

Supporting Information

Profile integration in the TCGA IDC cohort. To establish the integrated miRNA/mRNA expression profile we normalized mRNAs as RPKM(1) and miRNAs as reads per million of total aligned miRNA reads. Two expression filters were used to include only the genes with highest variability in the integrated profile. One filter was towards minimum fold change: 20% of a gene's expression values must have at least 1.5 fold-change in either direction from the gene's median value. The second filter was towards log-intensity variation: the variance of the \log_2 normalized reads for each gene was compared to the median of all the variances and only the genes more variable than the median gene passed the filter ($p < 0.05$). After these two variability filters were used, 7735 mRNAs and 247 miRNAs were retained in the integrated RNA profile (RNA expression matrix in Supplemental Information). The Catalogue Of Somatic Mutations In Cancer database (ver. 60)(2) was used to identify the genes which are known to harbor functional somatic mutations in cancer. The breast cancer dataset was incremented with the highly related ovarian cancer dataset in order to evaluate a large tumors sample size. The genes with at least two validated somatic mutations resulting in alteration of the primary protein structure were identified. DNA methylation (meDNA) was studied using the Infinium 450K platform on 296 patients from the same IDC cohort. The M value,(3) i.e. the \log_2 ratio of the intensities of a methylated probe versus its corresponding un-methylated probe, was used to measure CpG methylation (DNA methylation dataset in Supporting Information).

Risk stratification and ROC curves. The ability of the RNA models to predict outcome was assessed by comparing the AUC of the respective ROC curves. Analysis of area under curve (AUC) for the Receiver Operating Characteristic (ROC) curve was conducted using the *survivalROC* package in R, allowing for time dependent ROC curve estimation with censored data(4). Since in all of the survival analyses, fewer events occurred after 60 months (see Kaplan-Meier curves), we compared the ability of models to predict outcome at, and around, this time point. The ROC curve plots the true-positive vs. false-positive predictions, thus higher AUC indicates better model performance (with AUC=0.5 indicating random performance). RNA risk scores and groups (risk-high or -low defined above) were based on weightings in the linear risk predictor.

Biological processes associated to risk genes. The functional association of the genes in the integrated signature with key clinical covariates was assessed in the TCGA cohort. *DAAM1*, thought to function as a scaffolding protein for the Wnt-induced assembly of a disheveled (Dvl)-Rho complex, was the prognostic gene harboring the highest correlation with lymph node involvement (N stage) in the TCGA dataset (Spearman correlation test, $p < 0.001$, FDR=0.01). *CXCR7* and *FAM199X* were also strongly associated to N stage ($p < 0.01$, FDR=0.1). *MCM10* was down-regulated in tumors with mutations in the *PIK3CA/AKT1/PTEN* axis ($p\text{-value} < 1E-07$) and *TRIM23* was down-regulated in *TP53* mutants ($p\text{-value} < 1E-07$). Among these genes, *CXCR7* was retained in the final multivariate model where N stage, *PIK3CA* axis mutations and ER status were also present (Table S6).

The BiNGO(5) plugin in Cytoscape(6) was used to retrieve the GO annotations for the BC risk genes and propagate them upwards through the GO hierarchy. The

hypergeometric test, in which sampling occurs without replacement, was used to assess the enrichment of gene ontology (GO) terms in the gene-set in the form of a p-value. The GO p-values were corrected using Benjamini and Hochberg method(7). Lipid modification (*PIK3CA*, *SMG1* and *CPT1A*) and regulation of nitric oxide biosynthesis (*HSP90AA1* and *GLA*) had corrected p-values lower than 0.05. There were no other functional enrichments when the coding genes in the risk predictor were considered together. This finding was in agreement with the risk genes impacting on a range of independent pathways and cellular processes.

We then investigated each single gene in detail, whether an mRNA or a miRNA, by performing GO analysis on the mRNAs with whom it correlated in the integrated RNA profile. Genes, involved in mitotic cell cycle and nuclear division, were positively associated with *miR-484*. *miR-328* was correlated with genes of M phase and DNA repair, *miR-874* with genes involved in cell adhesion. *miR-484* was negatively correlated with genes in morphogenesis and angiogenesis, and also in the development of epidermis and the assembly of hemidesmosomes, which anchor epithelial cells to extracellular matrix components such as the basal laminae. *CPT1A* was associated with the mammary gland branching involved in thelarche, the onset of postnatal breast development, usually occurring at the beginning of puberty, as well as Ral GTPase regulation. The expression of *PIK3CA* was associated with activation of protein phosphorylation and transcription initiation and with the repression of mitochondrial ATP synthesis coupled proton transport.

Table S1. The TCGA IDC Cohort of Patients with Invasive Ductal Carcinoma

Covariates		n=466	
Disease Stage, no (%)	I	94	20.2
	IIA	157	33.8
	IIB	97	20.9
	IIIA	65	14.0
	IIIB	12	2.6
	IIIC	17	3.7
	IV	12	2.6
Lymph Node Involvement, no (%)	N0	197	44.4
	N1	168	37.8
	N2	55	12.4
	N3	24	5.4
M stage, no (%)	M0	419	89.9
	M1	12	2.6
T stage, no (%)	T1	146	31.3
	T2	267	56.2
	T3	38	8.2
	T4	18	3.6
ER, no (%)	Negative	110	24.3
	Positive	342	75.7
PR, no (%)	Negative	151	33.4
	Positive	301	66.6
HER2, no (%)	Negative	255	79.7
	Positive	65	20.3
TNBC, no (%)	Non TN	256	80.0
	TN	64	20.0
Age, years, no (%)	<=45	102	21.9
	46-55	105	22.5
	56-65	141	30.3
	>65	118	25.3
TP53 Mutation, no (%)	Mutated	165	35.4
	Wild-type	301	64.6
PIK3CA Mutation, no (%)	Mutated	142	30.5
	Wild-type	324	69.5
Molecular subtype, no (%)	Basal-like	104	22.4
	HER2-enriched	49	10.5
	Luminal A	110	23.7
	Luminal B	148	31.8

ER, estrogen receptor; PR, progesterone receptor

Table S2. The negative correlation between mRNA expression and CpG DNA methylation of *PIK3CA* (FDR <0.001)

Correlation Coefficient (Spearman)	Parametric P-value	FDR	Unique ID (Illumina 450K)	Genomic Coordinates
-0.217	3.06E-05	0.0004	cg01792902	chr3:178864927-178864976
-0.249	1.70E-06	6.53E-05	cg22806557	chr3:178865282-178865331
-0.242	8.76E-05	0.0009	cg00658016	chr3:178865679-178865728
0.256	8.00E-07	4.30E-05	cg19522332	chr3:178866005-178866054
-0.337	< 1e-07	< 1e-07	cg22384366	chr3:178867005-178867054
-0.289	2.00E-07	2.31E-05	cg01753796	chr3:178867106-178867155

Table S3. DNA methylated CpG sites associated with OS in the TCGA cohort.

Gene Symbol	Probe	Cox Regression p value	Hazard Ratio
<i>ADAT1</i>	cg02576610	0.033	0.60
<i>ANKRD52</i>	cg10803923	4.127E-04	0.18
<i>ANKRD52</i>	cg23759554	0.012	0.58
<i>APC</i>	cg11057897	0.004	0.47
<i>APC</i>	cg18315896	0.029	2.75
<i>APC</i>	cg25922032	0.036	2.94
<i>ARSK / TTC37</i>	cg01559436	0.009	6.02
<i>BIRC6</i>	cg12233946	0.003	0.35
<i>BIRC6</i>	cg20351608	0.015	3.70
<i>BIRC6</i>	cg11971852	0.020	0.45
<i>C10orf18</i>	cg09262442	0.013	0.65
<i>C10orf18</i>	cg08452525	0.015	0.52
<i>C10orf18</i>	cg21302771	0.032	1.85
<i>C2CD2</i>	cg00054774	0.021	0.76
<i>C2CD2</i>	cg04723534	0.049	0.83
<i>CEP350</i>	cg21944557	0.014	2.54
<i>CHD9</i>	cg04770195	0.001	0.57
<i>CHD9</i>	cg05788994	0.008	2.48
<i>CHD9</i>	cg08944236	0.028	0.78
<i>CHM</i>	cg19000468	0.049	0.53
<i>CPT1A</i>	cg03855388	8.100E-06	0.63
<i>CPT1A</i>	cg18262591	1.250E-04	0.57
<i>CPT1A</i>	cg20629021	1.751E-04	0.49
<i>CPT1A</i>	cg01926073	5.880E-04	0.41
<i>CPT1A</i>	cg20809737	6.307E-04	0.58
<i>CPT1A</i>	cg13491471	0.002	0.58
<i>CPT1A</i>	cg22911054	0.004	0.69
<i>CPT1A</i>	cg19964581	0.005	0.55
<i>CPT1A</i>	cg00574958	0.008	0.68
<i>CPT1A</i>	cg12031060	0.008	0.55
<i>CPT1A</i>	cg09737197	0.009	0.78
<i>CPT1A</i>	cg26192826	0.011	1.53
<i>CPT1A</i>	cg10970946	0.013	0.56
<i>CPT1A</i>	cg19081843	0.014	2.36
<i>CPT1A</i>	cg14249520	0.017	0.70
<i>CPT1A</i>	cg17058475	0.029	0.77
<i>CPT1A</i>	cg23756264	0.030	0.58
<i>CPT1A</i>	cg11867574	0.032	0.58
<i>CPT1A</i>	cg10553894	0.043	0.75
<i>CPT1A</i>	cg26989316	0.044	0.70

DAAM1	cg18819791	0.003	0.63
DAAM1	cg17326597	0.017	0.71
DAAM1	cg19768143	0.020	0.65
DAAM1	cg23315932	0.021	0.74
DAAM1	cg05342987	0.022	1.98
DAAM1	cg23862792	0.027	0.34
DIP2B	cg16248432	2.832E-04	0.60
DIP2B	cg23925540	0.009	0.52
DIP2B	cg05827208	0.013	0.16
DIP2B	ch.12.1034102F	0.038	0.51
DIP2B	cg05512157	0.047	0.72
DPY19L3	cg09577932	0.022	0.64
DPY19L3	cg11416386	0.026	0.24
DPY19L3	cg02871028	0.033	3.71
FAM91A1	cg00189906	6.218E-04	0.13
GADD45GIP1	cg03539717	6.240E-05	0.41
GADD45GIP1	cg14393411	0.029	0.24
GBE1	cg02980152	0.011	3.15
GBE1	cg16472210	0.014	9.15
GBE1	cg03785807	0.036	0.40
GMCL1	cg04590956	0.019	0.25
GMCL1	cg19393587	0.021	0.65
GMCL1	cg19598605	0.036	0.58
GOLGA4	cg12435059	0.001	0.39
GOLGA4	cg09456388	0.009	7.80
HSP90AA1	cg11789371	0.021	0.76
MAN2A1	cg23500208	0.006	0.66
MAN2A1	cg07541465	0.015	1.59
MAN2A1	cg04020816	0.022	1.48
MAN2A1	cg19117777	0.025	3.06
MAN2A1	cg10569493	0.041	0.75
ME1	cg16701167	0.009	0.75
ME1	cg01354879	0.011	0.71
MPND	cg12985235	0.008	0.59
NCOA2	cg22304838	2.726E-04	0.41
NCOA2	cg16282160	8.596E-04	0.55
NCOA2	cg13312403	0.004	0.48
NCOA2	cg24536624	0.005	0.40
NCOA2	cg02492405	0.006	1.45
NCOA2	cg24857609	0.020	0.48
NCOA2	cg11197908	0.030	0.50
NCOA2	cg20679659	0.044	0.37
NCOA2	cg20699548	0.049	0.71

