Journal of Pathology
J Pathol December 2023; **261:** 477–489
Published online 22 September 2023 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.6205

ORIGINAL ARTICLE

Canonical Kaiso target genes define a functional signature that associates with breast cancer survival and the invasive lobular carcinoma histological type

Thijmen Sijnesael[1†], François Richard[2†], Max AK Rätze[1†], Thijs Kooman[1], Blessing Bassey-Archibong[3], Christa Rohof[1], Juliet Daniel[3], Christine Desmedt[2] and Patrick WB Derksen[1*]

[1] *Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands*
[2] *Laboratory for Translational Breast Cancer Research, Katholieke Universiteit Leuven, Leuven, Belgium*
[3] *Department of Biology, McMaster University, Hamilton, ON, Canada*

**Correspondence to: PWB Derksen, Department of Pathology, UMC Utrecht, H04.312, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands. E-mail: p.w.b.derksen@umcutrecht.nl*

[†]Equal contribution.

Abstract

Invasive lobular carcinoma (ILC) is a low- to intermediate-grade histological breast cancer type caused by mutational inactivation of E-cadherin function, resulting in the acquisition of anchorage independence (anoikis resistance). Most ILC cases express estrogen receptors, but options are limited in relapsed endocrine-refractory disease as ILC tends to be less responsive to standard chemotherapy. Moreover, ILC can relapse after >15 years, an event that currently cannot be predicted. E-cadherin inactivation leads to p120-catenin-dependent relief of the transcriptional repressor Kaiso (ZBTB33) and activation of canonical Kaiso target genes. Here, we examined whether an anchorage-independent and ILC-specific transcriptional program correlated with clinical parameters in breast cancer. Based on the presence of a canonical Kaiso-binding consensus sequence (cKBS) in the promoters of genes that are upregulated under anchorage-independent conditions, we defined an ILC-specific anoikis resistance transcriptome (ART). Converting the ART genes into human orthologs and adding published Kaiso target genes resulted in the Kaiso-specific ART (KART) 33-gene signature, used subsequently to study correlations with histological and clinical variables in primary breast cancer. Using publicly available data for $ER^{POS}Her2^{NEG}$ breast cancer, we found that expression of KART was positively associated with the histological ILC breast cancer type ($p < 2.7E-07$). KART expression associated with younger patients in all invasive breast cancers and smaller tumors in invasive ductal carcinoma of no special type (IDC-NST) (<2 cm, $p < 6.3E-10$). We observed associations with favorable long-term prognosis in both ILC (hazard ratio [HR] = 0.51, 95% CI = 0.29–0.91, $p < 3.4E-02$) and IDC-NST (HR = 0.79, 95% CI = 0.66–0.93, $p < 1.2E-04$). Our analysis thus defines a new mRNA expression signature for human breast cancer based on canonical Kaiso target genes that are upregulated in E-cadherin deficient ILC. The KART signature may enable a deeper understanding of ILC biology and etiology.

Keywords: lobular breast cancer; Kaiso; E-cadherin; gene signature; KART; $ER^{POS}Her2^{NEG}$ breast cancer; ILC

Received 10 January 2023; Revised 7 July 2023; Accepted 17 August 2023

No conflicts of interest were declared.

Introduction

Mutational inactivation of E-cadherin and a subsequent dismantling of the cellular adhesive structure called the adherens junction (AJ) causes the development and progression of invasive lobular carcinoma (ILC) (reviewed in [1]). Due to the functional inactivation of E-cadherin-based cell–cell contacts, ILC cells invade diffusely as noncohesive cells in trabecular structures or in a single-file pattern [2]. Recent evidence has established that E-cadherin inactivation leads to increased and concomitant activation of PI3K/AKT signals independently of activating mutations [3–5], which provide a mechanism for the previously observed AKT phosphorylation in lobular breast cancers [6]. In addition, E-cadherin loss leads to direct translocation of p120-catenin (p120) from the cell membrane to the cytoplasm and nucleus [7–9], a process that is further promoted by loss of cell-matrix contacts [10]. In healthy cells, the main function of p120 includes providing stability for all classical cadherin family members, including E-cadherin [11,12], and spatial regulation

of RhoA-dependent actomyosin contraction at the cleavage furrow during cytokinesis [13]. Importantly, p120 can also bind and relieve transcriptional repression of Kaiso (ZBTB33), a bimodal regulator of transcription [14]. Kaiso can repress mRNA transcription through binding to the canonical Kaiso-binding sequence (cKBS) *TCCTGCNA* or modulate a noncanonical function through the *CGCG*-containing consensus KBS *TCTCGCGAGA* [15,16].

The ability of p120 to relieve canonical Kaiso-dependent transcriptional repression relies on nuclear translocation of p120 through a conserved nuclear localization sequence (NLS) [17,18]. Canonical p120-dependent KBS targets such as Siamois [19], Wnt11 [20], and Cyclin D1 [19,21] have been associated with tissue homeostasis in frogs, flies, and/or mammalian cells. Although the exact mechanism remains unclear, nuclear influx of p120 increases upon transfer of cells to anchorage-independent culture conditions [10]. Under these conditions, the cKBS target Wnt11 is transcribed as a direct consequence of E-cadherin loss and Kaiso de-repression by p120, to subsequently promote autocrine RhoA-dependent-anoikis resistance in lobular breast cancer cells [10].

Although the histological subtypes of ILC and invasive ductal carcinoma of no special type (IDC-NST) display distinct molecular landscapes [6,22,23], this has not yet evolved into specific classifiers that can contribute to the development of a tailored clinical intervention for ILC. Here, we present a functional and experimentally derived transcriptional signature based on 33 canonical Kaiso target genes (KART) that associates with the lobular histological subtype. We show that high KART signatures associate with a favorable prognosis in breast cancer, irrespective of the histological type.

Materials and methods

ERBB3 promoter analysis

Promoter analysis of human *ERBB3* was performed using the Eukaryotic Promoter Database (EPD) [24], with evaluation of the 1,000 upstream base pairs of the gene and exploration of the ZBTB33 (Kaiso)-binding sites (full KBS sequence TCCTGCNA and core KBS sequence CTGCNA), and plotted using Adobe Illustrator (Adobe Inc, San Jose, CA, USA). Promoter analysis of mouse *Erbb3* gene was performed using the NCBI Genome Data Viewer (NCBI, Bethesda, MD, USA) [25].

KBS anoikis resistance transcriptome (KART)

The mouse ILC anoikis resistance transcriptome (ART) was described in van de Ven *et al* 2015 [10]; we derived the human counterparts from 27 candidate mouse KBS-containing ART genes through Ensemble (EMBL-EBI, Hinxton, UK) [26] and extracted the promoter regions using the EPD (SIB, Eclubens, Switzerland) [24].

Kaiso chromatin immunoprecipitations (ChIP) and PCR

ChIP procedures and PCR analyses were conducted as described previously [27]. The following primers were used for amplification of the target promoter sequences: cKBS Forward (5'–3') TTGAAATGCAAGGCCGTCTG and Reverse (5'–3') CCACAGAGACCGCGTGAAAT, non-cKBS Forward (5'–3') GTTGGGGTAAGGTC ACAGA, and Reverse (5'–3') GGAATAGAAAGGCG GGAAAG.

Cell culture

MCF7 cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA), STR-type verified by PCR, and cultured as described previously [9]. Generation of E-cadherin knockout MCF7::Δ*CDH1* cells using CRISPR-Cas9 editing has been described previously [5]. Designed primer pairs are shown in supplementary material, Table S1.

CRISPR/Cas9-mediated knockout of kaiso (ZBTB33)

Baculoviruses expressing Cas9 were produced as described previously [28]. Two guide (g)RNAs were tested in MCF7 cells, and after tracking of indels by decomposition analysis [29] the cells with the highest gRNA efficiency were clonally expanded and screened for Kaiso expression using western blotting.

Western blotting

Western blotting was performed as described previously [10]. The following primary antibodies were used for western blot analysis: monoclonal rabbit HER3/ERBB3 clone D22C5 (#12708; 1:1,000; Cell Signaling Technology, Leiden, The Netherlands), monoclonal mouse anti-Kaiso 6F/6F8 (#12723; 1:10000; Abcam, Amsterdam, The Netherlands), and polyclonal goat anti-Akt1 C-20 (sc-1618; 1:1,000; Santa Cruz Biotechnology, Heidelberg, Germany).

