Cu <sup>++</sup> transporting, beta polypeptide	
St8sia4	20452	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 4	<b>Glycan Metabolism</b>

UP- / DOWN- regulated in TICs

<https://doi.org/10.1371/journal.pone.0179223.t001>

combined data from 12 publically available datasets with microarray gene expression profiles and clinical outcome [23]. We found that leaving out each of the five pathways diminished the prognostic power of HTICS, albeit to different extents, indicating that all pathways contribute to its full prognostic power (Table 2). Most critical pathways were Cell cycle and Immune response, followed by Glycan metabolism, Homeostasis and Cell migration. Notably, by this analysis Glycan metabolism and Homeostasis (one gene each) are more critical than Cell migration (3 genes). However, the detailed analysis below reveals that Cell migration is more important than Glycan metabolism and Homeostasis, and that the three main pathways (Cell cycle, Immune response, Cell migration) are most significant, yet genes from all five pathways contribute to the prognostic capability of HTICS.

### Identification of substitutes for HTICS genes

Having demonstrated that HTICS monitors the activity of specific signaling pathways in HER2<sup>+</sup> breast cancer, we next asked whether genes in the same, rather than all other, pathways would preferentially substitute for HTICS genes. To identify such potential replacements, we initially performed correlation studies on each of the additional 266 TIC genes (i.e. from the original 283 differentially regulated genes in TICs, omitting the 17 HTICS genes) that are significantly regulated in enriched-TICs vs. non-TICs in MMTV-Her2/Neu tumors [23]. Specifically, for each of the 17 HTICS genes, a correlation analysis was performed to determine which of the 266 differentially regulated genes had the most similar expression pattern. Remarkably, 15 of the 17 top correlated genes shared the same pathways as the HTICS genes they could replace (Fig 1A). The only 2 exceptions were Cldn8 (Cell Migration), which was best replaced by Clu (Cell Death) and St8sia4 (Glycan Metabolism), which was best substituted by Ncf4 (Immune Response) (Fig 1B). Of the 15 genes with potential substitutes in the same pathway, two had the highest correlation with genes from different pathways: CCR2 (Immune Response) could be substituted by Rnase6 (RNA Catabolic Process), and Scrn1 (Immune Response) could be replaced by Mphosph6 (not associated with any identified pathway) (Fig 1; S2 Fig). Our observation that for most HTICS genes (15/17), the best substitutes share similar functional pathways underscores the importance of these pathways in predicting disease progression and survival of HER2<sup>+</sup>:ER<sup>-</sup> patients.

### Development of Nanostring probes for HTICS and detection in FFPE samples

To assess the clinical utility of HTICS, we designed Nanostring probes for the signature genes. As noted, HTICS was generated using RNA microarray data acquired from fresh/frozen tumor biopsies. However, in a clinical setting, the routine platform for detecting gene expression is formalin-fixed paraffin-embedded (FFPE) samples. RNA quality from FFPE tissues is poor, and differences in sample preparation may introduce artifacts that can render detection

**Table 2. Leave-one out analysis assessing the effect of each pathway on HTICS prognostic power.**

Pathway Removed	MFS	
	HR	AUC
<b>None</b>	<b>7.94</b>	<b>0.80</b>
<b>Cell Cycle</b>	<b>2.86</b>	<b>0.68</b>
Cell Migration	7.17	0.78
<b>Immunue Resp.</b>	<b>1.84</b>	<b>0.72</b>
Homeostasis	5.45	0.71
Glycan Metab.	5.13	0.71

<https://doi.org/10.1371/journal.pone.0179223.t002>

of some genes unreliable. Knowledge of signature pathways and substitutes could therefore enable the design of clinical prognostic kits with built-in backup genes and controls to expand the clinical utility of HTICS. To test this strategy, we employed NanoString technology, which can faithfully detect RNA expression from partially degraded and contaminated RNA because it does not use an amplification step [30]. To detect the same spliced RNA transcripts, we designed the 17 NanoString probes for the same regions in the mRNA as in the microarray platform used to generate HTICS. These NanoString probes detected expression of all 17 HTICS genes from FFPE tumor samples (Fig 2A). By comparing signals generated from human breast cancer cell lines and peripheral blood mononuclear cells (PBMC), we could determine the tissue from which signals of each HTICS gene originated. For example, *Cldn8* expression was predominantly detected in tumor epithelial cells, with very low signal in PBMC. In contrast, *CD74* was exclusively detected in PBMC (Fig 2A). *Vcam1* and *C1qb* were undetectable in both tumor epithelia and PBMC but readily detected in whole tumors, suggesting expression in other cell lineages/biological contexts such as endothelium/ coagulation (Fig 2A).

### Prognostic power of HTICS substitutes

To determine the ability of the potential replacements to predict clinical outcome, we designed NanoString probes for four substitutes (*Spink5*, *Ctss*, *Aim2* and *CD48*). These substitutes were selected because the original HTICS genes (*Npy*, *Ccr2*, *CD180* and *CD72*) gave low expression signals in FFPE sections (<100). The replacement genes exhibited comparable or even better expression (Fig 2A). Moreover, ROC analysis of the pattern of gene expression of each substitute compared to the original HTICS gene in all the 79 MFS patients revealed high AUCs of 0.83, 0.77, 0.80 and 0.82 when using the *Spink5*, *Ctss*, *Aim2* or *CD48* substitutes, respectively (Fig 2A and 2B). A modified HTICS signature (HTICSm) with all 4 genes substituted gave nearly identical score in predicting outcome as HTICS, with a correlation of 0.97 (Fig 2C). HTICSm and HTICS also segregated patients into the same risk groups with AUC of 0.99 via ROC analysis (Fig 2D). Thus, these substitutes should produce consistent results by the NanoString assay, and thus, may be included in a prognostic kit for HER<sup>2+</sup>:ER<sup>-</sup> patients.

To assess the prognostic power of HTICSm and HTICS in a non-biased way, we generated 1000 randomly selected sets of 17 genes and compared their prediction power to those of HTICS and HTICSm [35]. Both signatures were highly prognostic, ranking within the top 25 for HER<sup>2+</sup> patients and top 12 for HER<sup>2+</sup>:ER<sup>-</sup> patients (Fig 3A). Yet, HTICS performed consistently better than HTICSm and was ranked #1 for HER<sup>2+</sup>:ER<sup>-</sup> patients (Fig 3A). It had the highest prognostic power with hazard ratio (HR) of 7.94 for MFS and 5.57 for OS, as compared to HTICSm with HR of 5.20 for MFS, and HR of 4.01 for OS (Fig 3B).

When replacing all 17 HTICS genes with the substitutes, the alternative signature (AltHTICS) still made weak but significant prognosis for HER<sup>2+</sup>:ER<sup>-</sup> patients with HR of 2.83 for MFS (Fig 3C). In those cases where the highest correlated genes were from different pathways (*Rnase6* for *CCR2*; *Mphosph6* for *Scrn1*), substitution with other-pathway genes resulted in inconsistent (*Rnase6* in AltHTICsb), or insignificant (*Mphosph6* in AltHTICsc) prognosis (S2 Fig). Thus, while the substitutes are capable of significant prognostication, likely because they monitor similar pathways, they are not as effective as the original HTICS.

### Identifying HTICS core genes

To define a core gene set that is most critical within HTICS, we developed an algorithm in R to test the prognostic power of all 131,071 possible combinations of genes in HTICS using both MFS and OS cohorts. Of these, 15,417 gene combinations were found to produce significant prognosis ( $P < 0.05$ ) with an average signature length of 9.8 genes.