OTUD6B	cg06670766	8.910E-04	0.22
OTUD6B	cg25987774	0.017	3.74
OTUD6B	cg06216901	0.018	0.24
OTUD6B	cg23863900	0.020	2.32
OTUD6B	cg18888595	0.042	0.42
PDSS2	cg26954751	8.150E-04	0.59
PDSS2	cg06931193	0.004	2.19
PDSS2	cg14957497	0.005	0.13
PDSS2	cg01959896	0.006	2.72
PDSS2	cg01917202	0.007	0.13
PDSS2	cg26312839	0.009	0.15
PDSS2	cg09699384	0.020	0.21
PDSS2	cg21122684	0.033	0.23
PIGP / TTC3	cg27382599	0.003	0.14
PIGP / TTC3	cg14220272	0.034	0.42
PIK3CA	cg13435326	0.013	0.65
PIK3CA	cg22384366	0.029	0.66
PIK3CA	cg01753796	0.037	0.75
PIK3CA	cg22806557	0.038	0.39
RC3H2	cg14500628	0.019	1.95
SCYL2 / DEPDC4	cg18548043	0.012	20.06
SCYL2 / DEPDC4	cg02430089	0.016	0.33
SCYL2 / DEPDC4	cg11940847	0.041	5.80
SMG1	cg01079632	0.007	0.58
SMG1	cg27302229	0.009	4.02
SMG1	cg03614442	0.047	0.40
THAP2 / ZFC3H1	cg19966154	0.004	2.88
THAP2 / ZFC3H1	cg21830480	0.041	0.57
TOPBP1	cg12655621	0.001	3.38
TOPBP1	cg13024018	0.004	3.72
TRIM23 / C5orf44	cg08307039	0.005	0.21
TTC3	cg00302270	0.001	0.48
UBR5	cg12517050	1.500E-06	0.31
UBR5	cg14897263	0.018	0.71
UBR5	cg15001633	0.021	1.71
UBXN7	cg15925179	0.031	0.44
UBXN7	cg03191919	0.042	0.77
WDR20/HSP90AA1	cg04414295	0.009	1.73
ZFC3H1 / THAP2	cg07821172	0.008	0.37
ZFC3H1 / THAP2	cg03191116	0.028	0.35
ZNF292	cg15511114	0.044	2.10

* Cox regression, p value <0.05.

Table S4. Twenty four mRNAs and ten miRNAs commonly associated with clinical outcome across independent IDC subclasses.

Gene	Validated Subclasses
<i>ADAT1</i>	[StI, StIIA]
<i>ANKRD52</i>	[StI, StIIA, StIIB], [PI3K, TP53 PIK3CA]
<i>BIRC6</i>	[StI, StIIA]
<i>C10orf18</i>	[StI, StIIA]
<i>C2CD2</i>	[LumA, BL]
<i>CHD9</i>	[StI, StIIB]
<i>CHM</i>	[StI, StIIA]
<i>CPT1A</i>	[LumA, LumB], [LumB Marg Neg, BL Marg Neg], [N+ Horm Rec+ ERBB2- Marg Neg, TNBC Marg Neg], [N0, N+], [StIIB, StIIIA], [PI3K, noMajorMut], [Low MutRate, Med MutRate]
<i>DAAM1</i>	[StI, StIIA]
<i>DIP2B</i>	[StIIB, StIIIA], [TP53, PI3K]
<i>DPY19L3</i>	[StI, StIIB, StIIIA]
<i>FAM91A1</i>	[N0, N+]
<i>GMCL1</i>	[StIIA, StIIB, StIIIA], [PI3K, noMajorMut]
<i>hsa-miR-103</i>	[ERBB2+, TNBC]
<i>hsa-miR-1307</i>	[TP53, PI3K]
<i>hsa-miR-148b</i>	[TP53, MAPK]
<i>hsa-miR-324</i>	[Low MutRate, Med MutRate]
<i>hsa-miR-326</i>	[N0, N+]
<i>hsa-miR-328</i>	[N0, N+], [StI, StIIB, StIIIA], [Low MutRate, Med MutRate]
<i>hsa-miR-365</i>	[LumA, BL]
<i>hsa-miR-484</i>	[N0, N+], [N0 Marg Neg, N+ Marg Neg]
<i>hsa-miR-874</i>	[N0, N+], [N0 Marg Neg, N+ Marg Neg], [StI, StIIIA], [N1, N23]
<i>hsa-miR-93</i>	[N0 Marg Neg, N+ Marg Neg]
<i>ME1</i>	[LumA, LumB]
<i>NCOA2</i>	[LumB Marg Neg, BL Marg Neg]
<i>OTUD6B</i>	[N0, N+], [Low MutRate, Med MutRate]
<i>PDSS2</i>	[N0, N+], [Low MutRate, High MutRate]
<i>PIK3CA</i>	[LumA, BL], [Low MutRate, High MutRate]
<i>SMG1</i>	[StI, StIIA]
<i>TRIM23</i>	[StIIA, StIIB]
<i>TTC3</i>	[StI, StIIB]
<i>UBR5</i>	[StIIA, StIIB]
<i>UBXN7</i>	[HER2 Marg Neg, BL Marg Neg], [StI, StIIA], [PI3K, noMajorMut], [Low MutRate, High MutRate]
<i>ZFC3H1</i>	[StI, StIIA]

* The coding genes were restricted further by DNA methylation/OS analysis and presence of somatic mutations. Square brackets indicate the independent IDC subclasses used for validation. Marg Neg=Margin Negative. Horm Rec+ means ER+ and/or PR+ tumors. Mutation rate: Low <25 mutations in exome, 25<=Medium<=50, High>50. Mutations: PI3K (PIK3CA, AKT1, PTEN, PIK3R1), TP53 PIK3CA are double mutants. noMajorMut= others than PI3K, TP53, MAPK and GATA3.

Table S5. The Integrated miRNA/mRNA Linear Risk Predictor in the TCGA Cohort

RNA	p-value	Univariate HR (95% CI)	Multivariate HR (95% CI)	Weight (w_i)
ACSL1	0.002	1.43 (1.14-1.79)	0.84 (0.58-1.21)	0.151
ADAT1	<0.001	1.69 (1.26-2.26)	0.76 (0.41-1.43)	0.126
ANKRD52	<0.001	1.94 (1.34-2.83)	2.43 (1.09-5.39)	0.004
BIRC6	<0.001	2.23 (1.50-3.32)	2.19 (0.75-6.34)	0.062
CPT1A	<0.001	1.78 (1.38-2.30)	1.20 (0.79-1.84)	0.056
CXCR7	0.021	1.28 (1.04-1.58)	1.47 (1.03-2.10)	0.267
DAAM1	0.001	1.64 (1.22-2.22)	0.93 (0.55-1.57)	0.142
DIP2B	<0.001	2.55 (1.65-3.94)	1.10 (0.49-2.45)	0.078
FAM199X	<0.001	2.62 (1.66-4.14)	1.59 (0.59-4.27)	0.068
FAM91A1	<0.001	1.62 (1.24-2.12)	0.73 (0.35-1.50)	-0.021
FRZB	0.012	0.78 (0.64-0.95)	0.79 (0.57-1.10)	-0.239
GLA	0.005	0.58 (0.39-0.85)	0.35 (0.19-0.64)	-0.001
GMCL1	<0.001	2.23 (1.47-3.38)	0.34 (0.11-1.06)	0.076
HRASLS	0.015	1.31 (1.05-1.63)	2.24 (1.50-3.34)	0.165
hsa-miR-103	0.003	1.69 (1.19-2.40)	1.64 (0.82-3.27)	-0.090
hsa-miR-1307	0.001	1.44 (1.16-1.80)	0.63 (0.39-1.01)	0.090
hsa-miR-148b	<0.001	2.06 (1.40-3.02)	1.01 (0.54-1.88)	0.026
hsa-miR-328	<0.001	1.80 (1.36-2.38)	1.80 (0.96-3.35)	0.221
hsa-miR-484	<0.001	1.86 (1.37-2.52)	1.72 (0.92-3.22)	0.138
hsa-miR-874	<0.001	1.89 (1.41-2.53)	1.15 (0.70-1.89)	0.275
hsa-miR-93	0.02	1.40 (1.05-1.85)	0.87 (0.43-1.74)	-0.069
HSP90AA1	<0.001	1.90 (1.37-2.64)	1.05 (0.61-1.79)	0.117
MCM10	0.038	1.25 (1.01-1.54)	0.93 (0.56-1.55)	0.029
ME1	<0.001	1.37 (1.15-1.63)	1.05 (0.80-1.39)	0.035
NDRG1	0.035	1.21 (1.01-1.44)	0.95 (0.68-1.32)	0.040
NOTCH2NL	0.203	1.28 (0.88-1.87)	2.03 (1.12-3.67)	0.024
OTUD6B	0.001	1.54 (1.19-2.00)	0.90 (0.47-1.71)	-0.086
PDSS2	<0.001	1.80 (1.35-2.40)	1.43 (0.85-2.39)	0.089
PGK1	<0.001	2.47 (1.73-3.52)	2.52 (1.24-5.13)	0.144
PIK3CA	<0.001	2.09 (1.58-2.76)	2.11 (1.31-3.39)	0.062
PTAR1	0.001	1.84 (1.26-2.67)	1.17 (0.52-2.65)	-0.010
SMG1	0.002	1.86 (1.26-2.75)	0.55 (0.20-1.49)	0.029
TRIM23	0.001	1.77 (1.25-2.51)	1.56 (0.66-3.68)	0.034
TTC3	<0.001	2.25 (1.43-3.53)	1.12 (0.53-2.38)	0.064
UBR5	0.001	1.74 (1.25-2.41)	1.71 (0.70-4.15)	-0.036
UBXN7	<0.001	1.74 (1.31-2.32)	0.41 (0.18-0.96)	0.060
ZFC3H1	<0.001	1.88 (1.38-2.57)	1.33 (0.68-2.61)	0.019

*A new sample was predicted as high risk if its prognostic index was larger than 0.033. The prognostic index was computed by the formula $\sum_i w_i x_i - 8.877$ where w_i and x_i were respectively the weight and \log_2 RNA normalized reads for the i -th gene. The p-value was from the univariate Cox regression.

Table S6. Multivariate Cox proportional hazards model for OS in IDC.