RT-qPCR

Total RNA was extracted from cell or organoid pellets using Trizol (15596026, Thermo Fisher Scientific, Bleiswijk, The Netherlands). Poly-T primers and a cDNA reverse transcription kit (iScript Synthesis kit, 1708890, BioRad, Lunteren, The Netherlands) were used to generate cDNA. Two *ERBB3*-specific primer sets (supplementary material, Table S1) were used to evaluate expression values for human *ERBB3* using PCR. Primer efficiency was assessed by serial dilution. Expression values were generated using ΔΔCt values normalized to *GAPDH* for control and *ERBB3*. Experiments were performed in triplicate over three independent biological and technical settings using the BioRad CFX96 Real-Time System (Bio-Rad

Laboratories Hercules, CA, USA) and BioRad CFX manager software (Bio-Rad Laboratories). For each comparison, unpaired two-tailed Student's *t*-tests were used to determine statistical significance.

Bioinformatics analysis of patient data

The METABRIC [30] dataset, including clinical data and normalized gene expression, and the clinical data of The Cancer Genome Atlas (TCGA) dataset [31] were retrieved through cBioPortal [32] in August 2022. The TCGA normalized expression data were retrieved from the GDC data portal (https://gdc.cancer.gov/) in September 2021. Breast cancer types were assessed according to ER immunohistochemistry and ERBB2/HER2 immunohistochemistry or FISH status. Only patients with ER-positive/HER2-negative IDC-NST and ILC breast cancers were kept in the downstream analysis. This resulted in data from 227 IDC-NST and 98 ILC patients for TCGA and data from 979 IDC-NST and 118 ILC patients in METABRIC. Gene expression signatures were retrieved from the literature (ESR1_signature, AURKA, PLAU, STAT1 [33]; DCN [34]; gene expression grade index (GGI) [35]; SDDP [36]; Immune_Perez [37]; IRM [38]; immune cell signatures [39]; GENE21 [40]; LobSig [41], and computed as described in [42]). Wilcoxon tests were performed to compare continuous to categorical variables. Correlations were assessed using Spearman coefficients. In the heat maps, only significant correlations are colored: red, anticorrelated; blue correlated. Associations with disease-free survival (DFS) and overall survival (OS) were assessed using stratified log-rank test and Cox proportional hazard regression. Multivariable Cox models were adjusted for age (>50 versus ≤50), tumor size (≥2 versus <2), and nodal status (positive versus negative). The follow-up was curtailed at 8 and 20 years for the TCGA and METABRIC datasets, respectively, in consideration of the declining numbers of patients after that time point. *P* values were two-sided and statistical significance considered for $p < 0.05$. All analyses were performed using R version 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria) [43].

Results

Defining an anchorage-independent transcriptome driven by E-cadherin loss and canonical Kaiso target genes

We have used mouse ILC cells to couple anchorage independence upon loss of E-cadherin to the activation of a gene set in anchorage-independent (suspension) conditions [10,44]. Because mutational E-cadherin inactivation leads to a p120-dependent activation of Kaiso target genes (Figure 1A), we had used this gene set to define a set of 27 candidate mouse Kaiso targets based on the presence of a cKBS within a region 1 kb upstream of the transcriptional start site (TSS) [10]. These mouse ILC candidate Kaiso targets [10] were converted to human orthologs, and the presence of cKBS sites in these promoters was confirmed using the EPD (SIB) [24] (Table 1).

Because *ERBB3* activation recently was associated with ILC [23] and the *ERBB3* promoter contains a cKBS at position –904 (Figure 1B), we next verified whether *ERBB3* was a cKBS Kaiso target gene. For this we performed chromatin immunoprecipitations (ChIP) with monoclonal Kaiso antibodies, followed by an *ERBB3* promoter-specific PCR, which demonstrated that *ERBB3* was a *bona fide* canonical Kaiso target gene in E-cadherin-negative (MDA-MB-231) and E-cadherin-positive (MCF7) cells (Figure 1C). Next, we generated CRISPR/Cas9-mediated MCF7 Kaiso knockout cells (MCF7::ΔKaiso) using primers described in supplementary material, Table S1 (Figure 1D), and observed that Kaiso knockout induced a sixfold increase in *ERBB3* mRNA expression compared to control cells (Figure 1E). Moreover, E-cadherin loss or Kaiso knockout both increase ERBB3 protein expression by 2.1- and 5.0-fold, respectively (Figure 1F). In sum, we have identified *ERBB3* as a canonical Kaiso target, which we added to the ART signature (Table 1).

To develop an mRNA signature for human breast cancer, we added the published p120-dependent cKBS target genes such as *WNT11* [20], *CCND1* [19,21], *CCNE1* [45], *FOS* [19], *MYC* [19], *MMP7* [46,47], and *ID2* [44] to define a comprehensive KART signature consisting of 33 genes (Table 1).

Expression of KART signature associates with histological subtype ILC and clinicopathological parameters in breast cancer

We next analyzed expression of the KART signature in the publicly available $ER^{POS}/HER2^{NEG}$ mRNA expression datasets from TCGA (IDC-NST: $n = 227$ and ILC: $n = 98$) and METABRIC (IDC-NST: $n = 979$ and ILC: $n = 118$) to probe for possible associations with clinicopathological parameters. We found that the KART signature was positively associated with the ILC subtype in both the TCGA ($p = 1.1E-11$) and METABRIC ($p = 2.7E-07$) datasets (Figure 2A). Interestingly, KART is also associated with younger age at diagnosis (<50 years), both in patients with ILC ($p = 8.3E-02$ and $p = 1.8E-02$ in TCGA and METABRIC, respectively) and IDC-NST ($p = 1.1E-02$ and $p = 5.5E-08$ in TCGA and METABRIC, respectively) (Figure 2B,C), and smaller tumors in IDC-NST only ($p = 2.1E-02$ and $p = 6.3E-10$ in TCGA and METABRIC, respectively) (Figure 2B,C). In METABRIC, KART positively associates with low-grade tumors in both the ILC and IDC-NST cohorts ($p = 9.8E-03$ and $p = 6.7E-05$) (Figure 2C). In short, we show that the KART signature associates with the histological subtype ILC, younger patients, lower-grade tumors in all breast cancers, and smaller tumor size in the histological group IDC-NST.