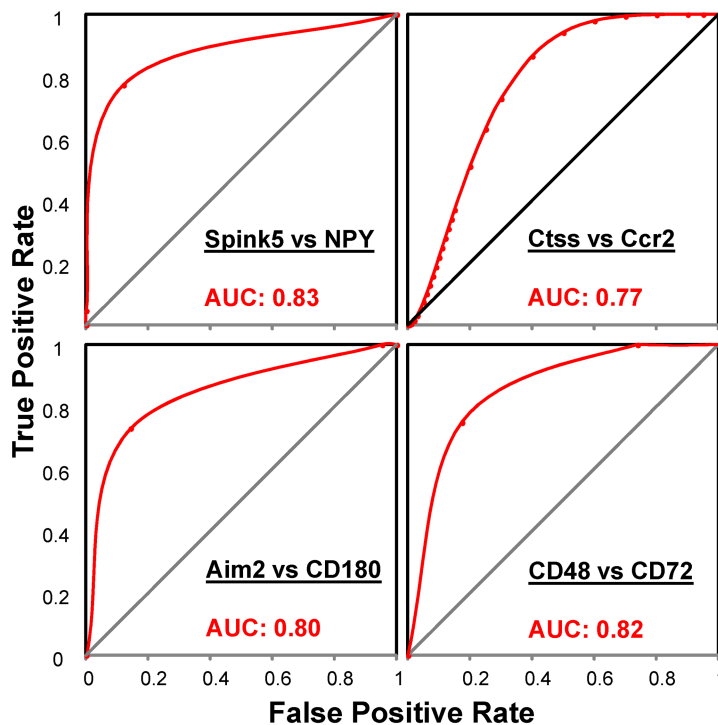
**A**

**Nanostring Probe Detection**

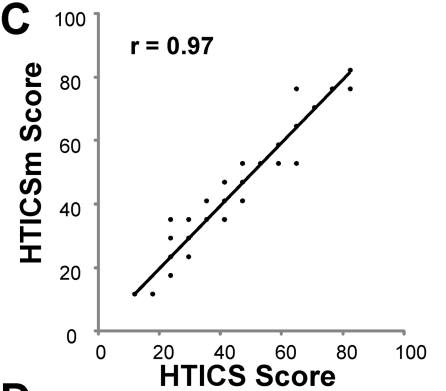
Background Controls		Fresh Sample			FFPE Samples	HTICSm Backup Genes Signal																															
		HTICS	Cell Lines	Human PBMC	Tumor																																
<table border="1"> <tr><td>Bkg.</td><td>Signal</td></tr> <tr><td>A</td><td>7.67</td></tr> <tr><td>B</td><td>3.11</td></tr> <tr><td>C</td><td>1.33</td></tr> <tr><td>D</td><td>1.44</td></tr> <tr><td>E</td><td>7.44</td></tr> <tr><td>F</td><td>6.56</td></tr> <tr><td>G</td><td>0.22</td></tr> <tr><td>H</td><td>1.22</td></tr> <tr><td>Ave</td><td>3.63</td></tr> <tr><td>S.D.</td><td>3.10</td></tr> <tr><td>95% CI up</td><td>10.13</td></tr> </table>	Bkg.	Signal	A	7.67	B	3.11	C	1.33	D	1.44	E	7.44	F	6.56	G	0.22	H	1.22	Ave	3.63	S.D.	3.10	95% CI up	10.13	AURKB	733.20	18.83	200.50	<table border="1"> <tr><td>Spink5</td><td>38.17</td></tr> <tr><td>Ctss</td><td>1300.83</td></tr> <tr><td>Aim2</td><td>69.94</td></tr> <tr><td>CD48</td><td>233.90</td></tr> </table>	Spink5	38.17	Ctss	1300.83	Aim2	69.94	CD48	233.90
	Bkg.	Signal																																			
	A	7.67																																			
	B	3.11																																			
	C	1.33																																			
	D	1.44																																			
	E	7.44																																			
	F	6.56																																			
	G	0.22																																			
	H	1.22																																			
	Ave	3.63																																			
	S.D.	3.10																																			
	95% CI up	10.13																																			
	Spink5	38.17																																			
	Ctss	1300.83																																			
	Aim2	69.94																																			
CD48	233.90																																				
CCNA2	923.80	19.67	313.31																																		
SCRN1	124.80	34.83	311.38																																		
NPY	53.80	0.83	33.58																																		
ATP7B	62.70	7.33	133.33																																		
CHAF1B	236.00	26.17	173.56																																		
CCNB1	2837.60	34.67	622.40																																		
CLDN8	68.80	7.67	263.63																																		
NRP1	266.20	44.50	1238.67																																		
CCR2	3.20	10.67	74.58																																		
C1QB	3.60	15.00	1547.63																																		
CD74	17.40	21815.00	21552.88																																		
VCAM1	1.10	1.67	205.79																																		
CD180	1.40	84.17	83.56																																		
ITGB2	28.00	4541.33	1076.54																																		
CD72	4.50	66.33	36.94																																		
ST8SIA4	17.50	343.50	236.96																																		

UP- / DOWN-regulated in TICs

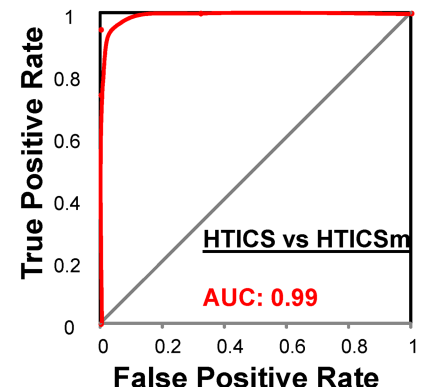
**B**



**C**



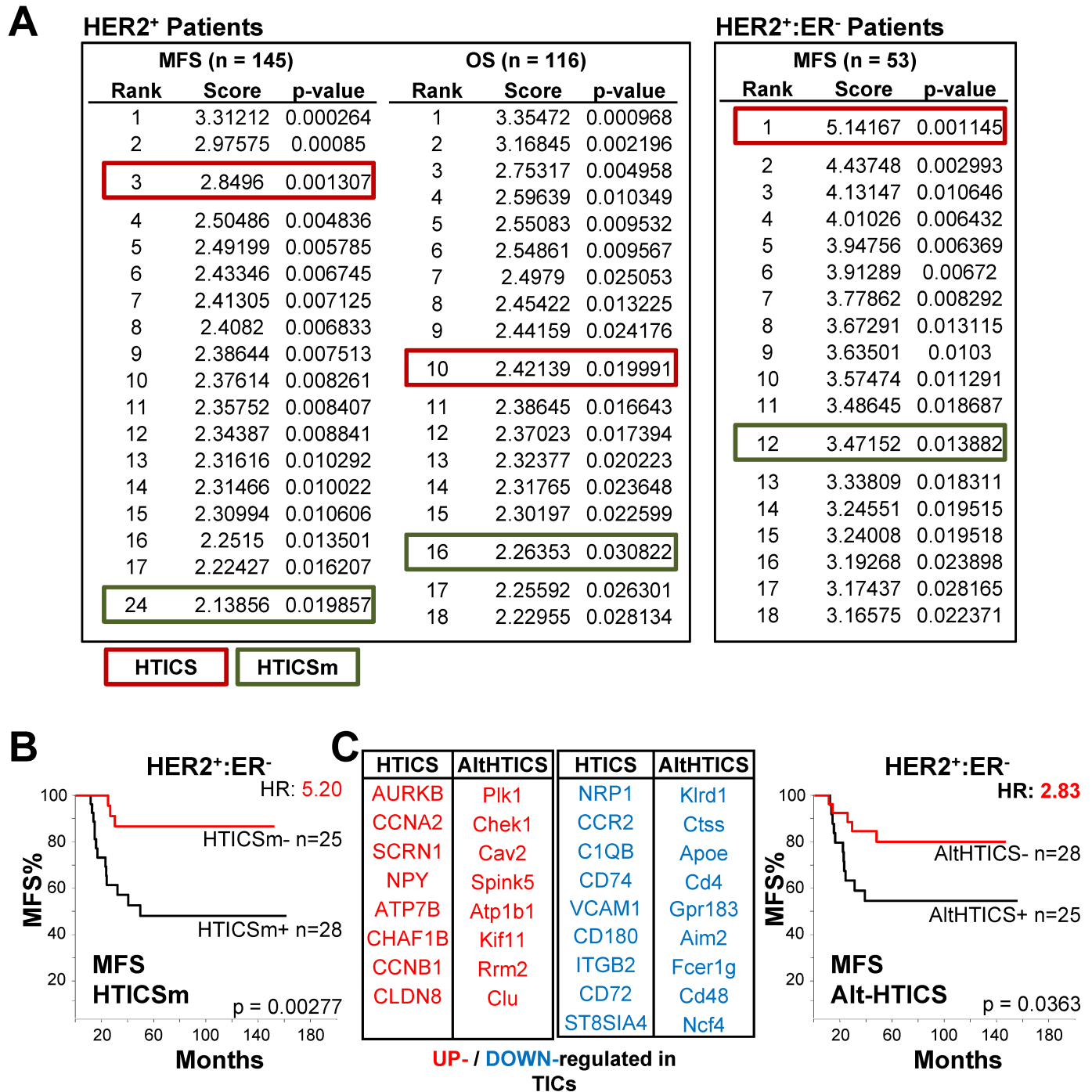
**D**



**Fig 2. Analysis of substitute genes versus HTICS.** (A) Reliable detection of gene expression of HTICS and substitutes in cell lines, PBMC and FFPE tumor samples using NanoString probes. (B) ROC analysis of individual substitute gene in comparison with

the original HTICS gene. (C) High degree of correlation ( $r = 0.97$ ) between HTICS and HTICSm. (D) ROC analysis demonstrating near perfect agreement between HTICS and HTICSm in segregating patients with AUC of 0.99.