Steps	B	SE	Wald	df	Sig.	HR	95.0% CI for HR	
							Lower	Upper
Step 1								
ACSL1	-.006	.272	.000	1	.983	.994	.583	1.694
ADAT1	-.268	.393	.464	1	.496	.765	.354	1.652
ANKRD52	.361	.498	.527	1	.468	1.435	.541	3.805
BIRC6	1.140	.716	2.537	1	.111	3.127	.769	12.715
CPT1A	.495	.280	3.134	1	.077	1.640	.948	2.837
CXCR7	.704	.269	6.871	1	.009	2.022	1.194	3.423
DAAM1	-.368	.362	1.032	1	.310	.692	.340	1.408
DIP2B	-.020	.530	.001	1	.970	.980	.347	2.769
FAM199X	-.457	.639	.512	1	.474	.633	.181	2.215
FAM91A1	-.018	.451	.002	1	.969	.982	.406	2.376
FRZB	-.215	.228	.885	1	.347	.807	.516	1.262
GLA	-1.107	.469	5.567	1	.018	.331	.132	.829
GMCL1	-1.394	.714	3.810	1	.051	.248	.061	1.006
HRASLS	.611	.276	4.895	1	.027	1.841	1.072	3.163
hsa-miR-103-1	-.227	.462	.241	1	.623	.797	.322	1.971
hsa-miR-1307	-.197	.359	.300	1	.584	.822	.407	1.660
hsa-miR-148b	.528	.372	2.022	1	.155	1.696	.819	3.514
hsa-miR-328	.402	.427	.889	1	.346	1.495	.648	3.451
hsa-miR-484	.230	.423	.296	1	.586	1.259	.550	2.884
hsa-miR-874	.308	.355	.756	1	.385	1.361	.679	2.728
hsa-miR-93	.243	.429	.320	1	.572	1.275	.549	2.958
HSP90AA1	-.141	.432	.106	1	.745	.869	.372	2.028
MCM10	-.016	.338	.002	1	.962	.984	.507	1.908
ME1	-.100	.197	.260	1	.610	.905	.615	1.330
NDRG1	-.375	.271	1.916	1	.166	.687	.404	1.169
NOTCH2NL	.710	.394	3.244	1	.072	2.035	.939	4.409
OTUD6B	-.501	.453	1.226	1	.268	.606	.250	1.471
PDSS2	.280	.334	.702	1	.402	1.323	.687	2.548
PGK1	1.908	.523	13.328	1	.000	6.738	2.420	18.765
PIK3CA	.583	.302	3.735	1	.053	1.791	.992	3.234
PTAR1	-.010	.560	.000	1	.986	.990	.330	2.969
SMG1	.532	.772	.475	1	.491	1.703	.375	7.732
TRIM23	.710	.617	1.324	1	.250	2.033	.607	6.809
TTC3	.638	.441	2.093	1	.148	1.893	.797	4.494
UBR5	.314	.589	.283	1	.595	1.368	.431	4.342
UBXN7	-1.099	.543	4.091	1	.043	.333	.115	.967
ZFC3H1	.845	.555	2.313	1	.128	2.328	.784	6.914
<i>TP53</i> Mut	.093	.604	.024	1	.877	1.098	.336	3.589
<i>PIK3CA/AKT/PTEN</i> Mut	1.101	.554	3.944	1	.047	3.006	1.015	8.906
PAM50			3.940	3	.268			
PAM50 (1)	.072	1.186	.004	1	.952	1.075	.105	10.978
PAM50 (2)	-1.060	.723	2.151	1	.142	.346	.084	1.429
PAM50 (3)	.301	.939	.103	1	.748	1.351	.215	8.508
Disease Stage	.144	.409	.124	1	.725	1.155	.518	2.574
T stage	.014	.403	.001	1	.971	1.015	.461	2.235
Estrogen Receptor	1.902	.819	5.396	1	.020	6.699	1.346	33.342
N stage	.498	.422	1.391	1	.238	1.645	.720	3.759

Steps	B	SE	Wald	df	Sig.	HR	95.0% CI for HR	
							Lower	Upper
Step 30								
CPT1A	.586	.205	8.172	1	.004	1.797	1.202	2.686
CXCR7	.497	.187	7.086	1	.008	1.643	1.140	2.368
GLA	-.872	.310	7.906	1	.005	.418	.228	.768
HRASLS	.421	.185	5.169	1	.023	1.523	1.060	2.190
hsa-miR-1307	-.400	.224	3.197	1	.074	.670	.432	1.039
hsa-miR-328	.622	.243	6.556	1	.010	1.862	1.157	2.996
NOTCH2NL	1.002	.304	10.832	1	.001	2.724	1.500	4.946
PGK1	1.253	.280	20.049	1	.000	3.502	2.023	6.062
PIK3CA	.687	.243	7.986	1	.005	1.987	1.234	3.200
TTC3	.653	.307	4.529	1	.033	1.921	1.053	3.504
UBXN7	-.843	.328	6.627	1	.010	.430	.227	.818
ZFC3H1	.816	.353	5.354	1	.021	2.261	1.133	4.513
<i>PIK3CA/AKT/PTEN</i> Mut	1.126	.419	7.211	1	.007	3.084	1.356	7.016
Estrogen Receptor	1.106	.462	5.723	1	.017	3.021	1.221	7.472
N stage	.596	.184	10.538	1	.001	1.815	1.266	2.600

Stratified by age groups (143 patients ≤55 years, 195 patients >55 years). Method = Backward Stepwise (Wald)

Table S7. The Prognostic Values of Eight RNA Signatures in Nine Breast Cancer Cohorts

Cohort	Clinical End point	RNA profile	Integrated miRNA/mRNA	10-miRNA (8)	GGI 97-gene (9)	IGS 186-Gene (10)	95-gene Naoi (11)	76-gene Rotterdam (12)	NKI Mamma Print 70-gene (13, 14)	Oncotype DX (15)
TCGA IDC (n=466)	<i>OS</i>	<i>mRNA/miRNA</i>	0.74 (p<0.001)	n.s. [§]	0.62 (p=0.034)	0.61 (p=0.032)	0.61 (p=0.043)	n.s. [§]	n.s. [§]	n.s. [§]
TCGA IDC Early stages I and II (n=348)	<i>OS</i>	<i>mRNA/miRNA</i>	0.77 (p<0.001)	n.s. [§]	n.s. [§]	n.s. [§]	n.s. [§]	n.s. [§]	0.66 (p=0.028)	n.s. [§]
UK(8) (n=207)	<i>DRFS</i>	<i>mRNA/miRNA</i>	0.65 (p=0.004)	0.76 (p<0.001)	0.66 (p=0.001)	0.70 (p<0.001)	0.72 (p<0.001)	0.66 (p=0.003)	0.73 (p<0.001)	0.68 (p<0.001)
NKI(16) (n=295)	<i>OS</i>	<i>mRNA</i>	0.75* (p<0.001)	na [#]	0.73 (p<0.001)	0.75 (p<0.001)	0.74 (p<0.001)	0.67 (p<0.001)	0.76 (p<0.001)	0.76 (p<0.001)
Hatzis(17) (n=508)	<i>DRFS</i>	<i>mRNA</i>	0.65* (p<0.001)	na [#]	0.66 (p<0.001)	0.65 (p<0.001)	0.64 (p<0.001)	0.62 (p=0.001)	0.62 (p<0.001)	0.63 (p<0.001)
Kao(18) (n=327)	<i>OS</i>	<i>mRNA</i>	0.62* (p=0.006)	na [#]	0.58 (p=0.051)	0.66 (p<0.001)	0.66 (p<0.001)	0.58 (p=0.038)	0.64 (p=0.005)	0.65 (p<0.001)
Wang(12) (n=286)	<i>DRFS</i>	<i>mRNA</i>	0.59* (p=0.025)	na [#]	0.59 (p=0.017)	0.60 (p=0.006)	0.71 (p<0.001)	0.65 (p<0.001)	0.57 (p=0.051)	0.62 (p<0.001)
TRANSBIG(19) (n=198)	<i>OS</i>	<i>mRNA</i>	0.64* (p=0.015)	na [#]	0.70 (p=0.002)	0.63 (p=0.018)	n.s. [§]	0.64 (p=0.023)	n.s. [§]	0.65 (p<0.001)
Bos(20) (n=195)	<i>DRFS</i>	<i>mRNA</i>	0.68* (p=0.011)	na [#]	0.67 (p=0.031)	0.68 (p=0.016)	n.s. [§]	n.s. [§]	0.69 (p=0.016)	0.74 (p=0.003)
TNBC(21) (n=383)	<i>DRFS</i>	<i>mRNA</i>	0.69* (p<0.001)	na [#]	0.65 (p<0.001)	0.68 (p<0.001)	0.69 (p<0.001)	0.65 (p<0.001)	0.68 (p<0.001)	0.66 (p<0.001)

[§] n.s. , p>0.05. The permutation p value was computed for testing the null hypothesis (AUC=0.5) using 1000 permutations. The ROC and KM curves are shown in Supporting Information. [#]na, no assessment was possible, since the miRNA signature could not be applied to an mRNA only profile. *only the mRNA component of the miRNA/mRNA integrated signature was evaluated in the mRNA only datasets.

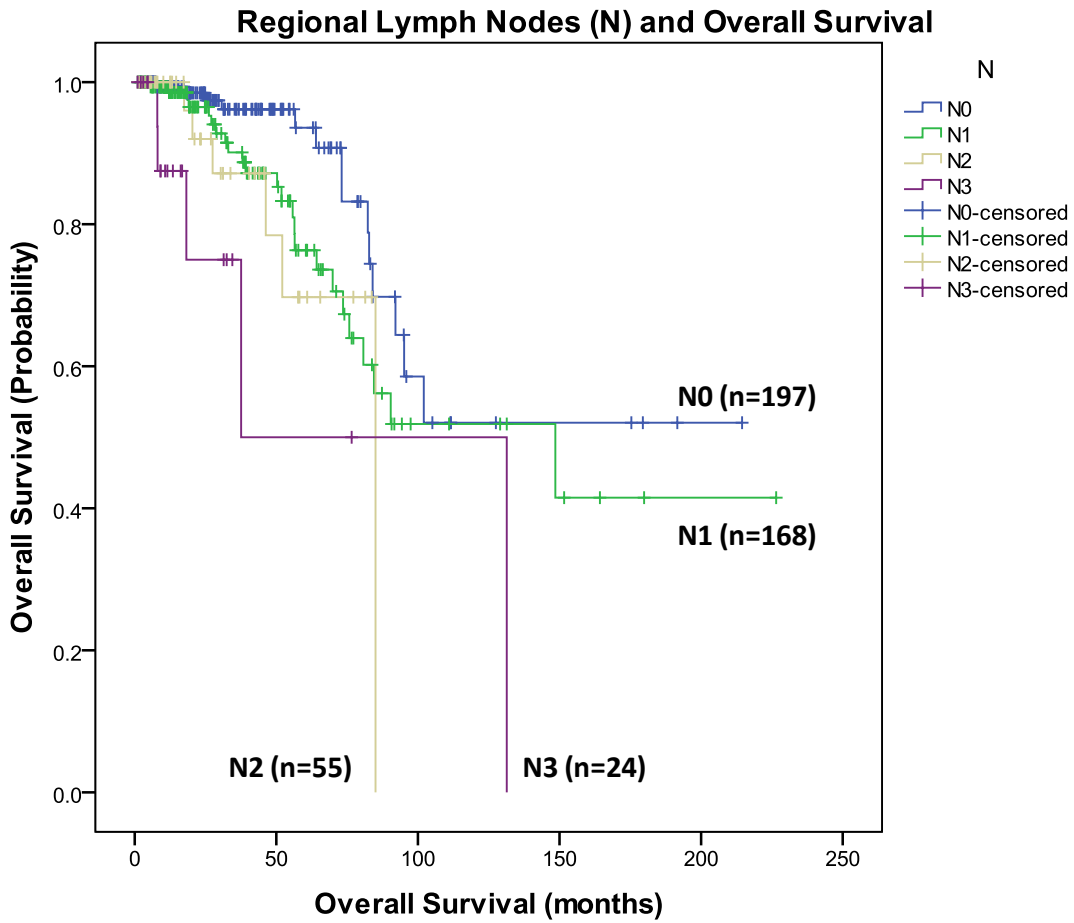


Figure S1. Kaplan Meier survival estimates by Regional Lymph Node involvement (N) in invasive ductal carcinoma (Overall log-rank test, p-value=0.005)