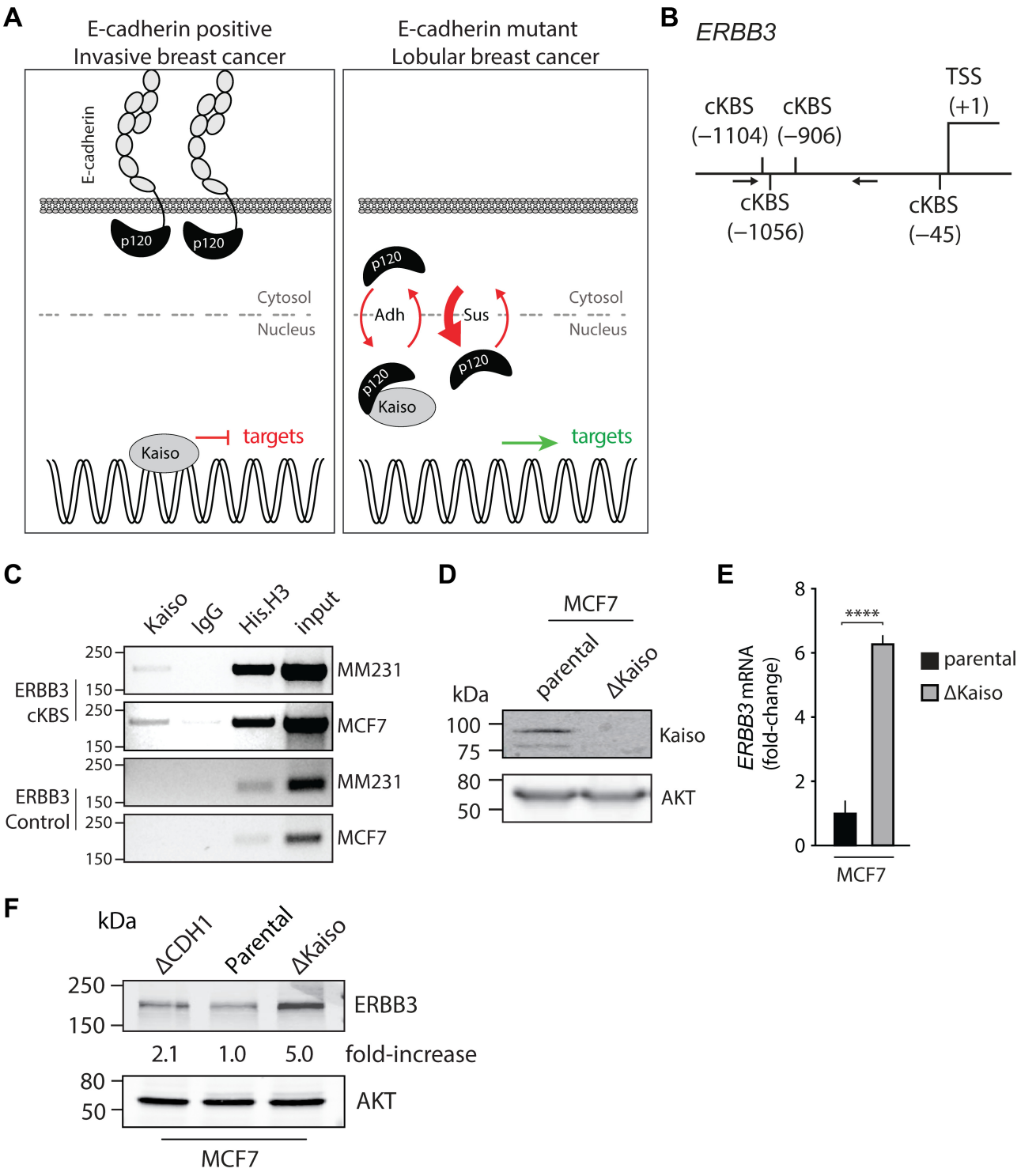


Figure 1. ERBB3 is a direct Kaiso transcriptional target gene in breast cancer cells. (A) Control over p120-dependent Kaiso transcriptional repression in the context of E-cadherin expression. Cartoon depicting E-cadherin-positive ductal breast cancer (left panel), where p120-catenin (p120) is retained at the membrane in the adherens junction (AJ) complex. In this setting, Kaiso represses expression of its target genes by KBS-dependent binding. In E-cadherin mutant lobular breast cancer (right panel), p120 is translocated to the cytosol and nucleus. Under anchorage-independent (suspension) conditions, nuclear influx of p120 is increased approximately twofold, leading to a relief of Kaiso transcriptional repression. (B) Shown is the *ERBB3* promoter region (up to −1,000 base pairs) that was analyzed for the presence of canonical Kaiso binding sites (cKBS; *CTGCNA*), plotted in base pairs upstream of the transcription start site (TSS). (C) Kaiso specific chromatin immunoprecipitation (ChIP) and subsequent PCR analyses on lysates from MDA-MB-231 (MM231) and MCF7 reveal that Kaiso binds the *ERBB3* promoter. (D) Western blot analysis depicting loss of Kaiso expression upon knockout in MCF7 cells (MCF7::ΔKaiso). AKT was used as loading control. (E) Bar graph showing *ERBB3* mRNA expression in parental MCF7 and MCF7::ΔKaiso. **** = $p < 0.0001$. (F) Kaiso controls ERBB3 protein levels. Shown are western blot analyses of ERBB3 expression levels in parental MCF7 cells, MCF7::ΔKaiso cells, and MCF7::ΔCDH1 cells grown under adherent conditions. The fold increase in ERBB3 expression relative to the loading control (AKT) is shown below the ERBB3 blot in bold typeface.

Table 1. Combined KBS anoikis resistance transcriptome (KART).

No.	Gene symbol	Ensembl entry	Origin	PMID reference
1	CCND1*	ENSG00000110092	*Xenopus laevis*/*Homo sapiens*	15935774/27694442 + 23226276
2	CCNE1*	ENSG00000105173	*Homo sapiens*	27694442
3	FOS*	ENSG00000170345	*Xenopus laevis*	15935774
4	MMP7*	ENSG00000137673	*Homo sapiens*	23251453 + 18653469
5	MYC*	ENSG00000136997	*Xenopus laevis*	15935774
6	WNT11*	ENSG00000085741	*Xenopus laevis*/*Mus musculus*	15543138/25713299
7	ID2*	ENSG00000115738	*Homo sapiens*/*Mus musculus*	35610485/25713299
8	ERBB3*	ENSG00000065361	*Homo sapiens*	Described in this article
9	AGT	ENSG00000135744	*Mus musculus*	25713299
10	ALDH1A3	ENSG00000184254		
11	ALDH3B2	ENSG00000132746		
12	ASPRV1	ENSG00000244617		
13	CAMK2N1	ENSG00000162545		
14	CCPG1	ENSG00000260916		
15	CDSN	ENSG00000204539		
16	DNAJB9	ENSG00000128590		
17	FABP5	ENSG00000164687		
18	GABRP	ENSG00000094755		
19	IL33	ENSG00000137033		
20	KRT78	ENSG00000170423		
21	LGR4	ENSG00000205213		
22	LY6D	ENSG00000167656		
23	MAN1C1	ENSG00000117643		
24	MXD1	ENSG00000059728		
25	MYLIP	ENSG00000007944		
26	RORC	ENSG00000143365		
27	SECTM1	ENSG00000141574		
28	SLC43A2	ENSG00000167703		
29	TACSTD2	ENSG00000184292		
30	TGFB3	ENSG00000119699		
31	TMEM176A	ENSG00000002933		
32	TMEM176B	ENSG00000106565		
33	TRIM29	ENSG00000137699		

*Validated by ChIP and/or functional analyses.

Individual contributions of KART signature genes to clinical associations

To investigate how much individual KART genes contribute to breast cancer histological subtypes, we analyzed the expression of the separate genes associated with ductal or lobular diagnosis in the TCGA breast cancer dataset. From the nine most abundantly expressed KART genes, seven individual genes (*TACSTD2*, *ERBB3*, *FOS*, *MYC*, *ID2*, *ALDH3B2*, and *CAMK2N1*) are significantly associated with ILC ($p \leq 0.05$), while expression levels of *CCND1* and *TMEM176B* have no specific association with either IDC-NST or ILC ($p \geq 0.05$) (Figure 3). In total, there are 22 upregulated genes within KART that contribute to the specific association with the histological subtype ILC (Figure 3 and supplementary material, Figure S2).

KART expression correlates with stromal and immune expression signatures in breast cancer

Next, we investigated whether the KART signature correlates with specific known transcriptional signatures related to estrogen signaling, proliferation, and immunity [33,35–39]. We defined correlations based on Spearman coefficients, where red squares represent anticorrelations and blue squares represent positive correlations, as previously described [48]. In TCGA, KART positively correlated with the stroma derived prognostic predictor (SDPP) and DCN.up stromal signatures (0.51 and 0.41), the Perez immune signature (0.58), and the PLAU invasion signature (0.41), and anticorrelates with ESR1 (−0.34) and AURKA proliferation signatures (−0.43) in IDC-NST (Figure 4 and supplementary material, Figure S3). Similar correlations were found in the METABRIC dataset, but with an additional correlation with the DCN stromal signature (0.49) (Figure 4). Anticorrelations with the ESR1 (−0.53) and AURKA (−0.49) modules were also noted for ILC in TCGA, while the SDPP stromal module (0.41) and the Perez immune module (0.53) positively correlated with KART for ILC (Figure 4 and supplementary material, Figure S3). In METABRIC, KART was anticorrelated with GGI grading (−0.5) while positively correlating with the DCN stroma signature for ILC (0.46) (Figure 4 and supplementary material, Figure S3). We then examined the correlations of the established LobSig signature [41] to the aforementioned signatures including KART. We found a uniform inverse correlation between KART and LobSig for both the TCGA (supplementary material, Figure S3A) and METABRIC datasets (supplementary material,