<https://doi.org/10.1371/journal.pone.0179223.g002>



**Fig 3. Comparison to 1000 random signatures and survival analysis of HTICSm and AltHTICS.** (A) HTICS and HTICSm tested against 1000 randomly generated signatures of equal numbers of up- and down-regulated genes. (B) Kaplan-Meier analysis with HTICSm on MFS cohorts. (C) Prognostic power of AltHTICS with 17 gene substitutes, on MFS cohorts.

<https://doi.org/10.1371/journal.pone.0179223.g003>

We used two criteria to select for gene combinations with prognostic power. First, we asked how frequently a gene appears with other genes within the 15,417 significant prognostic signatures. For this, we performed a pair-wise analysis for each of the 17 HTICS genes to determine the percentage of overlaps with the other 16 genes. The average % of overlap was 39.52% with standard deviation (SD) of 7.5%. If the percentage overlap of a pair of genes was greater than 1 SD above the average, i.e. 47.02%, the two genes were considered to require each other for prognosis (Fig 4A). To further increase the stringency of selecting critical genes, we only considered those that paired with two or more other genes. 8 such genes highlighted in pink in Fig 5A were identified: Ccr2, Nrp1, Scrn1, Vcam1, CD74, Chaf1b, Npy and CD180, and were designated Core1 (Fig 4B). Core1 gave a HR of 4.53 for HER2<sup>+</sup>:ERα<sup>-</sup> patients (Fig 4C), and AUC of 0.61 with high false positive rate by ROC analysis (Fig 4D).

Interestingly, of the 15,417 gene combinations with significant prognosis, 92 (with average length of 11.1 genes), gave similar or even better score than HTICS. Composition of the top 10 signatures and corresponding HRs in OS and MFS cohorts are depicted in Fig 5A. The 92 signatures maintained the cell migration, cell cycle and immune-response pathways; a relatively high percentage of them also included the glycan metabolism gene St8sia4, but not the homeostasis gene, Atp7b (Fig 5A and 5B). At least 3 of the 5 HTICS pathways were found in 100% of these 92 highly prognostic sub-signatures (Fig 5C). Whether these HTICS-derived sub-signatures are indeed superior to HTICS or whether they represent over-fitting awaits analysis of additional/new cohorts. Here, we used these 92 top combinations in a second approach to identify core genes for HTICS employing the pair-wise analysis described above (Fig 5D). This analysis also resulted in a core of 8 genes (designated Core2): Scrn1, Nrp1, C1qb, CD74, Npy, Chaf1b, St8sia4 and Ccr2 (Fig 5B). Core2 was a better predictor than Core1 with HR of 5.33 for HER2<sup>+</sup>:ERα<sup>-</sup> patients and AUC of 0.72 by ROC analysis (Fig 5E and 5F).

Comparison of Core1 and Core2 revealed 6 common genes, 3 of which (Chaf1b, Scrn1, Npy) were up-regulated and 3 (Ccr2, CD74, Nrp1) down-regulated in enriched TICs (Fig 6A). The 6-gene “Core” monitored Cell Cycle, Cell Migration and Immune Response pathways but excluded the two other minor pathways/genes (Atp7b, St8sia4; Fig 6A and 6B). It significantly predicted survival of HER2<sup>+</sup>:ERα<sup>-</sup> patients with HR of 4.43 (Fig 6C), but exhibited a relatively weak AUC of 0.58 by ROC analysis due to high rate of false positives (Fig 6D).

## Effect of substitutions on the prognostic power of HTICS Core

We next examined the ability of the replacement genes to substitute for the Core. By substituting all 6 genes in the Core (Fig 7A), the resulting AltCore produced better HR of 6.37 than the Core in OS analysis for HER2<sup>+</sup>:ERα<sup>-</sup> patients (Fig 7B). However, like the Core, AltCore had a low AUC of 0.59 (Fig 7C). Comparing the Core and AltCore with 1000 randomly selected sets of 6-genes (with 3 gene-up; 3 gene-down), the Core produced consistently significant result (ranked #15 for MFS and #12 for OS for HER2<sup>+</sup> patients; #2 for MFS in HER2<sup>+</sup>:ERα<sup>-</sup> patients), whereas AltCore was very unstable ranking #2 for MFS but #84 for OS (Fig 7D). Therefore, while the substitutes were able to provide prognostication, they are less stable in predicting patient outcome than the original HTICS or optimized Core.

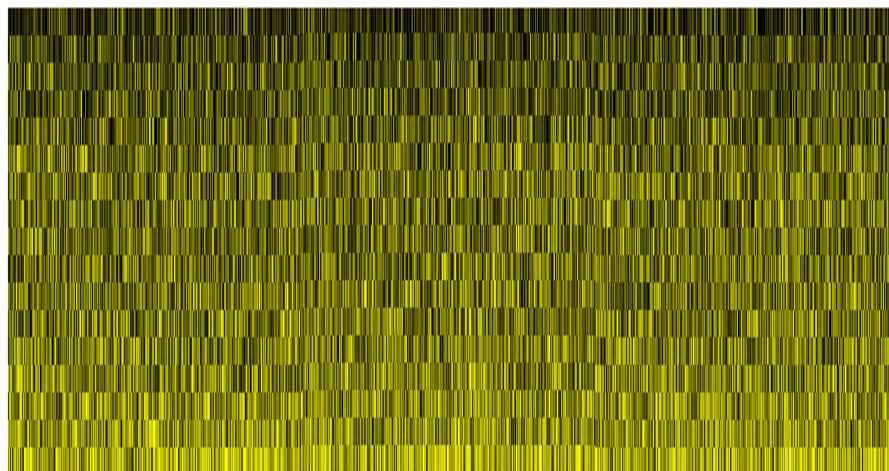
## Cross-platform analysis of the HTICS Core

Finally, to validate the efficacy of the Core, we analyzed RNA-Sequencing profiles of invasive HER2<sup>+</sup> breast cancer patients with OS data from The Cancer Genome Atlas (TCGA) [2]. Survival analysis demonstrated that HTICS Core significantly predicted OS with HR of 4.15 (Fig 7E). The AltCore, Core1 and Core2 gave similar trends but none was significant (Fig 7E).