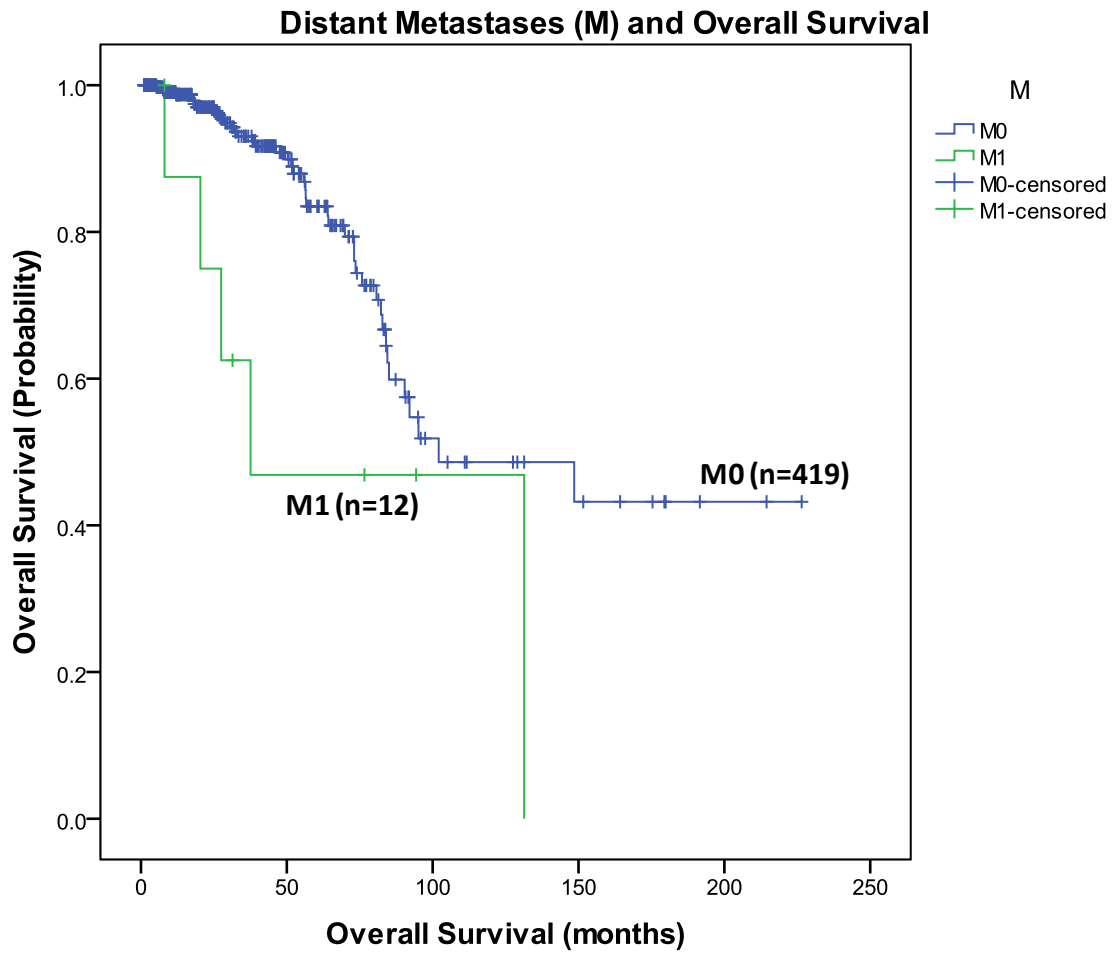


Figure S2. Kaplan Meier survival estimates by Distant Metastases (M) in invasive ductal carcinoma (Overall log-rank test, p-value=0.026)

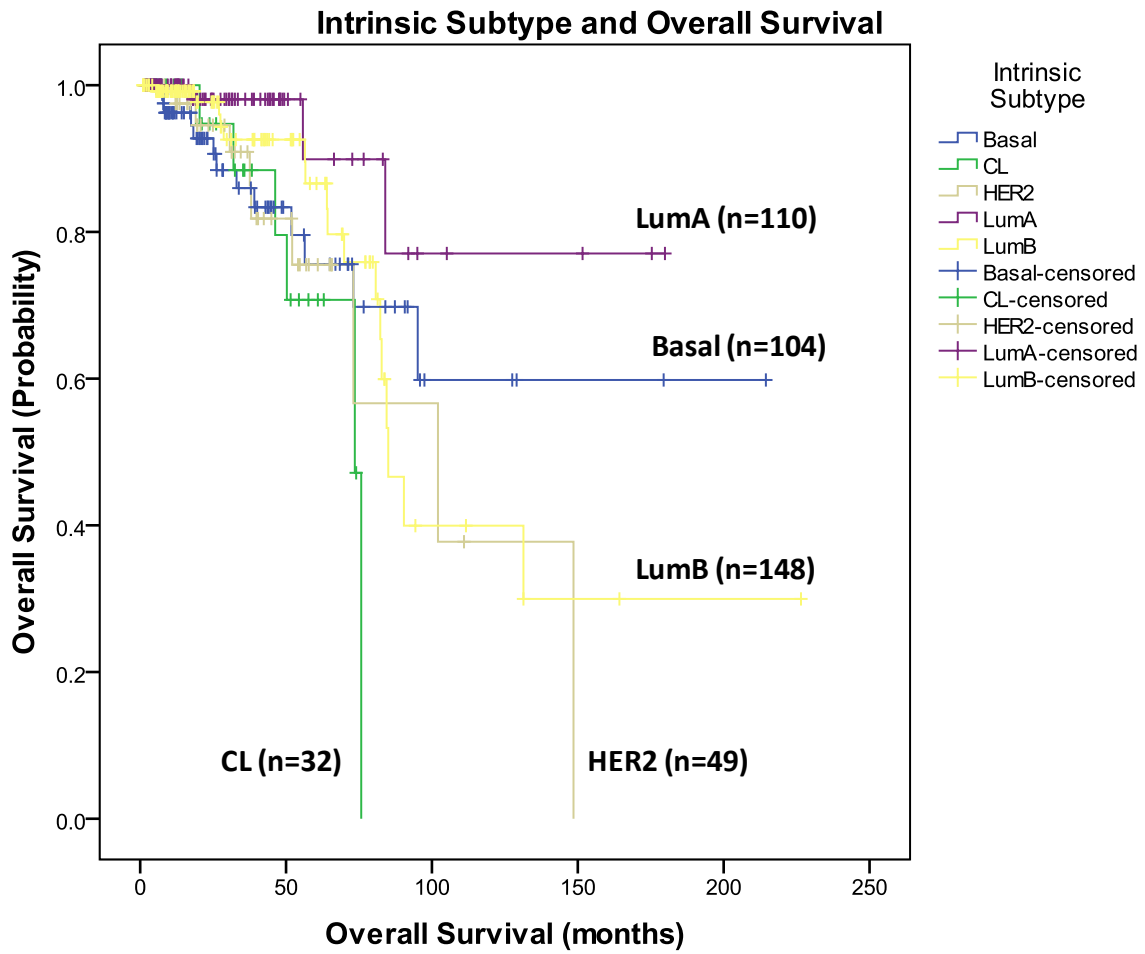


Figure S3. Kaplan Meier survival estimates by intrinsic subtypes in invasive ductal carcinoma (Overall log-rank test, p-value=0.042).

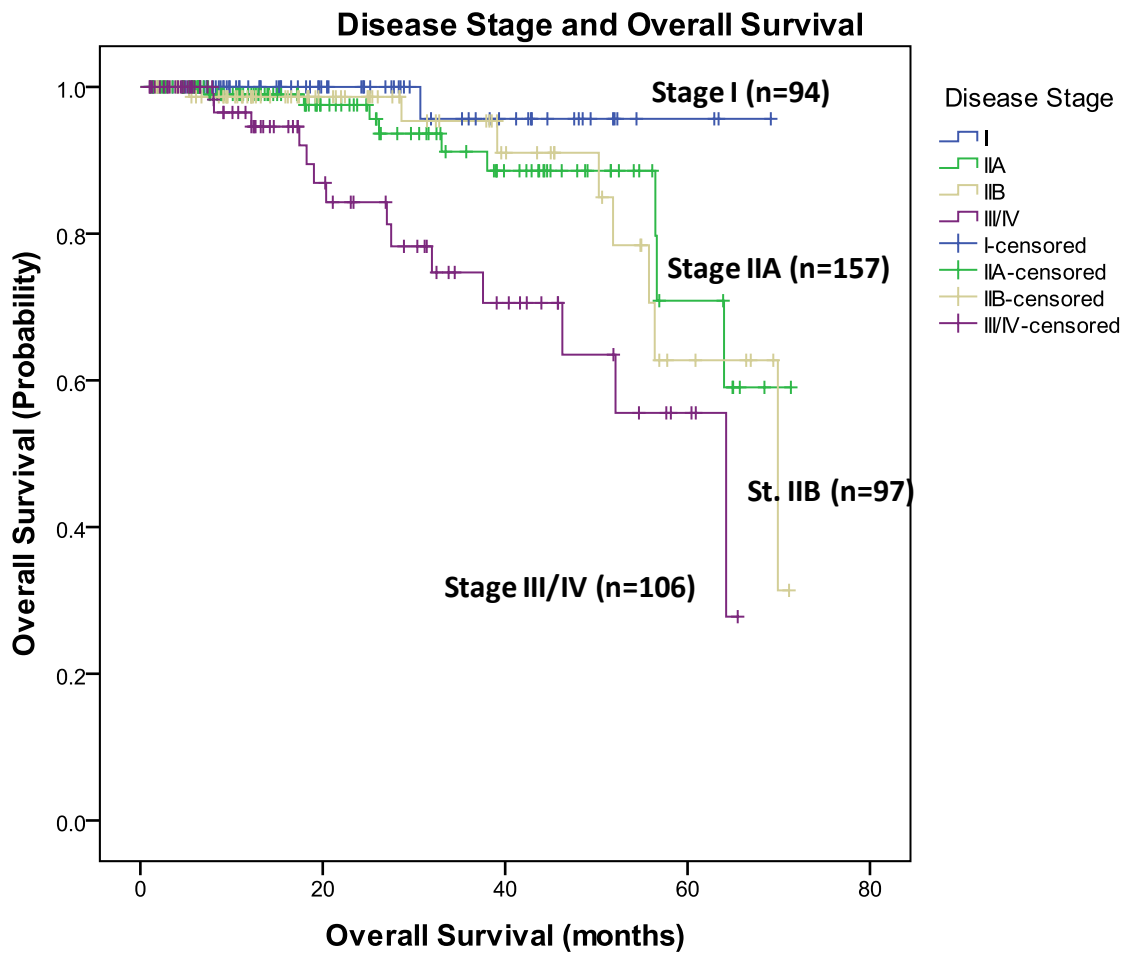


Figure S4. Kaplan Meier survival estimates by the AJCC Disease Stage in invasive ductal carcinoma (Overall log-rank test, p-value=0.002)