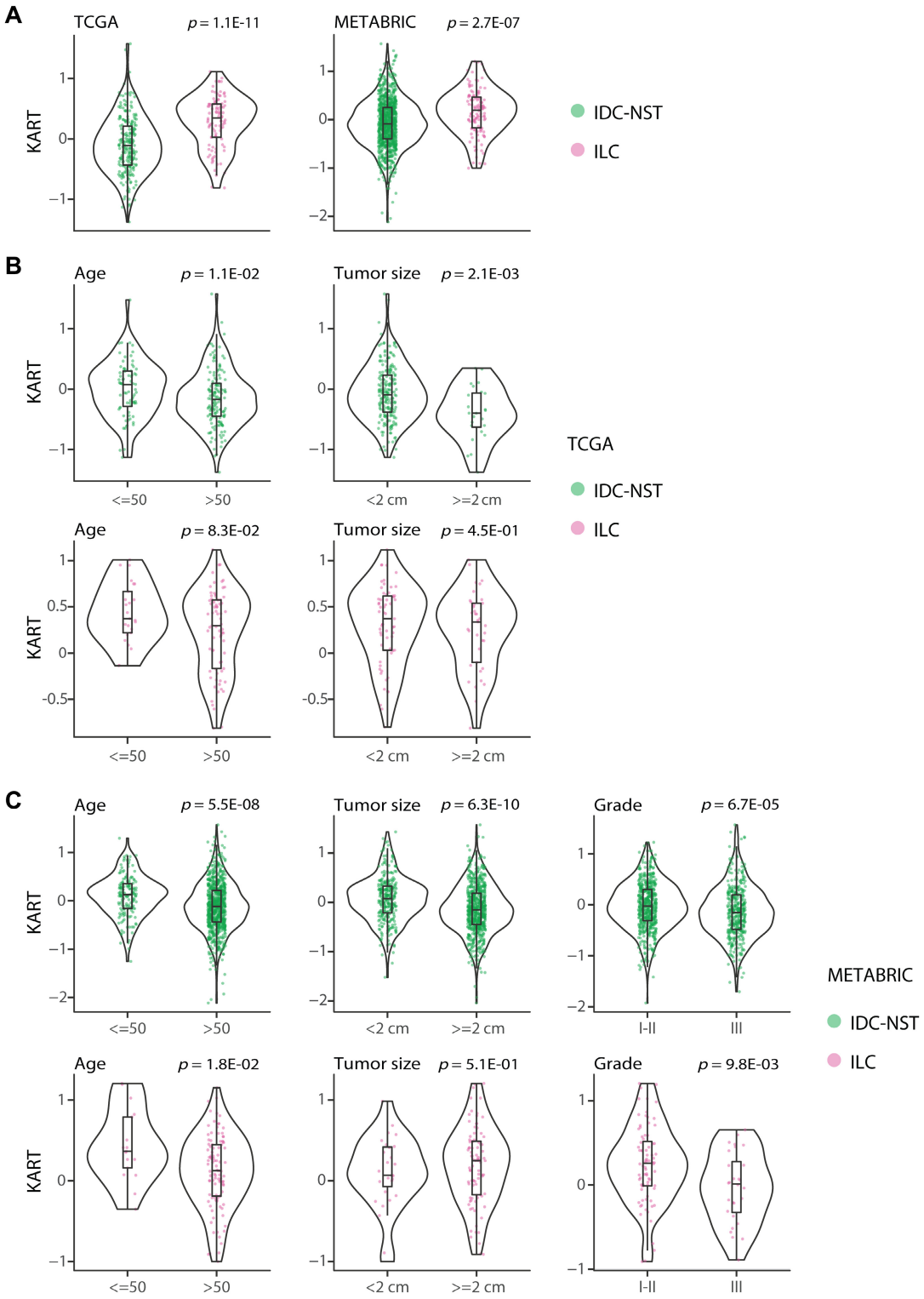


Figure 2. The KART signature associates with the histological subtype ILC, with young patients, small tumors, and low- to intermediate-grade tumors. (A) Shown are the associations of normalized KART expression with IDC-NST and ILC in the TCGA (left) and METABRIC (right) datasets. All points are confirmed ER^{POS} and $HER2^{NEG}$ breast cancers. TCGA: IDC-NST $n = 227$; ILC $n = 98$. (B and C) Associations of normalized KART expression with clinicopathological features in breast cancer. Shown are violin plots for age, tumor size, and grading associations for IDC-NST (green bullets) and ILC (pink bullets) in the TCGA (B) and METABRIC (C) cohorts. All statistics were performed using the Wilcoxon signed-rank test.

Figure S3B). Overall, in both datasets, positive correlation for KART in invasion and immune signatures are linked to negative correlations for LobSig, whereas the negative correlations for KART with proliferation and estrogen signature are positively linked to LobSig (supplementary material, Figure S3).

In short, KART correlates, independently of the histopathological subtype, with stromal, immune,

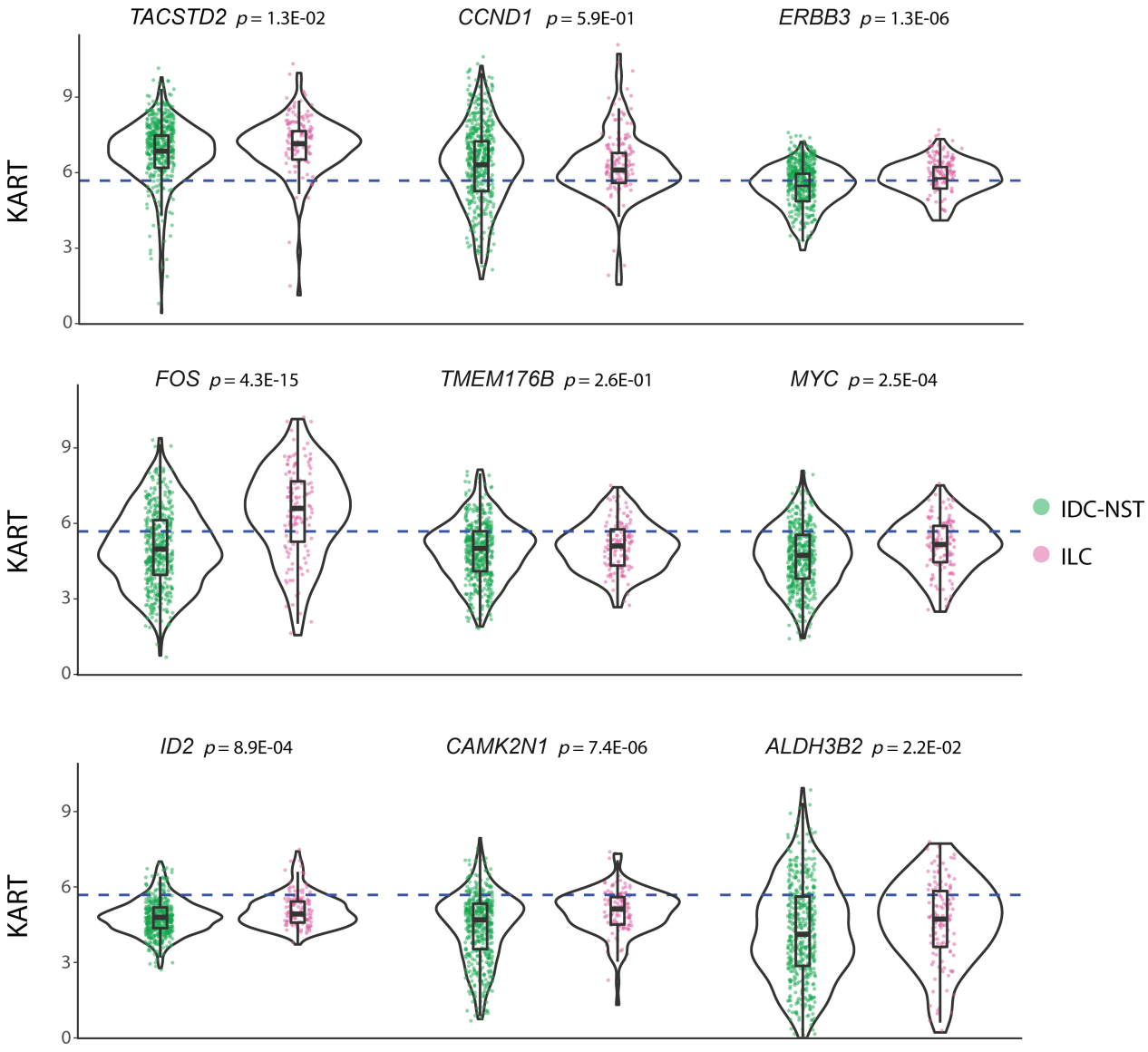


Figure 3. Individual gene contributions to KART signature and their associations to ILC histological phenotype. Violin plots showing nine highest expressed genes within KART in TCGA database in IDC-NST (green bullets) and ILC (pink bullets), ranked on overall mRNA expression levels. The dashed blue lines represent the 97th percentile of the sequenced mRNA. Statistics were performed using the Wilcoxon signed-rank test.

and invasion signatures, but anticorrelated with proliferation, grading, and the estrogen receptor signature.