**A** Pair-wise Analysis of HTICS Genes Within All 15,417 Significant Combinations

%Overlap	Ccna <sub>2</sub>	Scrn1	Npy	Atp7b	Chaf1b	Ccnb1	Cldn8	Nrp1	Ccr2	C1qb	Cd74	Vcam1	Cd180	Itgb2	Cd72	St8sia <sub>4</sub>
<b>Aurkb</b>	36.33	44.93	41.33	26.63	39.74	32.12	32.45	45.15	46.96	38.58	42.18	46.38	39.71	38.80	36.59	39.43
<b>Ccna2</b>		39.27	40.98	25.65	36.53	27.59	33.95	47.15	42.88	40.93	37.77	43.90	36.72	34.03	36.89	37.18
<b>Scrn1</b>			47.12	30.99	47.13	40.85	42.28	55.92	55.78	46.19	48.10	49.90	44.65	39.85	38.58	44.99
<b>Npy</b>				25.26	37.12	32.08	38.92	44.45	50.52	40.59	45.44	41.86	38.25	38.34	31.52	42.96
<b>Atp7b</b>					23.82	23.10	28.79	31.33	33.52	28.37	30.05	30.50	26.83	24.86	21.35	30.31
<b>Chaf1b</b>						34.85	37.75	44.87	48.36	45.09	43.26	44.61	42.34	39.95	32.89	40.27
<b>Ccnb1</b>							32.38	38.16	40.33	38.21	35.72	38.29	31.80	33.09	29.09	35.62
<b>Cldn8</b>								47.44	49.37	40.42	40.64	46.75	36.34	34.66	37.10	40.93
<b>Nrp1</b>									58.92	43.69	57.77	51.97	49.75	43.79	40.36	44.75
<b>Ccr2</b>										51.74	51.15	57.63	48.21	43.14	40.24	49.71
<b>C1qb</b>											45.40	43.23	37.27	36.13	31.95	38.75
<b>Cd74</b>												46.62	39.51	37.49	38.20	39.89
<b>Vcam1</b>		Ave	SD	Ave+SD									43.38	42.20	34.94	46.79
<b>Cd180</b>														35.09	31.36	37.48
<b>Itgb2</b>															26.05	35.45
<b>Cd72</b>																33.24

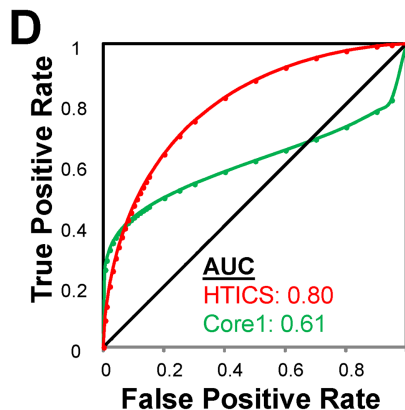
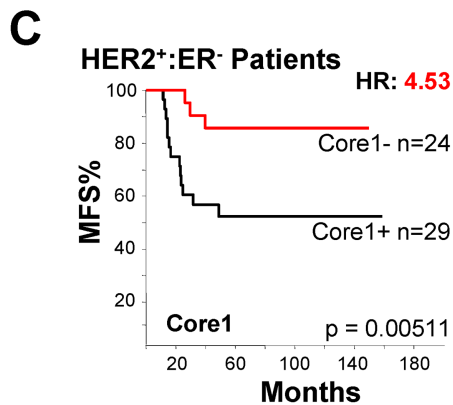
**B** 15,417 Significant Combinations



HTICS Genes	Freq%	#Genes Paired
Ccr2	80	10
Nrp1	73	7
Scrn1	69	6
Vcam1	69	3
Cd74	64	3
Chaf1b	58	2
Npy	57	2
Cd180	57	2
C1qb	59	1
St8sia4	57	1
Aurkb	56	0
Cldn8	54	2
Ccna2	51	1
Itgb2	50	0
Cd72	45	0
Ccnb1	44	0
Atp7b	34	0

**Core1**

UP- / DOWN- regulated in TICs



**Fig 4. Identification of Core 1 by analysis of all 15,417 HTICS gene combinations.** (A) Pair-wise analysis using 15,417 combinations that generated significant prognosis in both MFS and OS datasets. The frequency of combinations containing both genes in a pair is shown;

pairs with high frequencies are highlighted in red. (B) Frequency of appearance of each HTICS gene in the 15,417 significant combinations and identification of Core 1. Vertical black lines indicate a gene is present in each combination; yellow lines point to its absence. Frequency and number of HTICS genes paired are listed. (C) Kaplan-Meier analysis of Core1 on MFS cohorts. (D) ROC analysis of Core1.

<https://doi.org/10.1371/journal.pone.0179223.g004>

Therefore the 6 gene Core identified here was consistent in making prognostic predictions on independent datasets generated by different technologies (RNA microarray and RNA-seq).

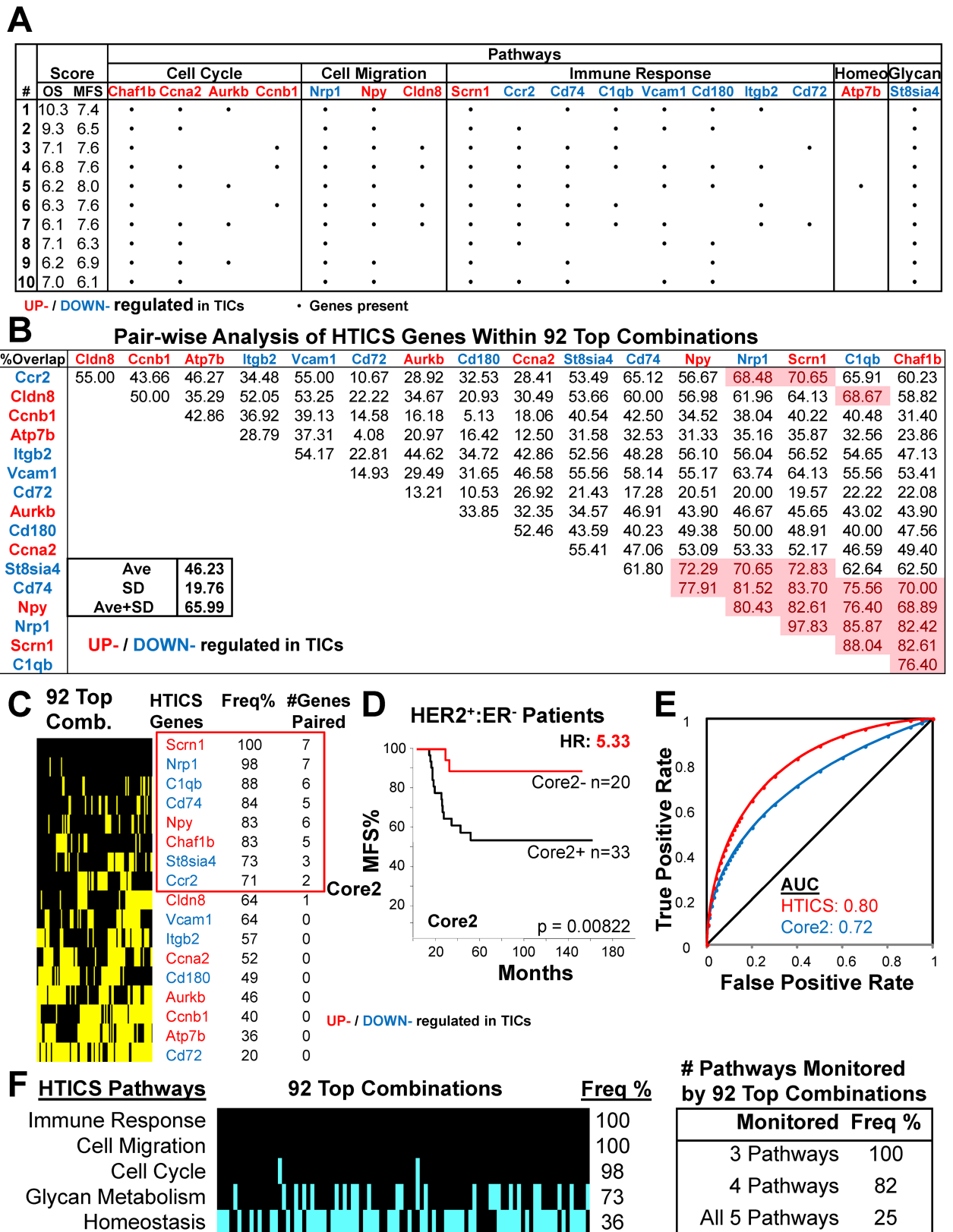
## Comparison of HTICS and Core signatures by empirical false detection rate

To compare all the signatures described herein we used Empirical False Detection Rate (eFDR). This was calculated by 10,000 permutations using the entire microarray dataset. Each permutation generated a randomly selected gene sets for HR calculation and the rate of random gene sets with significantly better HR than HTICS was recorded (hence eFDR). After 10,000 permutations, only 13 random gene sets had better HR than the original signature and thus, HTICS had eFDR of 0.00129 (Fig 8A). All six other signatures developed herein were analyzed in a similar way. The results clearly show that while all core signatures were prognostic, significant and with low eFDR, HTICS was by far the most powerful by all criteria. Thus, while the various pathways, in particularly Cell Cycle, Immune Response and Cell Migration, were sufficient to predict clinical outcome, the other pathways (Homeostasis and Glycan Metabolism) fine-tuned the signature and maximized its prognostic power (Fig 8B).