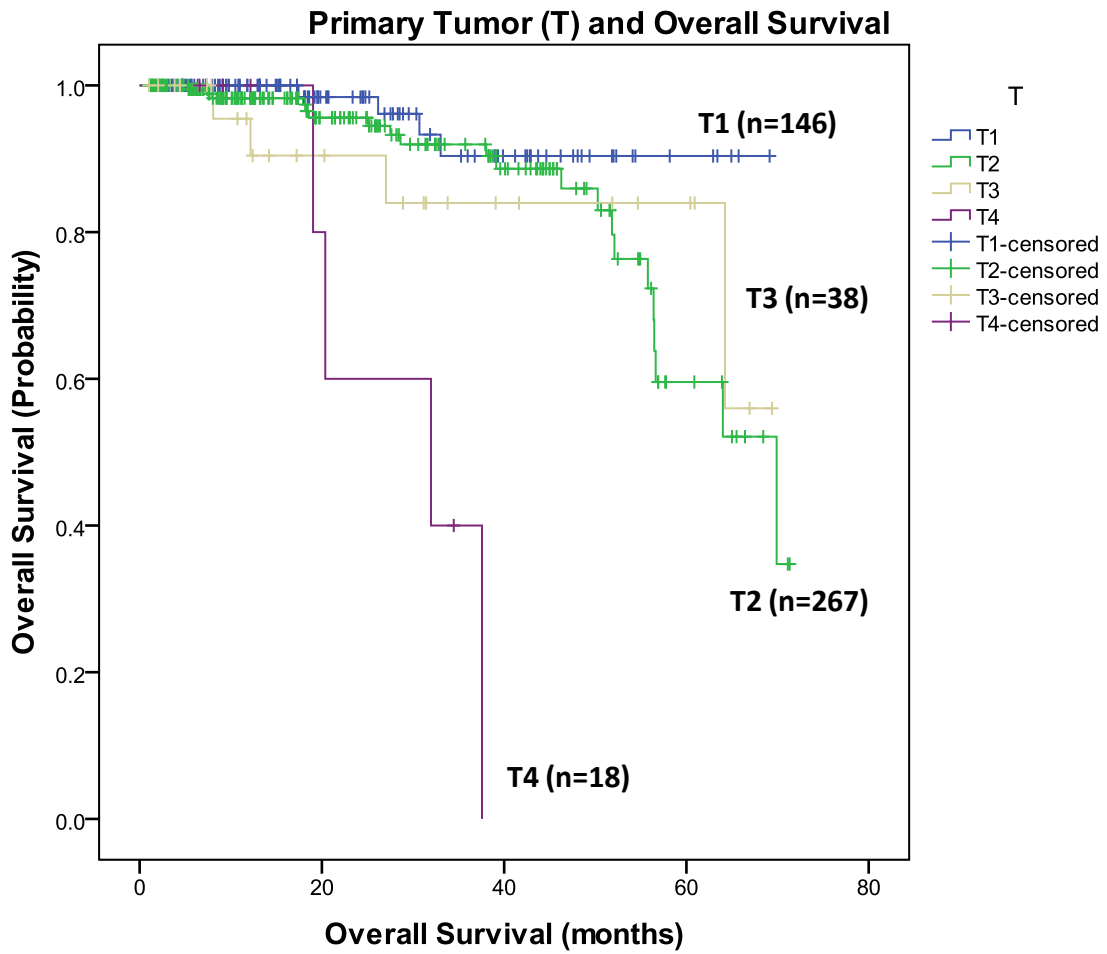


Figure S5. Kaplan Meier survival estimates by the T stage in invasive ductal carcinoma (Overall log-rank test, p-value <0.001)

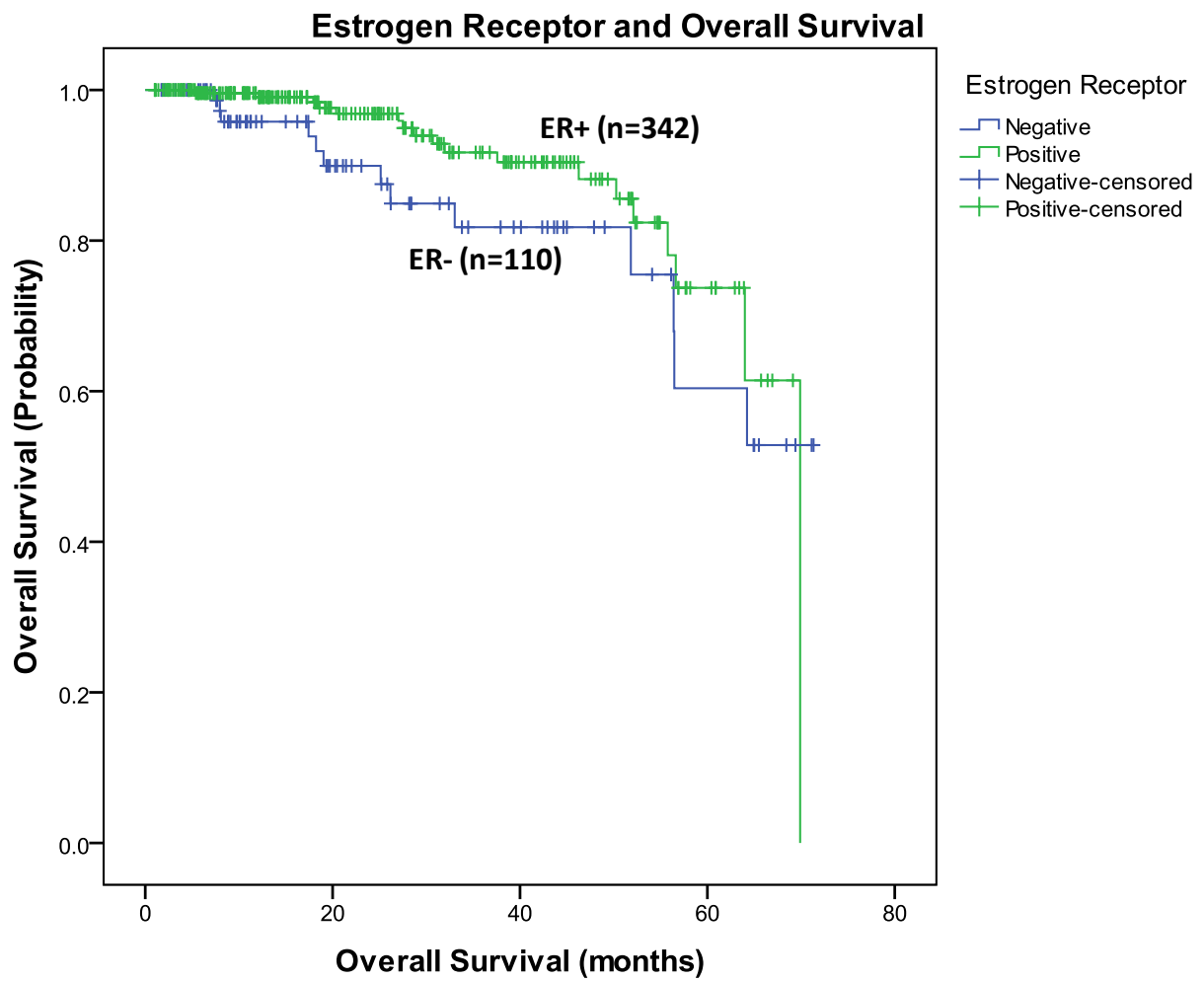


Figure S6. Kaplan Meier survival estimates by the Estrogen Receptor (ER) status in invasive ductal carcinoma (Breslow test, p-value=0.016)

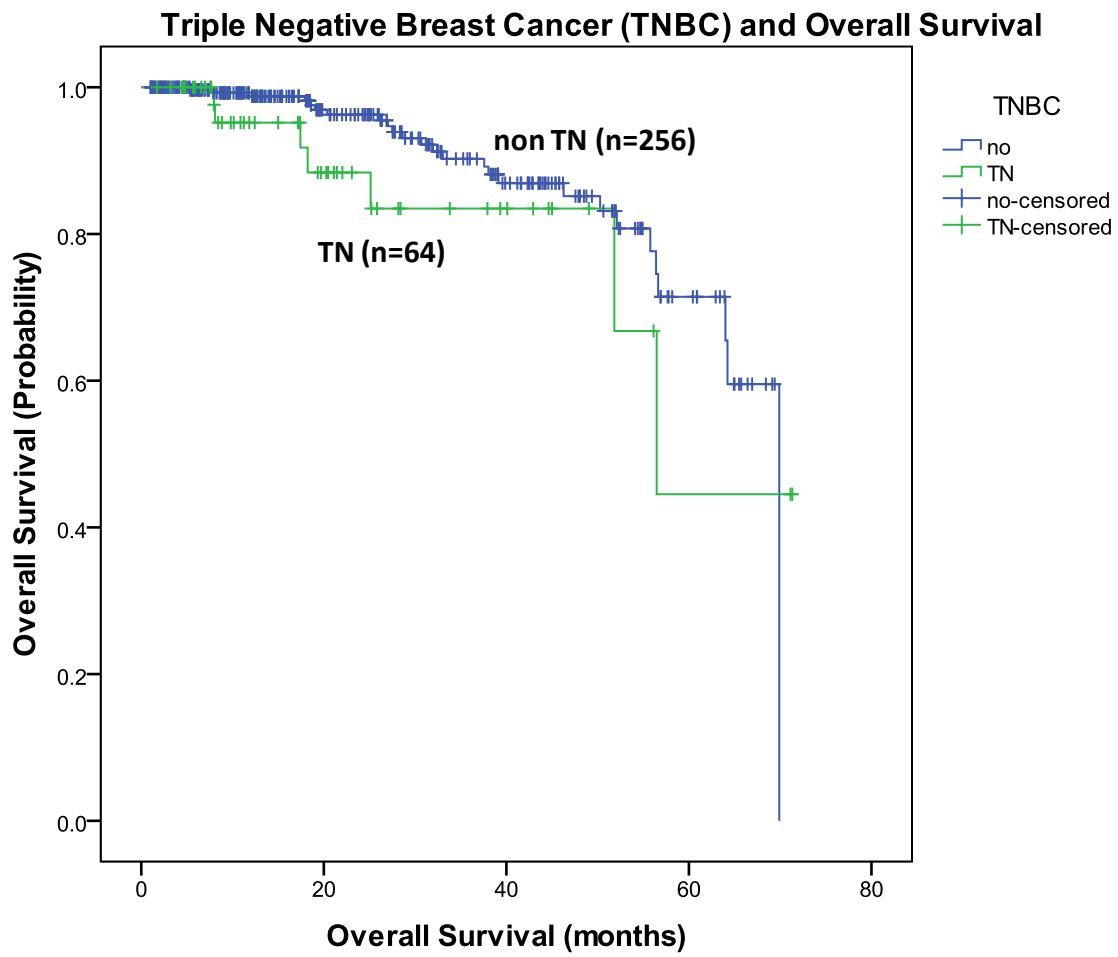


Figure S7. Kaplan Meier survival estimates by the Triple Negative (TNBC) status in invasive ductal carcinoma (Breslow test, p-value=0.041)

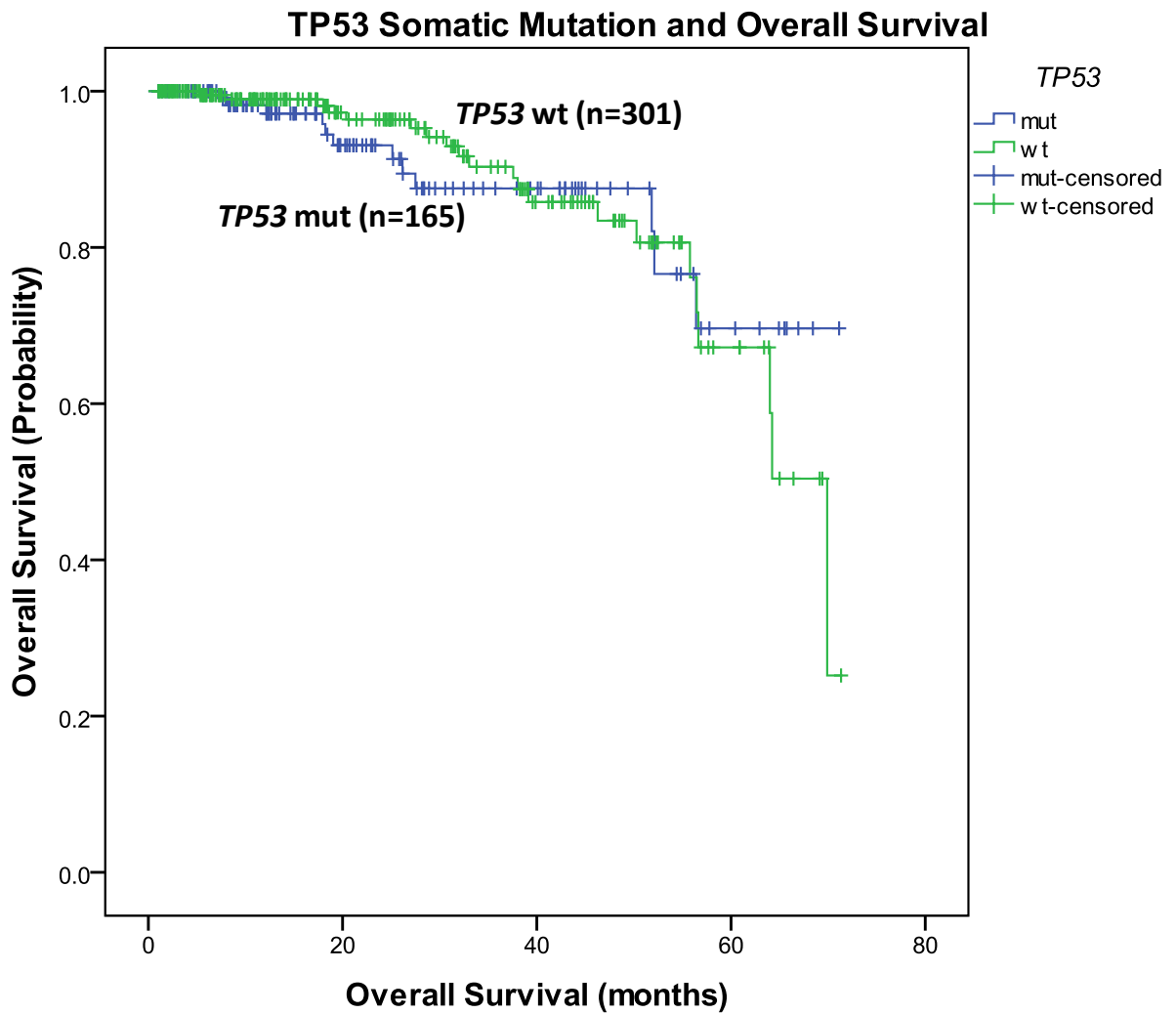


Figure S8. Kaplan Meier survival estimates by the *TP53* somatic mutation status in invasive ductal carcinoma (log-rank test, non significant).