KART associates with a favorable DFS and OS in breast cancer

The association of the KART signature with the ILC histological type and the fact that ILC is a generally low-grade disease prompted us to assess whether KART was associated with favorable prognosis in breast cancer survival. We observed that KART was prognostic for improved DFS in the TCGA IDC-NST dataset in multivariable analyses (hazard ratio [HR] = 0.50, 95% CI = 0.23–1.09, p = 7.1E-03) (Figure 5A and supplementary material, Table S2) when considering KART expression around the median. The direction of this association was conserved for OS; however, significance was not reached (HR = 0.72, 95% CI = 0.28–1.81, p = 1.5E-01) (Figure 5B and supplementary material, Table S2). For ILC, higher KART expression was significantly associated with better prognosis in multivariable analyses for DFS in TCGA (HR = 0.18, 95% CI = 0.04–0.76, p = 2.5E-02) (Figure 5C and supplementary material, Table S2). This trend was also observed for OS in TCGA (HR = 0.10, 95% CI = 0.02–1.00, p = 7.6E-03) (Figure 5D and supplementary material, Table S2). In the METABRIC dataset, we found that higher KART expression associated with better OS in multivariable analyses for both the IDC-NST (HR = 0.79, 95% CI = 0.66–0.93, p = 1.2E-04) and the ILC (HR = 0.51, 95% CI = 0.29–0.91, p = 3.4E-02) cohorts (Figure 6A–C and supplementary material, Table S2). In short, we found that high KART expression predicted favorable prognosis for

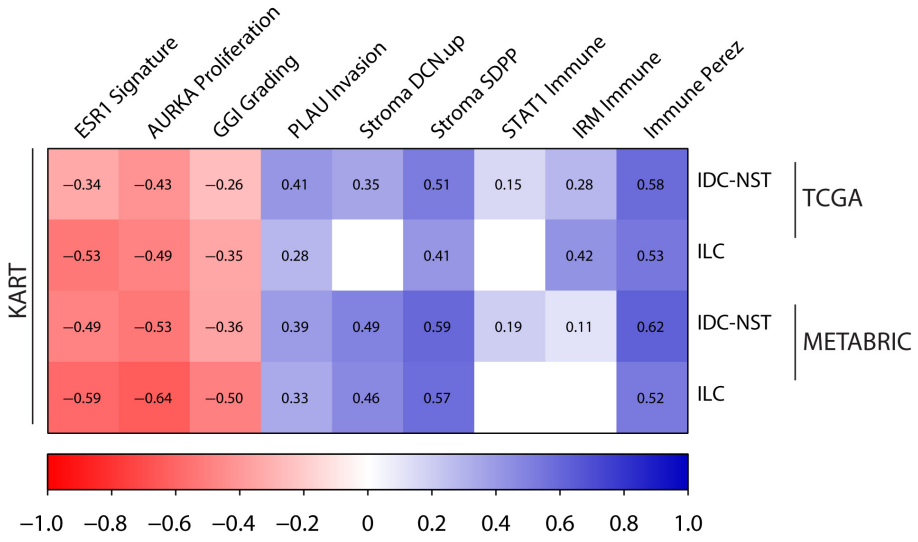


Figure 4. The KART signature positively correlates with stroma, invasion, and immune expression signatures and negatively with proliferation expression profiles. Condensed heat maps showing significant gene expression signature correlations for IDC-NST and ILC in the TCGA and METABRIC databases. Numbers inside squares represent Pearson coefficients. Red indicates negative correlations, blue represents positive correlations, white represents no significant association.

both DFS and OS in IDC-NST and ILC patients. When comparing KART to the performance of the LobSig signature, we again found an inverse association in both datasets. Interestingly, in contrast to KART, we found that high LobSig inversely associated with the histological ILC type when compared to KART, with a lower significance in TCGA ($p = 8.3E-11$ versus $p = 1.1E-11$) and METABRIC ($p = 2.1E-05$, versus $p = 2.7E-07$) (supplementary material, Figures S4A and S5A). Moreover, while high KART associates with good survival, high LobSig associates with poor survival in both the TCGA and METABRIC datasets (supplementary material, Figures S4B and S5B). Although LobSig showed a slightly stronger association with OS in ILC than KART in TCGA (HR = 6.94, 95% CI = 1.75–27.57, $p = 1.7E-04$, versus HR = 0.10, 95% CI = 0.01–0.68, $p = 7.6E-03$) and METABRIC (HR = 2.52, 95% CI = 1.55–4.09, $p = 2.4E-04$ versus HR = 0.44, 95% CI = 0.25–0.76, $p = 3.4E-02$), we observed an inverse but similar significance for OS and DFS between the two signatures (supplementary material, Figure S4B,C). Both LobSig and KART showed an association with DFS in ILC (HR = 5.02, 95% CI = 1.93–13.04, $p = 1.8E-03$ versus HR = 0.15, 95% CI = 0.04–0.62, $p = 2.5E-02$). In short, KART and LobSig showed inverse associations of similar significance with the main histological breast cancer types IDC-NST and ILC and clinicopathological parameters in TCGA and METABRIC datasets.

Discussion

The transcriptional modifier Kaiso has been implicated in cancer due to its role as a transcriptional regulator of genes such as *CCND1* [19,21], *MYC* [19], *WNT11* [10,20], *MMP7* [46,47], and *ID2* [44]. In this study, we combined candidate and established Kaiso target genes that were either detected in the context of anchorage-independent mouse ILC cells or Kaiso targets that were functionally identified and verified. Interestingly, several canonical Kaiso targets (genes regulated through the cKBS consensus) from these studies play a role in the regulation of Wnt signaling. Both canonical and noncanonical Wnt targets are mostly associated with cellular differentiation, development, or specific functions of terminally differentiated cells. Examples are Rapsyn, a critical effector of acetylcholine receptor signals in the neuromuscular junction [49], and Wnt11, Cyclin D1, and Myc, which control a plethora of developmental and differentiation processes in breast and other tissues [50–53].

Although Cyclin D1 is essential for cell cycle progression, it is well known that high expression of Cyclin D1 is strongly associated with low-grade luminal-type breast cancers [54,55]. Given that Cyclin D1 is the second most abundantly expressed cKBS target in all breast cancers, its high expression likely strongly contributes to our finding that KART is positively associated with better OS and DFS in breast cancer regardless of the histological type. The relatively weak individual contribution of Cyclin D1 to histological differentiation between IDC-NST and ILC also corroborates this assumption, although a study by Tobin *et al* indicated that low Cyclin D1 protein expression in ILC associated with improved outcome [56]. In contrast, ID2 showed a lower level of expression (ranked 7/33) but strongly associated with the ILC phenotype. These findings are in line with reports that link ID2 as an inhibitor of ILC cell proliferation while facilitating anoikis resistance, a hallmark of metastatic ILC [44,57].

Concomitant high ID2 and Cyclin D1 expression in ILC [54,56,58,59] might be explained by the finding that