## Discussion

We report the identification of critical pathways mediating the prognostic power of HTICS, a 17-gene signature for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer [23]. We identified cell proliferation, cell migration and immune response pathways as most critical. Two other pathways, homeostasis and Glycan Metabolism, albeit not as important, contributed to optimal prognostication. We found that 15 of 17 (88%) substituting genes that could best replace the 17 HTICS genes were from the same biological pathways, highlighting their importance in predicting disease-free and overall survival for these patients. We also identified a core signature of 6 genes with high prognostic power and showed that they also include genes from the three major pathways. Thus, not all genes in HTICS equally contribute to its prognostic power. Our results suggest that IHC-based detection of the three major pathways—cell proliferation, cell migration and immune response—may successfully predict clinical outcome for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer patients. Indeed, we are currently performing retrospective analysis on large cohorts of HER2<sup>+</sup>:ERα<sup>-</sup> patients using the NanoString assay developed herein. In addition, drugs/inhibitors that can antagonize the Core or substituting genes identified herein may have therapeutic benefits.

Our finding of immune-response gene pathway in HTICS, especially of those associated exclusively with PBMC, seems counterintuitive given that the signature was derived from mouse MMTV-Her2/Neu tumor epithelial cells that were depleted of other cell lineages (endothelial and hematopoietic cells[23]). However, the non-TIC fraction was evidently contaminated with immune-cells, possibly due to incomplete lineage depletion or to lack of expression of cell surface markers used for lineage depletion (CD31; CD45-TER119, respectively). In hindsight, the inclusion of these cells, reflecting normal anti-cancer responses by the host, allowed us to generate a signature that includes immune-response genes that is critically important for the prognostic power of HTICS. Indeed, a meta analysis revealed that immune-response is an important determinant in the prognosis of HER2<sup>+</sup> breast cancer relative to



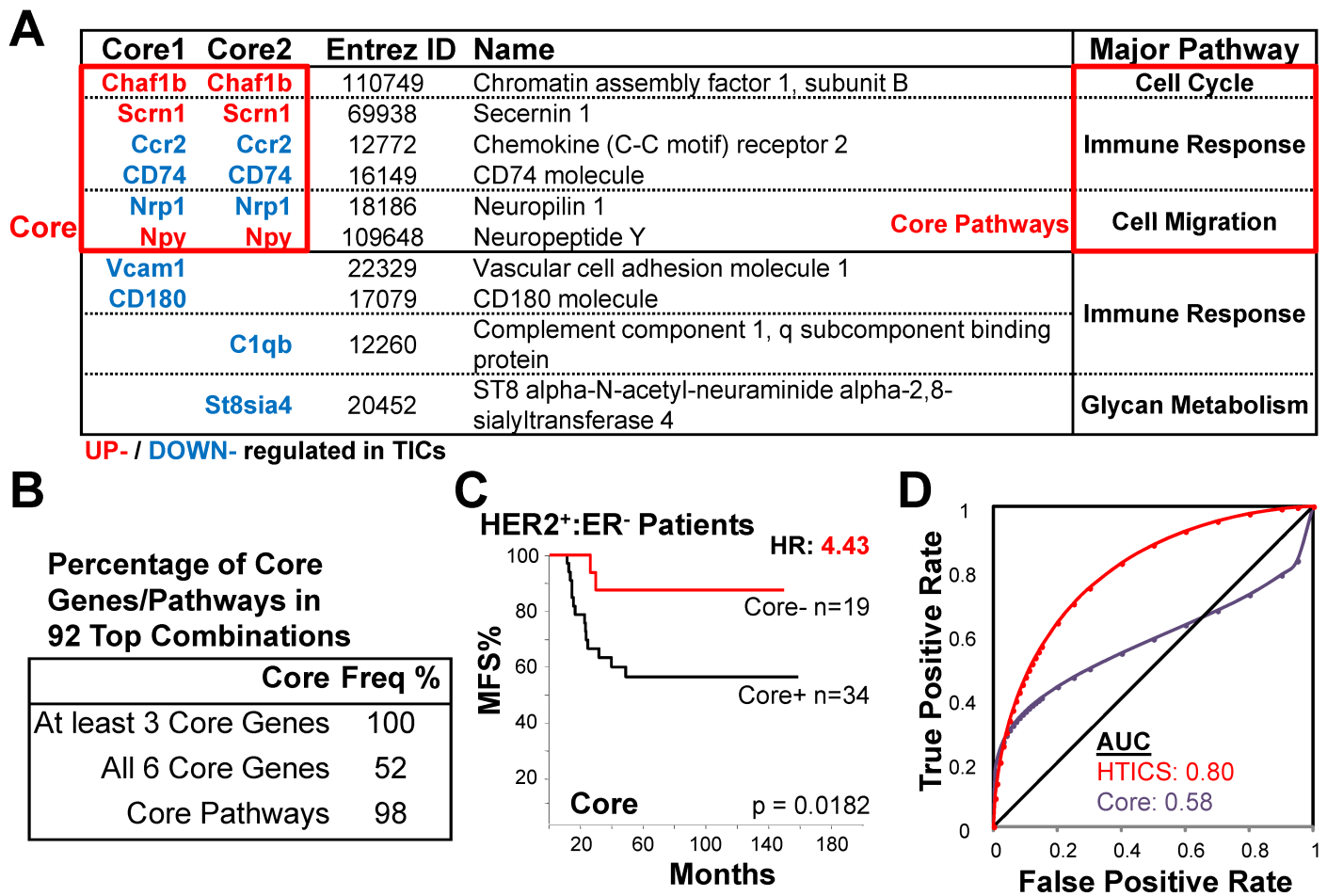
**Fig 5. Identification of Core 2 by analysis of Top 92 HTICS gene combinations.** (A) Top 10 gene combinations (of 92) with superior prognostic power. • denotes the presence of a gene in the combination. (B) Pair-wise analysis using the 92 superior combinations. Genes



that paired with at least 2 HTICS genes with >1 standard deviation above average are highlighted in red. (C) Frequency of appearance of each HTICS gene in the 92 top combinations. Vertical black lines indicate the gene is present in each combination; yellow lines indicate its absence. Core2 includes genes with higher than median frequency that paired with at least 2 genes. (D) Kaplan-Meier analysis of Core2 on MFS cohorts. (E) ROC analysis of Core2. (F) Frequency of the 5 HTICS pathways in the 92 superior combinations. Black lines indicate a pathway is present in each combination; cyan lines indicate its absence. All combinations monitored at least 3 pathways as indicated.