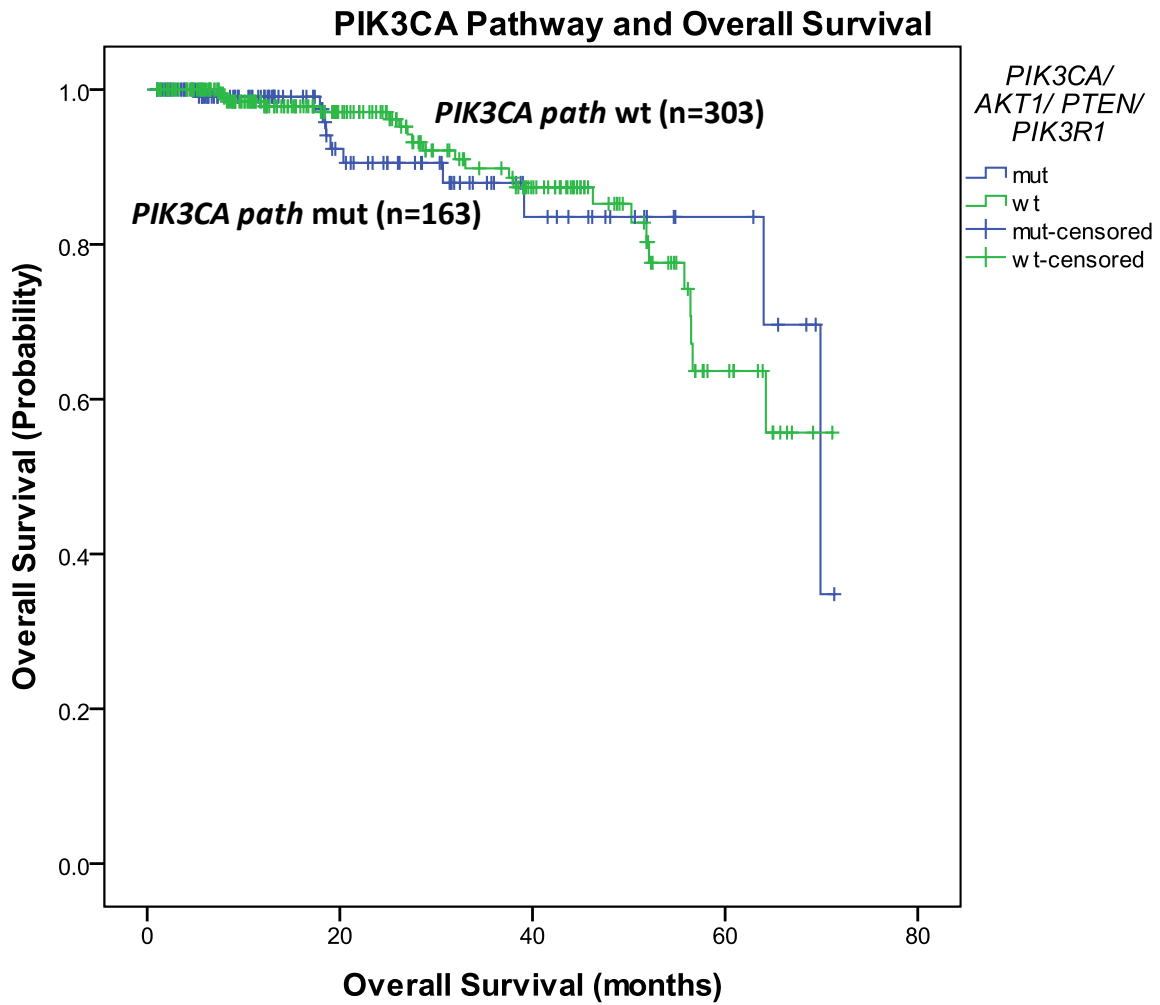


Figure S9. Kaplan Meier survival estimates by the *PIK3CA* pathway somatic mutation status in invasive ductal carcinoma (log-rank test, non significant).

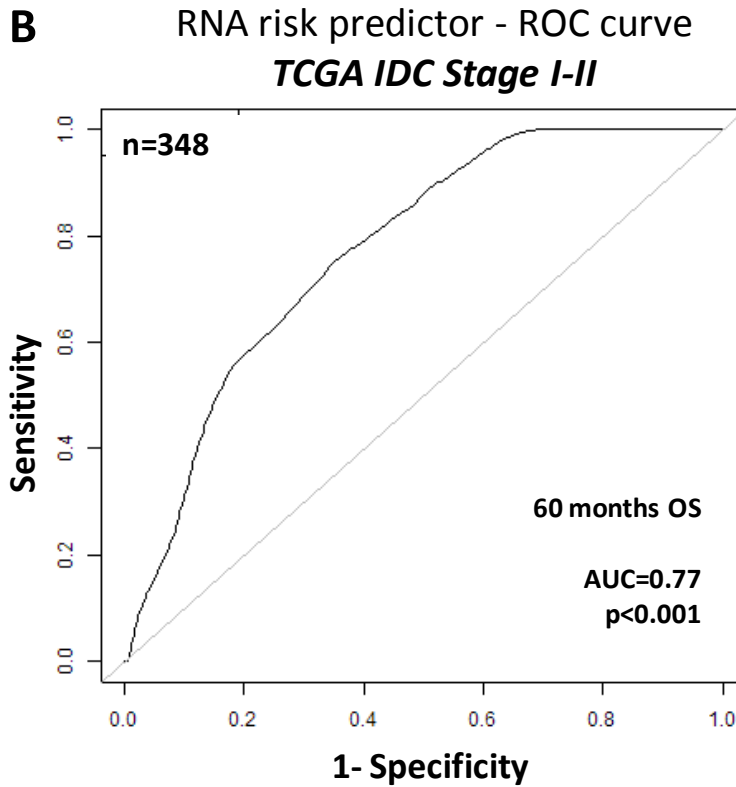
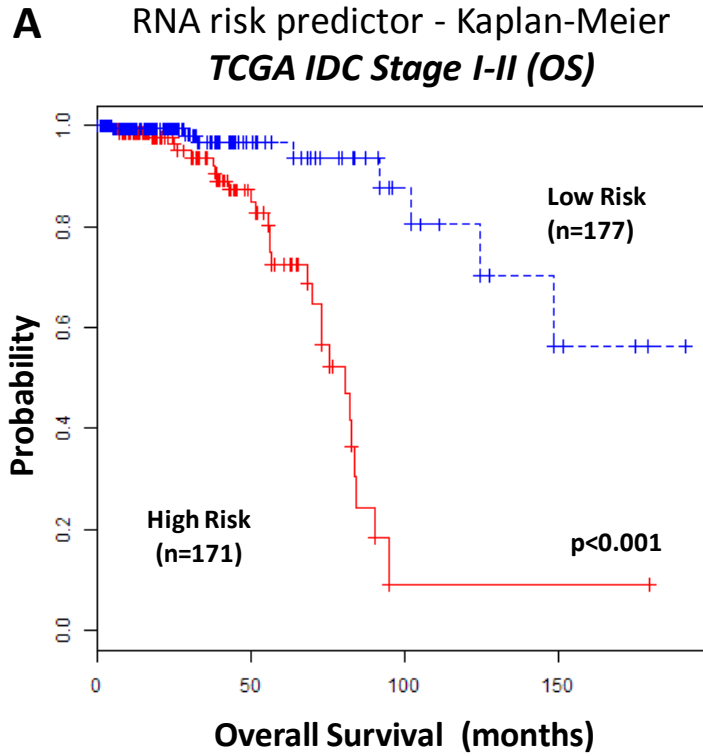
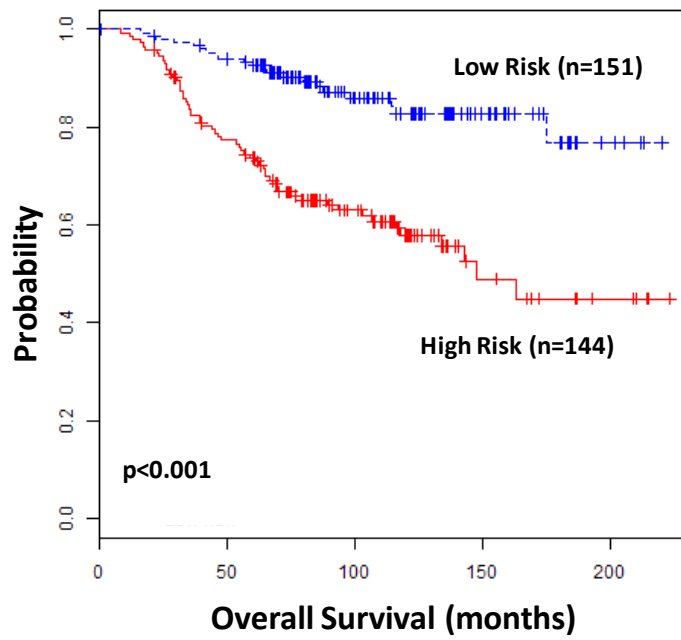


Figure S10. A) KM and B) ROC curves for the miRNA/mRNA signature in early stage I and II IDC tumors of the TCGA cohort.