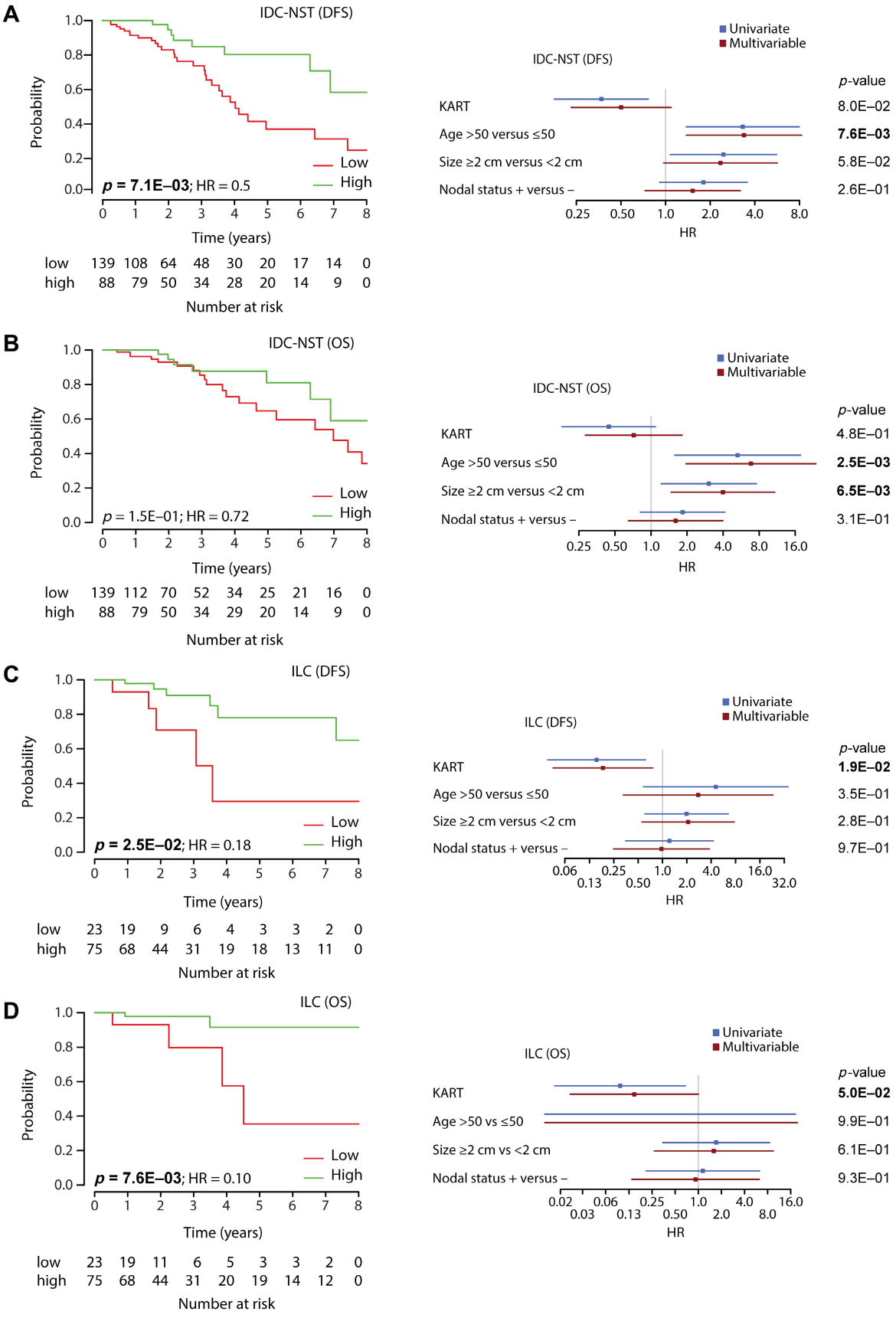
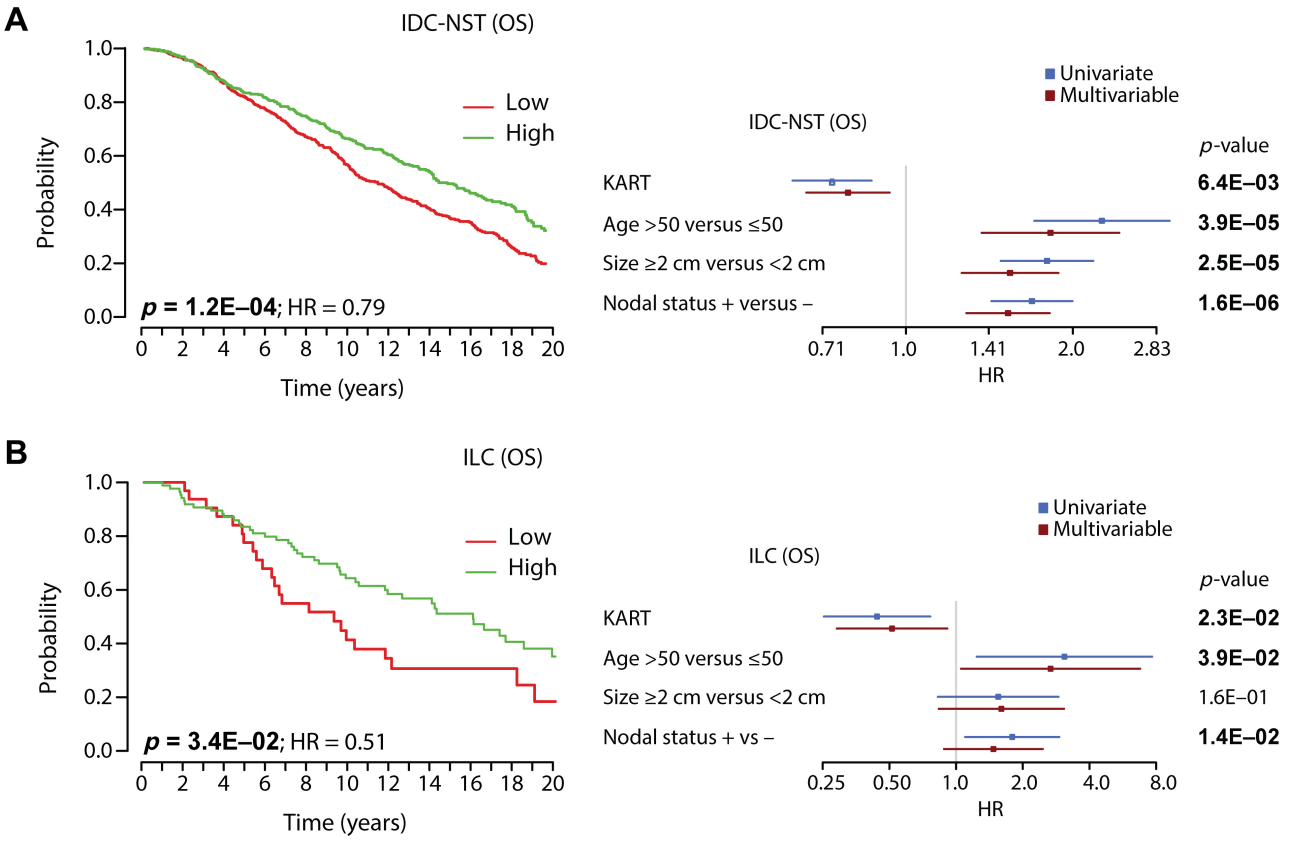


Figure 5. KART expression correlates with DFS and OS for IDC-NST and ILC in TCGA dataset. (A–D) Kaplan–Meier graphs depicting DFS (A and C) and OS (B and D) for IDC-NST (A and B) or ILC patients respectively (C and D) regarding KART expression categorized around median with corresponding forest plots. Significant *p* values are depicted in bold. IDC-NST: $n = 227$ and ILC: $n = 98$.



C

Number at risk	Years	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
IDC-NST	Low	516	507	489	472	441	415	389	362	326	301	263	223	200	177	153	133	112	93	80	56	0
	High	463	456	443	420	398	371	350	329	296	273	235	208	184	160	151	129	109	86	75	59	0
ILC	Low	32	32	31	29	27	24	21	17	17	15	12	11	10	8	7	7	6	5	5	3	0
	High	86	85	79	76	73	69	65	63	57	54	46	41	39	33	29	27	25	20	16	15	0

Figure 6. KART expression associates with long-term OS for IDC-NST and ILC in METABRIC dataset. (A and B) OS Kaplan–Meier curves for IDC-NST (A) and ILC (B) cohorts in METABRIC dataset regarding KART expression categorized around median with corresponding forest plots. Significant *p* values are depicted in bold. (C) Table depicting longitudinal patient numbers of the data shown in (A) and (B). IDC-NST: $n = 979$ and ILC: $n = 118$.

cytosolic ID2 functions as a CDK4/6 antagonist in ILC cells through binding to hypo-phosphorylated Rb and subsequent dampening of cell cycle progression [44]. In this context, ID2 functions as a CDK4/6 inhibitor, which potentially induces a compensatory upregulation of Cyclin D1. Because high Cyclin D1 expression levels mostly associate with favorable short-term prognosis in ILC [54] and CDK4/6 inhibitors induce an upregulation of Cyclin D1 in lobular-type breast cancer cell lines [60], we postulate that the Kaiso target ID2 may propel further compensatory upregulation of Cyclin D1 expression. This codependence may therefore partly underpin the observed associations of the KART signature with good prognosis and low proliferation in ILC.

Our work has identified *ERBB3* as a cKBS Kaiso target gene. Interestingly, somatic oncogenic *ERBB3* mutations have recently been linked to ILC by multiple studies [23,41,61,62]. It is well established that loss of E-cadherin leads to autocrine activation of growth factor receptor signals [3,4]. Since E-cadherin-dependent cell–cell adhesion inhibits activation of multiple receptor tyrosine kinases [3,4,63,64] and ILC is a slow proliferating disease, we hypothesize that ERBB3 mostly propels pro-survival cues in ILC through PI3K/AKT. Because ERBB3 signaling depends on heterodimerization and activation through other ERBB family members [65,66], it is tempting to speculate that the Kaiso target ERBB3 may conspire with low levels of ERBB2 to foster sustained AKT activation and subsequent anchorage independence in metastatic ILC.