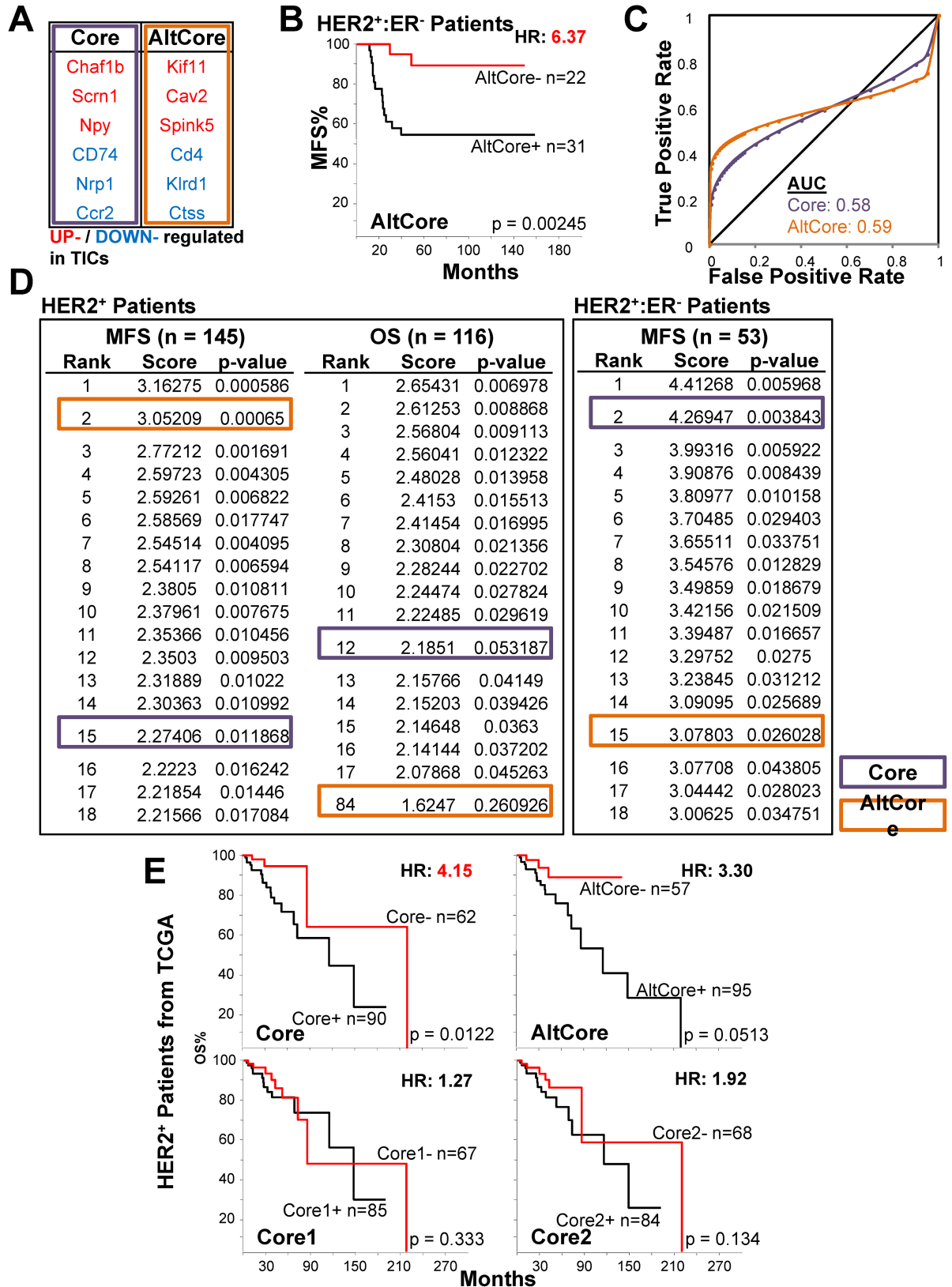
<https://doi.org/10.1371/journal.pone.0179223.g005>

other subtypes [28]. Notably, the identification of immune-related genes as part of HTICS was only made possible because we identified MMTV-Her2/Neu TICs in isogenic (FvB) mice with intact immune system. Had we attempted to generate HTICS from human HER2<sup>+</sup> TICs injected into immunocompromised mice, we would have not had the immune component of HTICs, and thus, not a strong prognostic signature. As noted, the presence of immune response pathway even in the minimal 6-gene Core, is particularly interesting given the effect of the immune system on the therapeutic effects of trastuzumab [31]. Combination therapy with trastuzumab and immunomodulatory drugs may therefore be highly synergistic for HER2<sup>+</sup>:ER<sup>-</sup> patients.



**Fig 6. Identification of 6 common Core genes for HTICS.** (A) Comparison between Core1 and Core2 and identification of a common 6 gene Core. (B) Frequency of the 6 Core genes and 3 essential pathways in the 92 good combinations. (C) Kaplan-Meier analysis of Core on MFS cohorts. (D) ROC analysis of Core.

<https://doi.org/10.1371/journal.pone.0179223.g006>



**Fig 7. Analysis of HTICS Core.** (A) AltCore with all 6 Core gene substitutions. (B) Kaplan-Meier analysis of AltCore on MFS cohorts. (C) ROC analysis of AltCore. (D) Core and AltCore tested against 1000 randomly generated signatures of equal numbers

of up- and down-regulated genes. (E) Kaplan-Meier OS analyses of Core, AltCore, Core1 and Core2 on HER2<sup>+</sup> breast cancer samples with RNA-Seq data (TCGA).

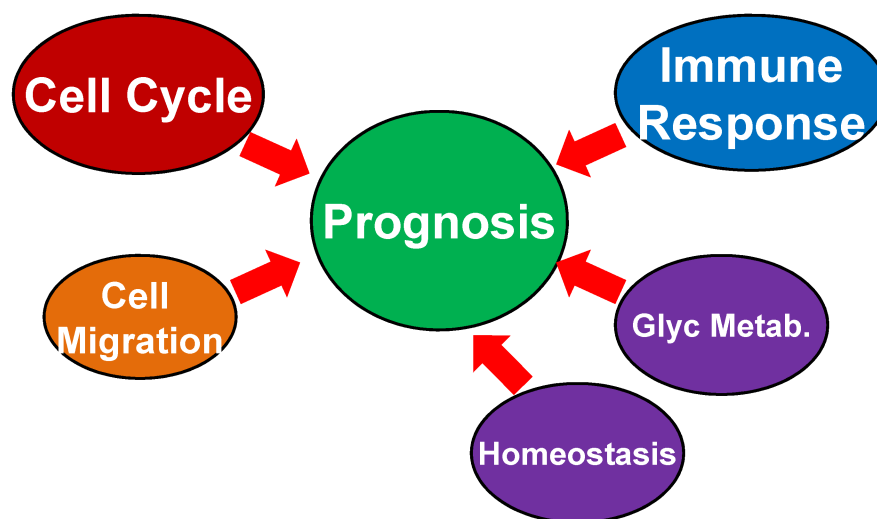
<https://doi.org/10.1371/journal.pone.0179223.g007>

One important question is whether the HTICS or Core signature genes identified here merely monitor aggressive behaviour of HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer, or whether they also

**A**

Signature	Description	HER2 <sup>+</sup> ER <sup>-</sup> MFS		
		HR	P-value	eFDR
HTICS	Original 17 Gene Signature	7.94	0.000843	0.0013
HTICS <sub>m</sub>	HTICS with 4 replacement genes	5.20	0.00277	0.0036
AltHTICS	HTICS consists of all replacement genes	2.83	0.0363	0.0143
Core1	8 core genes by all significant combinations	4.53	0.00511	0.0055
Core2	8 core genes by top 92 combinations	5.33	0.00822	0.0027
Core	Genes shared by Core1 and Core2	4.43	0.0182	0.0068
AltCore	Core consists of all replacement genes	6.37	0.00245	0.003

**B**



**Fig 8. A Model for signaling pathways that drive the prognostic power of HTICS.** (A) Assessment of HTICS and derived signatures by hazard ratio, significance and Empirical False Detection Rate (eFDR) analysis. (B) A schematic model for pathways that drive the prognostic power of HTICS. Our analysis reveals that cell cycle and immune response pathways as well as cell migration are critical components, but that additional pathways—homeostasis and Glycan metabolism—also contribute to the full prognostic power of HTICS.

<https://doi.org/10.1371/journal.pone.0179223.g008>

actively participate in and drive metastatic growth. The identification of a 6 Core gene set as well as substitutes may simplify this question by focusing on the most critical genes. For example, inhibition/knockdown of the only Core cell cycle gene Chaf1b, (Chromatin assembly factor 1, subunit B), which is required for assembly of histone octamers onto newly-replicated DNA, and/or its substitutes Kif11 (Kinesin Family Member 11), which is required for chromosome positioning, centrosome separation and bipolar spindle formation during cell mitosis, may diminish proliferation or self-renewal of HER2<sup>+</sup>:ERα<sup>-</sup> TICs.

Over-expression of Nrp1 (Neuropilin 1), a vascular endothelial cell growth factor (VEGF) 165 Receptor, has been implicated in the vascularization and metastasis of many cancer types, including breast cancer [37]. Our results suggest that anti-NRP1 therapy such as the synthetic peptide, EG3287, [38], may be assessed against HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer. The Nrp1 substitute, Klrd1 (Killer Cell Lectin-Like Receptor Subfamily D, Member 1), an antigen expressed on Natural Killer cells, may also be considered as target to block tumor dissemination. Finally, as noted, the homeostasis gene, Atp7b is required for efflux of copper out of the cells. Interestingly, disulfiram, a potent anti-cancer agent, kills tumor cells by increasing intracellular copper concentrations [39, 40].