A RNA risk predictor - Kaplan-Meier
NKI (OS)



B RNA risk predictor - ROC curve
NKI validation cohort

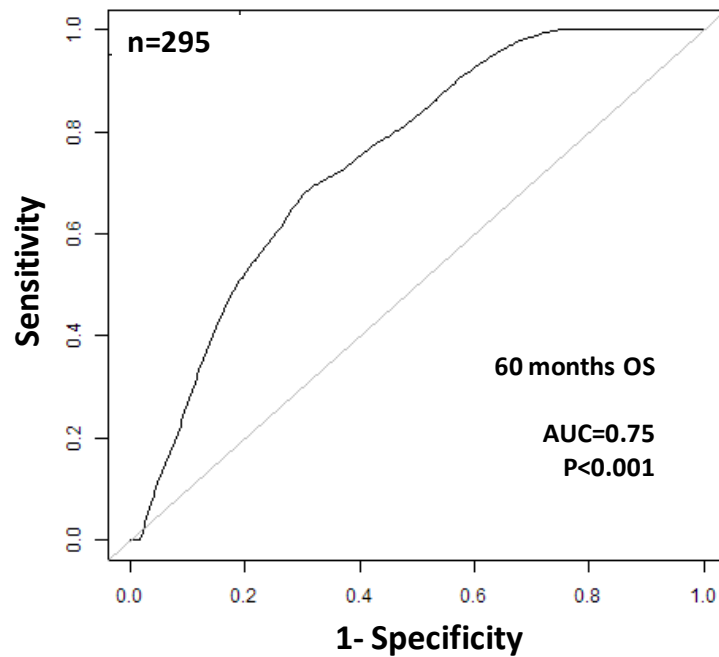
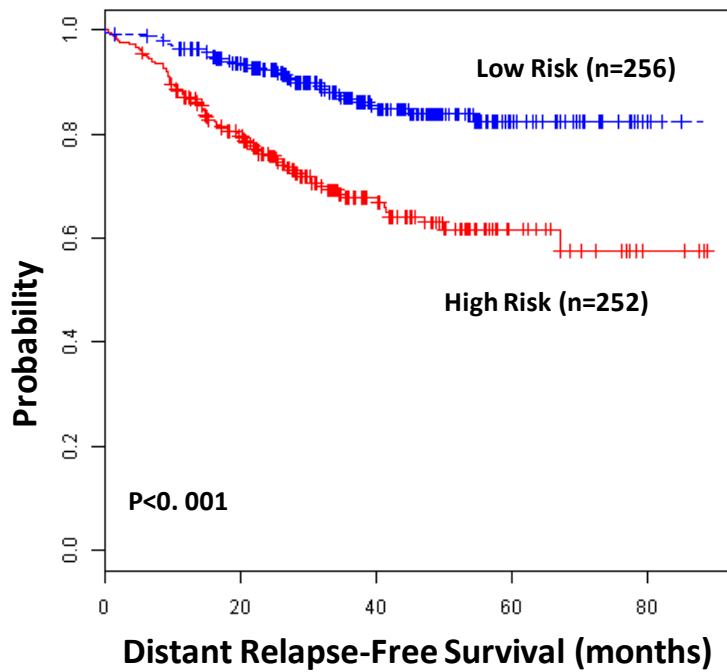


Figure S11. A) KM and B) ROC curves for the miRNA/mRNA signature in the NKI cohort.

A RNA risk predictor - Kaplan-Meier
Hatzis (DRFS)



B RNA risk predictor - ROC curve
Hatzis validation cohort

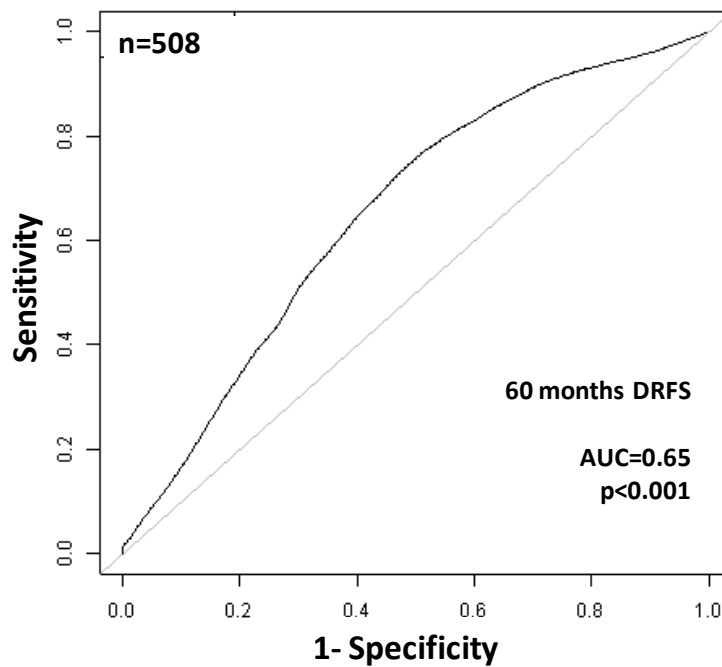


Figure S12. A) KM and B) ROC curves for the miRNA/mRNA signature in the Hatzis cohort.

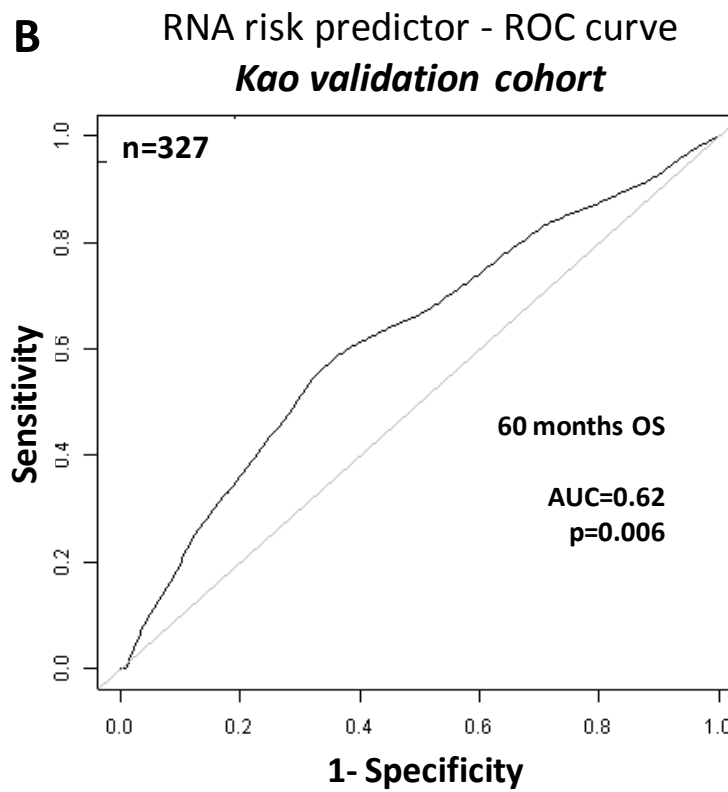
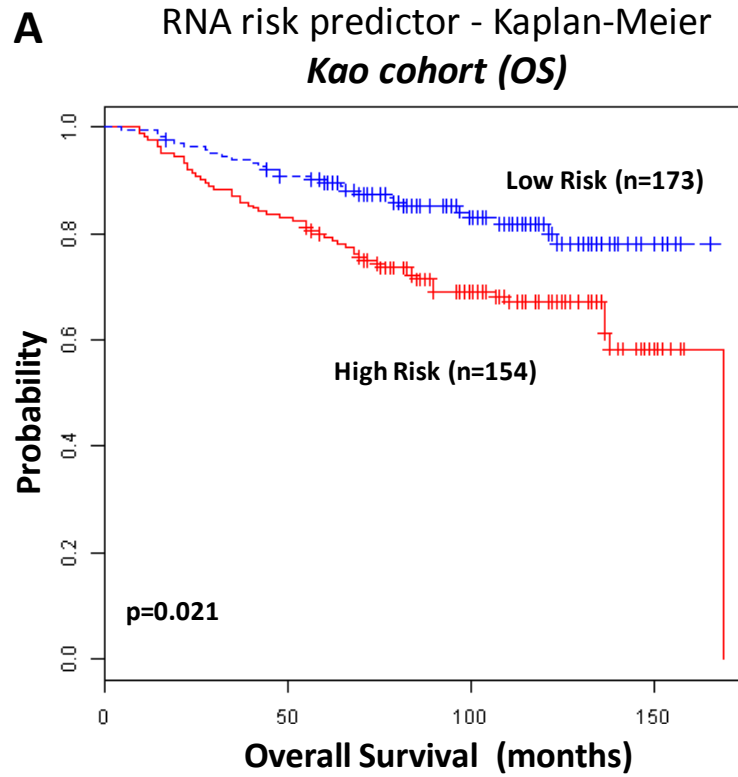
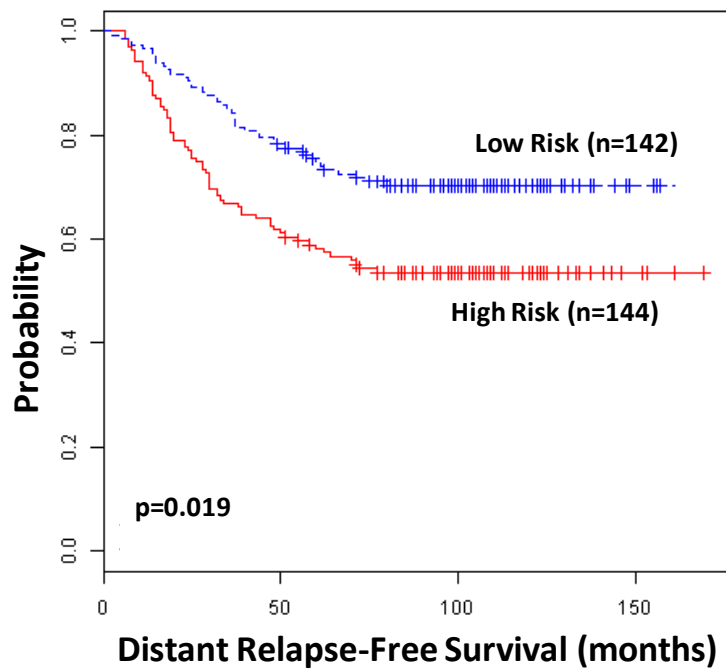


Figure S13. A) KM and B) ROC curves for the miRNA/mRNA signature in the Kao cohort.

A RNA risk predictor - Kaplan-Meier
Wang (DRFS)



B RNA risk predictor - ROC curve
Wang validation cohort

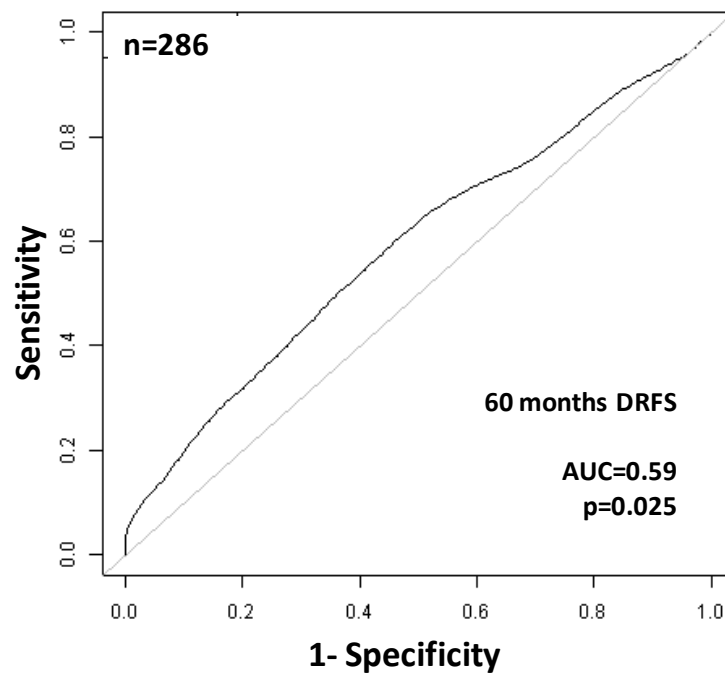


Figure S14. A) KM and B) ROC curves for the miRNA/mRNA signature in the Wang cohort.