Although ILC is unfortunately still mostly defined as a histomorphological breast cancer type, it has become evident that lobular carcinoma is in fact a unique entity within the breast cancer spectrum. Apart from the obvious phenotypical differences, multiple studies have shown that ILC presents a specific genomic and

transcriptional landscape [23,67–69] that drives a distinct biochemistry [3,4,6,70,71]. As a result, the identification of targetable oncological cues has instigated ILC-specific trials that require reproducible inclusion criteria for ILC. Interestingly, a recent multicenter concordance study showed that interobserver agreement for histological differential breast cancer diagnosis for ILC is moderate if no additional immunohistochemistry for E-cadherin status is provided [72]. Given the essential roles of E-cadherin loss within ILC etiology, we think that the KART signature represents a functional supportive tool that, on a biological basis and through unbiased selection and clinical associations, results in a better understanding of ILC biology and, potentially, future predictions regarding treatment responses.

To conclude, we present the cKBS KART, a novel 33-gene signature based on a combined set of functional and experimentally established, as well as candidate canonical Kaiso target genes. High KART expression associates with the ILC breast cancer type and is associated with improved long-term OS and DFS in IDC-NST and ILC.

Acknowledgements

We thank members of the Derksen, Desmedt, and Juliet laboratories for help and critical input. This work was made possible through financial support from the Netherlands Organization for Scientific Research (NWO/ZonMW-VIDI 016.096.318), the Dutch Cancer Society (KWF-UU-2011-5230, KWF-UU-2014-7201 and KWF-UU-2016-10456), and Breast Cancer Now (2018NovPCC1297). F.R. is funded by the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO: 1297322N). This publication is based upon work from COST Action LOBSTERPOT (CA19138), supported by COST (European Cooperation in Science and Technology). COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks https://www.cost.eu.

Author contributions statement

TS, MAKR, TK and PWBD, designed the study, performed experiments, and wrote the manuscript. BB and JD designed and performed the ChIP experiments and analysis. TK and CR performed biochemical experiments. FR and CD designed and performed bioinformatic analysis on breast cancer data. All authors read the manuscript and were given the opportunity to provide input.

Data availability statement

The METABRIC and TCGA data, including both clinical and gene expression data, can be found online on cBioPortal (https://www.cbioportal.org/study/summary?id=brca_metabric) and the GDC data portal (https://portal.gdc.cancer.gov/projects/TCGA-BRCA).

References

1. Bruner HC, Derksen PWB. Loss of E-cadherin-dependent cell-cell adhesion and the development and progression of cancer. *Cold Spring Harb Perspect Biol* 2018; **10:** a029330.
2. McCart Reed AE, Kutasovic JR, Lakhani SR, *et al*. Invasive lobular carcinoma of the breast: morphology, biomarkers and omics. *Breast Cancer Res* 2015; **17:** 12.
3. Teo K, Gómez-Cuadrado L, Tenhagen M, *et al*. E-cadherin loss induces targetable autocrine activation of growth factor signalling in lobular breast cancer. *Sci Rep* 2018; **8:** 15454.
4. Nagle AM, Levine KM, Tasdemir N, *et al*. Loss of E-cadherin enhances IGF1-IGF1R pathway activation and sensitizes breast cancers to anti-IGF1R/InsR inhibitors. *Clin Cancer Res* 2018; **24:** 5165–5177.
5. Hornsveld M, Tenhagen M, van de Ven RA, *et al*. Restraining FOXO3-dependent transcriptional BMF activation underpins tumour growth and metastasis of E-cadherin-negative breast cancer. *Cell Death Differ* 2016; **23:** 1483–1492.
6. Ciriello G, Gatza ML, Beck AH, *et al*. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 2015; **163:** 506–519.
7. Sarrió D, Pérez-Mies B, Hardisson D, *et al*. Cytoplasmic localization of p120ctn and E-cadherin loss characterize lobular breast carcinoma from preinvasive to metastatic lesions. *Oncogene* 2004; **23:** 3272–3283.
8. Dabbs DJ, Bhargava R, Chivukula M. Lobular versus ductal breast neoplasms: the diagnostic utility of p120 catenin. *Am J Surg Pathol* 2007; **31:** 427–437.
9. Schackmann RCJ, van Amersfoort M, Haarhuis JHI, *et al*. Cytosolic p120-catenin regulates growth of metastatic lobular carcinoma through Rock1-mediated anoikis resistance. *J Clin Invest* 2011; **121:** 3176–3188.
10. van de Ven RAH, Tenhagen M, Meuleman W, *et al*. Nuclear p120-catenin regulates the anoikis resistance of mouse lobular breast cancer cells through Kaiso-dependent Wnt11 expression. *Dis Model Mech* 2015; **8:** 373–384.
11. Davis MA, Ireton RC, Reynolds AB. A core function for p120-catenin in cadherin turnover. *J Cell Biol* 2003; **163:** 525–534.
12. Ireton RC, Davis MA, van Hengel J, *et al*. A novel role for p120 catenin in E-cadherin function. *J Cell Biol* 2002; **159:** 465–476.
13. van de Ven RAH, de Groot JS, Park D, *et al*. p120-catenin prevents multinucleation through control of MKLP1-dependent RhoA activity during cytokinesis. *Nat Commun* 2016; **7:** 13874.
14. Daniel JM, Reynolds AB. The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol Cell Biol* 1999; **19:** 3614–3623.
15. Daniel JM, Spring CM, Crawford HC, *et al*. The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res* 2002; **30:** 2911–2919.
16. Prokhortchouk A, Hendrich B, Jørgensen H, *et al*. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* 2001; **15:** 1613–1618.
17. Kelly KF, Spring CM, Otchere AA, *et al*. NLS-dependent nuclear localization of p120ctn is necessary to relieve Kaiso-mediated transcriptional repression. *J Cell Sci* 2004; **117:** 2675–2686.
18. Kelly KF, Otchere AA, Graham M, *et al*. Nuclear import of the BTB/POZ transcriptional regulator Kaiso. *J Cell Sci* 2004; **117:** 6143–6152.