The identification of cell proliferation, migration and immune-response pathways as key prognostic determinants in HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer suggests that IHC assays that can faithfully detect alterations in these pathways may be highly prognostic. While proliferation and immune-response can readily be tested by IHC, identification of cell migration markers may be more challenging. A starting point may be the HTICS Cell migration pathway genes (Nrp1, Npy, Cldn8) and their substitutes (Klrd1, Spink5, Itga8) identified herein—but other known markers of cancer cell dissemination or circulating tumor cells may be equally informative [41, 42].

## Conclusion

In conclusion, using bioinformatic analysis we established a rationale for identifying critical pathways as well as substitute genes and Core Signature genes that would facilitate clinical application of HTICS and may uncover new therapeutic targets for HER2<sup>+</sup>:ERα<sup>-</sup> breast cancer. Our analysis should be applicable to other types of multigene-based prognostic signatures for breast and other cancer types, and may facilitate the development of a universal breast cancer prognostic test for multiple breast cancer subtypes.

## Supporting information

### **S1 Fig. ROC analysis comparing HER2 status as determined by amplicon genes vs IHC, and comparing the prognostic power of HTICS in HER2<sup>+</sup>:ERα<sup>-</sup> vs HER2<sup>+</sup>:ERα<sup>+</sup> patients.**

(A) ROC analysis comparing HER2 status as determined by amplicon genes vs IHC. (B) ROC analysis demonstrating that HTICS is most effective in prognostication of HER2<sup>+</sup>:ERα<sup>-</sup> patients. (C) Heatmap for HTICS gene expression in HER2<sup>+</sup> patients with corresponding SSM scores and incidents of metastasis.

(TIF)

**S2 Fig. Effect of substituting other-pathway genes for AltHTICS.** Using the highest correlated genes from different pathways instead of genes in the same pathway (Rnase6 for CCR2 and Mphosph6 for Scrn1) (A), the prognostic ability of the signatures: Rnase6 in AltHTICSb and Mphosph6 in AltHTICSc, were compared with the original AltHTICS by Kaplan-Meier analysis of HER2<sup>+</sup>:ERα<sup>-</sup> patients with MFS data (B).

(TIF)

**S3 Fig. Effect of removing individual functional pathway on the prognostic power of HTICS.** (A) Kaplan-Meier analysis of HER2<sup>+</sup>:ER $\alpha$ <sup>-</sup> patients with MFS data in leave-one-out analysis of the HTICS pathways. (B) ROC analysis for HER2<sup>+</sup>:ER $\alpha$ <sup>-</sup> patients with MFS data in leave-one-out analysis of the HTICS pathways.  
(TIF)

**S1 Table. Datasets and samples with microarray and HER2 IHC for 5-gene amplicon assessments.**  
(XLSX)

**S2 Table. Overall and metastasis-free survival datasets with no herceptin treatment and having all breast cancer subtypes represented.**  
(XLSX)

**S3 Table. Metastasis-free survival incidences and time for all samples.**  
(XLSX)

**S4 Table. Nanostring counts for HTICS, backup, and control genes in a panel of FFPE breast cancer samples.**  
(XLSX)

## Acknowledgments

We thank Dr. Shannon Dunn, UHN, for human peripheral blood mononuclear cell (PBMC) samples, and Drs. Rod Bremner and Sean McCurdy for advise on ROC analysis.

## Author Contributions

**Conceptualization:** JCL EZ.

**Data curation:** JCL.

**Formal analysis:** JCL MZ.

**Funding acquisition:** EZ AE SNM.

**Investigation:** JCL EZ.

**Methodology:** JCL.

**Resources:** JCL EZ AE SNM.

**Software:** JCL.

**Supervision:** EZ.

**Validation:** JCL.

**Visualization:** JCL.

**Writing – original draft:** JCL EZ.

**Writing – review & editing:** JCL EZ.

## References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000; 406(6797):747–52. <https://doi.org/10.1038/35021093> PMID: 10963602

2. Network TCGA. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490(7418):61–70. Epub 2012/09/25. <https://doi.org/10.1038/nature11412> PMID: 23000897; PubMed Central PMCID: PMC3465532.
3. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012; 486(7403):346–52. Epub 2012/04/24. <https://doi.org/10.1038/nature10983> PMID: 22522925.
4. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012; 486(7403):405–9. Epub 2012/06/23. <https://doi.org/10.1038/nature11154> PMID: 22722202; PubMed Central PMCID: PMC4148686.
5. Eswaran J, Cyanam D, Mudvari P, Reddy SD, Pakala SB, Nair SS, et al. Transcriptomic landscape of breast cancers through mRNA sequencing. *Scientific reports*. 2012; 2:264. Epub 2012/02/23. <https://doi.org/10.1038/srep00264> PMID: 22355776; PubMed Central PMCID: PMC3278922.
6. Gray J, Druker B. Genomics: the breast cancer landscape. *Nature*. 2012; 486(7403):328–9. Epub 2012/06/23. <https://doi.org/10.1038/486328a> PMID: 22722187.
7. Goss PE, Strasser-Weippl K. Prevention strategies with aromatase inhibitors. *Clin Cancer Res*. 2004; 10(1 Pt 2):372S–9S. Epub 2004/01/22. PMID: 14734494.
8. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 2001; 344(11):783–92. PMID: 11248153. <https://doi.org/10.1056/NEJM200103153441101>
9. Abramson V, Arteaga CL. New strategies in HER2-overexpressing breast cancer: many combinations of targeted drugs available. *Clin Cancer Res*. 2011; 17(5):952–8. PMID: 21248299. <https://doi.org/10.1158/1078-0432.CCR-09-1947>
10. Dean-Colomb W, Esteva FJ. Her2-positive breast cancer: herceptin and beyond. *Eur J Cancer*. 2008; 44(18):2806–12. PMID: 19022660. <https://doi.org/10.1016/j.ejca.2008.09.013>
11. Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. *The New England journal of medicine*. 2012; 367(19):1783–91. Epub 2012/10/02. <https://doi.org/10.1056/NEJMoa1209124> PMID: 23020162.
12. Swain SM, Baselga J, Kim SB, Ro J, Semiglazov V, Campone M, et al. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *The New England journal of medicine*. 2015; 372(8):724–34. Epub 2015/02/19. <https://doi.org/10.1056/NEJMoa1413513> PMID: 25693012.
13. Baselga J, Cortes J, Kim SB, Im SA, Hegg R, Im YH, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *The New England journal of medicine*. 2012; 366(2):109–19. Epub 2011/12/14. <https://doi.org/10.1056/NEJMoa1113216> PMID: 22149875.
14. O'Sullivan CC, Bradbury I, Campbell C, Spielmann M, Perez EA, Joensuu H, et al. Efficacy of Adjuvant Trastuzumab for Patients With Human Epidermal Growth Factor Receptor 2-Positive Early Breast Cancer and Tumors  $\leq 2$  cm: A Meta-Analysis of the Randomized Trastuzumab Trials. *J Clin Oncol*. 2015; 33(24):2600–8. <https://doi.org/10.1200/JCO.2015.60.8620> PMID: 26101239; PubMed Central PMCID: PMC4534523.
15. Tolaney SM, Barry WT, Dang CT, Yardley DA, Moy B, Marcom PK, et al. Adjuvant paclitaxel and trastuzumab for node-negative, HER2-positive breast cancer. *N Engl J Med*. 2015; 372(2):134–41. <https://doi.org/10.1056/NEJMoa1406281> PMID: 25564897; PubMed Central PMCID: PMC4313867.
16. Liu JC, Egan SE, Zacksenhaus E. A Tumor initiating cell-enriched prognostic signature for HER2+: ER $\alpha$ -breast cancer; rationale, new features, controversies and future directions. *Oncotarget*. 2013; 4(8):1317–28. Epub 2013/08/16. PMID: 23945331; PubMed Central PMCID: PMC3787160. <https://doi.org/10.18632/oncotarget.1170>
17. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine*. 2004; 351(27):2817–26. Epub 2004/12/14. <https://doi.org/10.1056/NEJMoa041588> PMID: 15591335.
18. Liu R, Wang X, Chen GY, Dalerba P, Gurney A, Hoey T, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med*. 2007; 356(3):217–26. PMID: 17229949. <https://doi.org/10.1056/NEJMoa063994>
19. Finak G, Bertos N, Pepin F, Sadekova S, Souleimanova M, Zhao H, et al. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med*. 2008; 14(5):518–27. PMID: 18438415. <https://doi.org/10.1038/nm1764>
20. Tofigh A, Suderman M, Paquet ER, Livingstone J, Bertos N, Saleh SM, et al. The prognostic ease and difficulty of invasive breast carcinoma. *Cell reports*. 2014; 9(1):129–42. Epub 2014/10/07. <https://doi.org/10.1016/j.celrep.2014.08.073> PMID: 25284793.