19. Park J, Kim S, Lyons J, *et al*. Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Dev Cell* 2005; **8:** 843–854.
20. Kim SW, Park J-I, Spring CM, *et al*. Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat Cell Biol* 2004; **6:** 1212–1220.
21. Donaldson NS, Pierre CC, Anstey MI, *et al*. Kaiso represses the cell cycle gene cyclin D1 via sequence-specific and methyl-CpG-dependent mechanisms. *PLoS One* 2012; **7:** e50398.
22. Michaut M, Chin S-F, Majewski I, *et al*. Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* 2016; **6:** 18517.
23. Desmedt C, Zoppoli G, Gundem G, *et al*. Genomic characterization of primary invasive lobular breast cancer. *J Clin Oncol* 2016; **34:** 1872–1881.
24. Dreos R, Ambrosini G, Groux R, *et al*. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res* 2017; **45:** D51–D55.
25. Rangwala SH, Kuznetsov A, Ananiev V, *et al*. Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Res* 2020; **31:** 159–169.
26. Howe KL, Achuthan P, Allen J, *et al*. Ensembl 2021. *Nucleic Acids Res* 2020; **49:** D884–D891.
27. Pierre CC, Longo J, Bassey-Archibong BI, *et al*. Methylation-dependent regulation of hypoxia inducible factor-1 alpha gene expression by the transcription factor Kaiso. *Biochim Biophys Acta* 2015; **1849:** 1432–1441.
28. Hindriksen S, Bramer AJ, Truong MA, *et al*. Baculoviral delivery of CRISPR/Cas9 facilitates efficient genome editing in human cells. *PLoS One* 2017; **12:** e0179514.
29. Brinkman EK, Chen T, Amendola M, *et al*. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 2014; **42:** e168.
30. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490:** 61–70.
31. Curtis C, Shah SP, Chin S-F, *et al*. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012; **486:** 346–352.
32. Gao J, Aksoy BA, Dogrusoz U, *et al*. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013; **6:** pl1.
33. Desmedt C, Haibe-Kains B, Wirapati P, *et al*. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res* 2008; **14:** 5158–5165.
34. Farmer P, Bonnefoi H, Anderle P, *et al*. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med* 2009; **15:** 68–74.
35. Sotiriou C, Wirapati P, Loi S, *et al*. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006; **98:** 262–272.
36. Finak G, Bertos N, Pepin F, *et al*. Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* 2008; **14:** 518–527.
37. Perez EA, Thompson EA, Ballman KV, *et al*. Genomic analysis reveals that immune function genes are strongly linked to clinical outcome in the north central cancer treatment group N9831 adjuvant trastuzumab trial. *J Clin Oncol* 2015; **33:** 701–708.
38. Teschendorff AE, Miremadi A, Pinder SE, *et al*. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol* 2007; **8:** R157.
39. Tamborero D, Rubio-Perez C, Muiños F, *et al*. A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations. *Clin Cancer Res* 2018; **24:** 3717–3728.
40. Gu-Trantien C, Loi S, Garaud S, *et al*. CD4+ follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest* 2013; **123:** 2873–2892.
41. Mccart Reed AEM, Lal S, Kutasovic JR, *et al*. LobSig is a multigene predictor of outcome in invasive lobular carcinoma. *NPJ Breast Cancer* 2019; **5:** 18.
42. Ignatiadis M, Singhal SK, Desmedt C, *et al*. Gene modules and response to neoadjuvant chemotherapy in breast cancer subtypes: a pooled analysis. *J Clin Oncol* 2012; **30:** 1996–2004.
43. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2018. Available from: https://www.R-project.org.
44. Rätze MAK, Koorman T, Sijnesael T, *et al*. Loss of E-cadherin leads to Id2-dependent inhibition of cell cycle progression in metastatic lobular breast cancer. *Oncogene* 2022; **41:** 3507–3509.
45. Pozner A, Terooatea TW, Buck-Koehntop BA. Cell-specific Kaiso (ZBTB33) regulation of cell cycle through cyclin D1 and cyclin E1. *J Biol Chem* 2016; **291:** 24538–24550.
46. Barrett CW, Smith JJ, Lu LC, *et al*. Kaiso directs the transcriptional corepressor MTG16 to the Kaiso binding site in target promoters. *PLoS One* 2012; **7:** e51205.
47. Ogden SR, Wroblewski LE, Weydig C, *et al*. p120 and Kaiso regulate *Helicobacter pylori*-induced expression of matrix metalloproteinase-7. *Mol Biol Cell* 2008; **19:** 4110–4121.
48. Alday-Parejo B, Richard F, Wörthmüller J, *et al*. MAGI1, a new potential tumor suppressor gene in estrogen receptor positive breast cancer. *Cancer* 2020; **12:** 223.
49. Rodova M, Kelly KF, VanSaun M, *et al*. Regulation of the rapsyn promoter by Kaiso and delta-catenin. *Mol Cell Biol* 2004; **24:** 7188–7196.
50. Christiansen J, Monkley S, Wainwright B. Murine WNT11 is a secreted glycoprotein that morphologically transforms mammary epithelial cells. *Oncogene* 1996; **12:** 2705–2711.
51. Sicinski P, Donaher JL, Parker SB, *et al*. Cyclin D1 provides a link between development and oncogenesis in the retina and breast. *Cell* 1995; **82:** 621–630.
52. Fantl V, Stamp G, Andrews A, *et al*. Mice lacking cyclin D1 are small and show defects in eye and mammary gland development. *Gene Dev* 1995; **9:** 2364–2372.
53. Godeau F, Persson H, Gray HE, *et al*. C-myc expression is dissociated from DNA synthesis and cell division in *Xenopus oocyte* and early embryonic development. *EMBO J* 1986; **5:** 3571–3577.
54. van Diest PJ, Michalides RJ, Jannink L, *et al*. Cyclin D1 expression in invasive breast cancer. Correlations and prognostic value. *Am J Pathol* 1997; **150:** 705–711.
55. El-Hafez AA, Shawky AEA, Hasan B. Cyclin D1 overexpression associates with favourable prognostic factors in invasive breast carcinoma. *Cancer Biomark* 2012; **12:** 149–154.
56. Tobin NP, Lundgren KL, Conway C, *et al*. Automated image analysis of cyclin D1 protein expression in invasive lobular breast carcinoma provides independent prognostic information. *Human Pathol* 2012; **43:** 2053–2061.
57. Derksen PWB, Liu X, Saridin F, *et al*. Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis. *Cancer Cell* 2006; **10:** 437–449.
58. Soslow RA, Carlson DL, Horenstein MG, *et al*. A comparison of cell cycle markers in well-differentiated lobular and ductal carcinomas. *Breast Cancer Res Treat* 2000; **61:** 161–170.
59. Stighall M, Manetopoulos C, Axelson H, *et al*. High ID2 protein expression correlates with a favourable prognosis in patients with primary breast cancer and reduces cellular invasiveness of breast cancer cells. *Int J Cancer* 2005; **115:** 403–411.
60. O'Brien N, Conklin D, Beckmann R, *et al*. Preclinical activity of abemaciclib alone or in combination with antimitotic and targeted therapies in breast cancer. *Mol Cancer Ther* 2018; **17:** 897–907.
61. Harrison BT, Nakhlis F, Dillon DA, *et al*. Genomic profiling of pleomorphic and florid lobular carcinoma in situ reveals highly recurrent ERBB2 and ERRB3 alterations. *Mod Pathol* 2020; **17:** 1287.

62. Shamir ER, Chen Y-Y, Krings G. Genetic analysis of pleomorphic and florid lobular carcinoma in situ variants: frequent ERBB2/ERBB3 alterations and clonal relationship to classic lobular carcinoma in situ and invasive lobular carcinoma. *Mod Pathol* 2020; **33:** 1078–1091.
63. Sullivan B, Light T, Vu V, *et al*. Mechanical disruption of E-cadherin complexes with epidermal growth factor receptor actuates growth factor-dependent signaling. *Proc Natl Acad Sci U S A* 2022; **119:** e2100679119.
64. Qian X, Karpova T, Sheppard AM, *et al*. E-cadherin-mediated adhesion inhibits ligand-dependent activation of diverse receptor tyrosine kinases. *EMBO J* 2004; **23:** 1739–1748.
65. Berger MB, Mendrola JM, Lemmon MA. ErbB3/HER3 does not homodimerize upon neuregulin binding at the cell surface. *FEBS Lett* 2004; **569:** 332–336.
66. Choi B, Cha M, Eun GS, *et al*. Single-molecule functional anatomy of endogenous HER2-HER3 heterodimers. *Elife* 2020; **9:** e53934.
67. Desmedt C, Zoppoli G, Sotiriou C, *et al*. Transcriptomic and genomic features of invasive lobular breast cancer. *Semin Cancer Biol* 2017; **44:** 98–105.
68. Weigelt B, Geyer FC, Natrajan R, *et al*. The molecular underpinning of lobular histological growth pattern: a genome-wide transcriptomic analysis of invasive lobular carcinomas and grade- and molecular subtype-matched invasive ductal carcinomas of no special type. *J Pathol* 2010; **220:** 45–57.
69. Pareja F, Ferrando L, Lee SSK, *et al*. The genomic landscape of metastatic histologic special types of invasive breast cancer. *NPJ Breast Cancer* 2020; **6:** 53.
70. Tasdemir N, Ding K, Savariau L, *et al*. Proteomic and transcriptomic profiling identifies mediators of anchorage-independent growth and roles of inhibitor of differentiation proteins in invasive lobular carcinoma. *Sci Rep* 2020; **10:** 11487.
71. Tasdemir N, Bossart EA, Li Z, *et al*. Comprehensive phenotypic characterization of human invasive lobular carcinoma cell lines in 2D and 3D cultures. *Cancer Res* 2018; **78:** 6209–6222.
72. Christgen M, Kandt LD, Antonopoulos W, *et al*. Inter-observer agreement for the histological diagnosis of invasive lobular breast carcinoma. *J Pathol Clin Res* 2021; **8:** 191–205.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. Genomic confirmation of Kaiso knockout

Figure S2. Individual gene contributions to KART signature and their associations to ILC histological phenotype

Figure S3. Full KART and LobSig signature correlations

Figure S4. Comparative correlation of KART and LobSig with clinicopathological parameters for ILC in TCGA dataset

Figure S5. Comparative correlation of KART and LobSig with clinicopathological parameters for ILC in METABRIC dataset

Table S1. Primer pairs used

Table S2. Full hazard ratio statistics of KART associations with DFS and OS in TCGA and METABRIC