21. Hallett RM, Dvorkin-Gheva A, Bane A, Hassell JA. A gene signature for predicting outcome in patients with basal-like breast cancer. *Sci Rep.* 2012; 2:227. Epub 2012/02/23. <https://doi.org/10.1038/srep00227> PMID: 22355741; PubMed Central PMCID: PMC3259129.
22. Liu JC, Voisin V, Wang S, Wang DY, Jones RA, Datti A, et al. Combined deletion of Pten and p53 in mammary epithelium accelerates triple-negative breast cancer with dependency on eEF2K. *EMBO molecular medicine.* 2014; 6(12):1542–60. Epub 2014/10/22. <https://doi.org/10.15252/emmm.201404402> PMID: 25330770; PubMed Central PMCID: PMC4287974.
23. Liu JC, Voisin V, Bader GD, Deng T, Pusztai L, Symmans WF, et al. Seventeen-gene signature from enriched Her2/Neu mammary tumor-initiating cells predicts clinical outcome for human HER2<sup>+</sup>:ER<sup>+</sup>-breast cancer. *Proc Natl Acad Sci U S A.* 2012; 109(15):5832–7. Epub 2012/03/31. <https://doi.org/10.1073/pnas.1201105109> PMID: 22460789; PubMed Central PMCID: PMC3326451.
24. Kreso A, Dick JE. Evolution of the Cancer Stem Cell Model. *Cell stem cell.* 2014; 14(3):275–91. Epub 2014/03/13. <https://doi.org/10.1016/j.stem.2014.02.006> PMID: 24607403.
25. Liu JC, Deng T, Lehal RS, Kim J, Zacksenhaus E. Identification of tumorsphere- and tumor-initiating cells in HER2/Neu-induced mammary tumors. *Cancer Res.* 2007; 67(18):8671–81. Epub 2007/09/19. <https://doi.org/10.1158/0008-5472.CAN-07-1486> PMID: 17875707.
26. Guy CT, Webster MA, Schaller M, Parsons TJ, Cardiff RD, Muller WJ. Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proceedings of the National Academy of Sciences of the United States of America.* 1992; 89(22):10578–82. Epub 1992/11/15. PMID: 1359541; PubMed Central PMCID: PMC50384.
27. Staaf J, Ringner M, Vallon-Christersson J, Jonsson G, Bendahl PO, Holm K, et al. Identification of subtypes in human epidermal growth factor receptor 2—positive breast cancer reveals a gene signature prognostic of outcome. *J Clin Oncol.* 2010; 28(11):1813–20. Epub 2010/03/17. <https://doi.org/10.1200/JCO.2009.22.8775> PMID: 20231686.
28. Desmedt C, Haibe-Kains B, Wirapati P, Buyse M, Larsimont D, Bontempi G, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res.* 2008; 14(16):5158–65. Epub 2008/08/14. <https://doi.org/10.1158/1078-0432.CCR-07-4756> PMID: 18698033.
29. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. *Nat Rev Cancer.* 2006; 6(2):99–106. Epub 2006/02/24. <https://doi.org/10.1038/nrc1802> PMID: 16491069.
30. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol.* 2008; 26(3):317–25. PMID: 18278033. <https://doi.org/10.1038/nbt1385>
31. Bianchini G, Gianni L. The immune system and response to HER2-targeted treatment in breast cancer. *The Lancet Oncology.* 2014; 15(2):e58–68. Epub 2014/02/01. [https://doi.org/10.1016/S1470-2045\(13\)70477-7](https://doi.org/10.1016/S1470-2045(13)70477-7) PMID: 24480556.
32. Deng T, Liu JC, Chung PE, Uehling D, Aman A, Joseph B, et al. shRNA Kinome Screen Identifies TBK1 as a Therapeutic Target for HER2<sup>+</sup> Breast Cancer. *Cancer Res.* 2014; 74(7):2119–30. Epub 2014/02/04. <https://doi.org/10.1158/0008-5472.CAN-13-2138> PMID: 24487029.
33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102(43):15545–50. PMID: 16199517. <https://doi.org/10.1073/pnas.0506580102>
34. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One.* 2011; 5(11):e13984. PMID: 21085593. <https://doi.org/10.1371/journal.pone.0013984>
35. Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol.* 2011; 7(10):e1002240. Epub 2011/10/27. <https://doi.org/10.1371/journal.pcbi.1002240> PMID: 22028643; PubMed Central PMCID: PMC3197658.
36. Mouttet D, Lae M, Caly M, Gentien D, Carpentier S, Peyro-Saint-Paul H, et al. Estrogen-Receptor, Progesterone-Receptor and HER2 Status Determination in Invasive Breast Cancer. Concordance between Immuno-Histochemistry and MapQuant Microarray Based Assay. *PLoS One.* 2016; 11(2):e0146474. <https://doi.org/10.1371/journal.pone.0146474> PMID: 26829108; PubMed Central PMCID: PMCPMC4735463.
37. Bachelder RE, Crago A, Chung J, Wendt MA, Shaw LM, Robinson G, et al. Vascular endothelial growth factor is an autocrine survival factor for neuropilin-expressing breast carcinoma cells. *Cancer Res.* 2001; 61(15):5736–40. PMID: 11479209.
38. Barr MP, Byrne AM, Duffy AM, Condron CM, Devocelle M, Harriott P, et al. A peptide corresponding to the neuropilin-1-binding site on VEGF(165) induces apoptosis of neuropilin-1-expressing breast tumour cells. *Br J Cancer.* 2005; 92(2):328–33. <https://doi.org/10.1038/sj.bjc.6602308> PMID: 15655556; PubMed Central PMCID: PMCPMC2361857.

39. Robinson TJ, Pai M, Liu JC, Vizeacoumar F, Sun T, Egan SE, et al. High-throughput screen identifies disulfiram as a potential therapeutic for triple-negative breast cancer cells: Interaction with IQ motif-containing factors. *Cell Cycle*. 2013; 12(18). Epub 2013/08/27. PMID: [23974104](https://pubmed.ncbi.nlm.nih.gov/23974104/). <https://doi.org/10.4161/cc.26063>
40. Allensworth JL, Evans MK, Bertucci F, Aldrich AJ, Festa RA, Finetti P, et al. Disulfiram (DSF) acts as a copper ionophore to induce copper-dependent oxidative stress and mediate anti-tumor efficacy in inflammatory breast cancer. *Molecular oncology*. 2015; 9(6):1155–68. <https://doi.org/10.1016/j.molonc.2015.02.007> PMID: [25769405](https://pubmed.ncbi.nlm.nih.gov/25769405/); PubMed Central PMCID: [PMCPMC4493866](https://pubmed.ncbi.nlm.nih.gov/PMC4493866/).
41. Kang Y, Pantel K. Tumor cell dissemination: emerging biological insights from animal models and cancer patients. *Cancer cell*. 2013; 23(5):573–81. Epub 2013/05/18. <https://doi.org/10.1016/j.ccr.2013.04.017> PMID: [23680145](https://pubmed.ncbi.nlm.nih.gov/23680145/); PubMed Central PMCID: [PMC3667710](https://pubmed.ncbi.nlm.nih.gov/PMC3667710/).
42. Sosa MS, Bragado P, Aguirre-Ghiso JA. Mechanisms of disseminated cancer cell dormancy: an awakening field. *Nature reviews Cancer*. 2014; 14(9):611–22. Epub 2014/08/15. <https://doi.org/10.1038/nrc3793> PMID: [25118602](https://pubmed.ncbi.nlm.nih.gov/25118602/); PubMed Central PMCID: [PMC4230700](https://pubmed.ncbi.nlm.nih.gov/PMC4230700/).