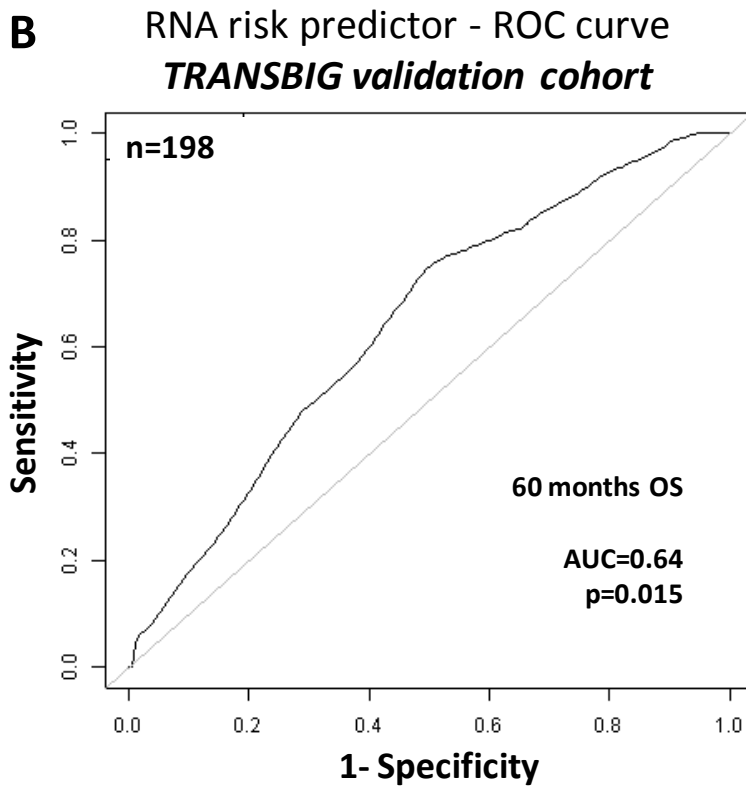
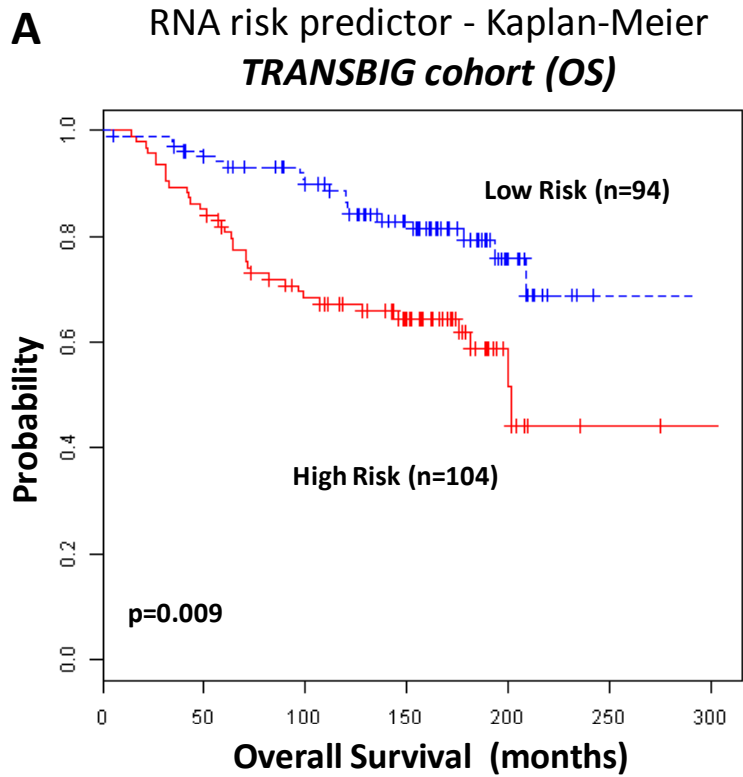


Figure S15. A) KM and B) ROC curves for the miRNA/mRNA signature in the TRANSBIG cohort.

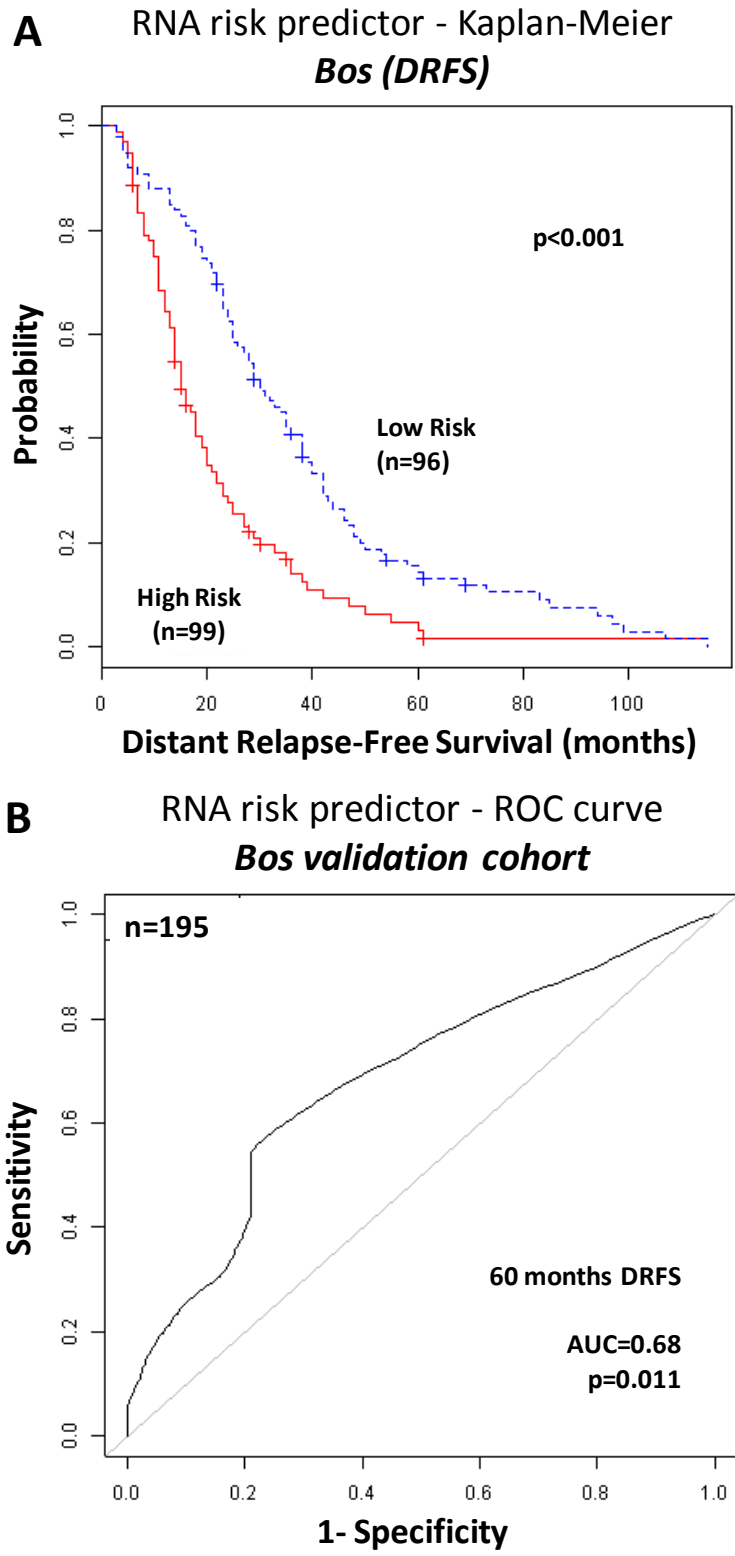


Figure S16. A) KM and B) ROC curves for the miRNA/mRNA signature in the Bos cohort.

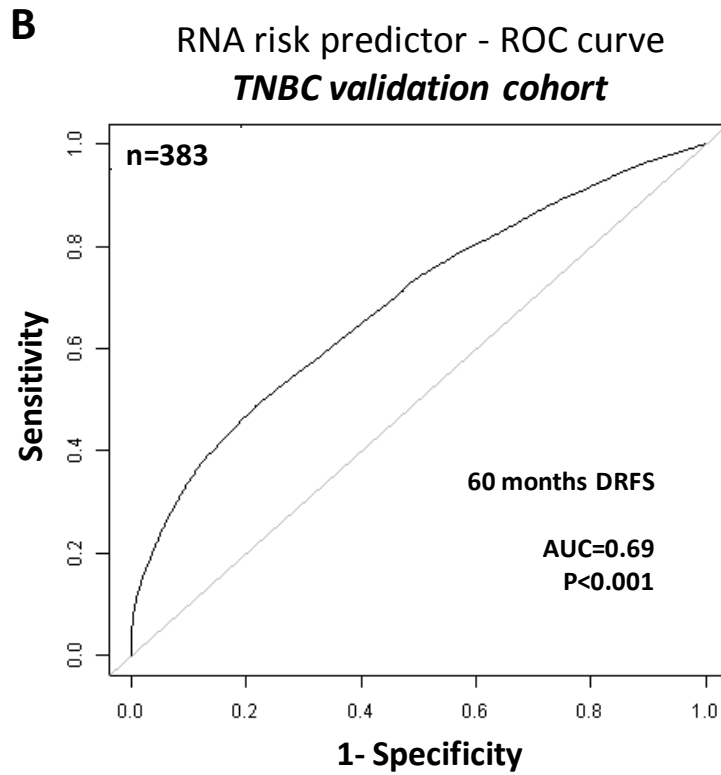
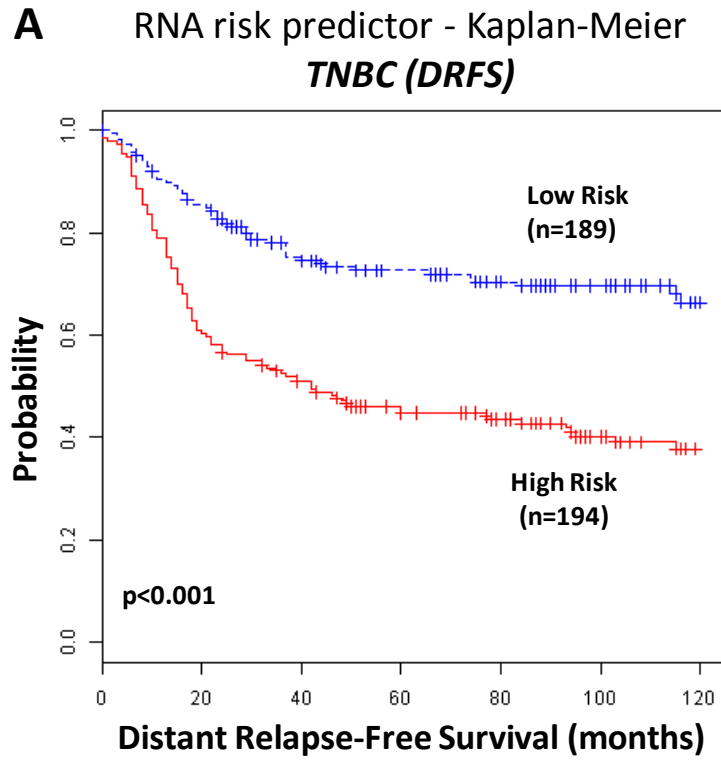


Figure S17. A) KM and B) ROC curves for the miRNA/mRNA signature in the TNBC cohort.

References

1. Mortazavi A, Williams BA, McCue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5(7):621-628.
2. Bamford S, *et al.* (2004) The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 91(2):355-358.
3. Du P, *et al.* (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11:587.
4. Heagerty PJ, Lumley T, & Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56(2):337-344.
5. Maere S, Heymans K, & Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21(16):3448-3449.
6. Shannon P, *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.
7. Storey JD & Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* 100(16):9440-9445.
8. Buffa FM, *et al.* (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer research* 71(17):5635-5645.
9. Sotiriou C, *et al.* (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98(4):262-272.
10. Liu R, *et al.* (2007) The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 356(3):217-226.
11. Naoi Y, *et al.* (2011) Development of 95-gene classifier as a powerful predictor of recurrences in node-negative and ER-positive breast cancer patients. *Breast Cancer Res Treat* 128(3):633-641.
12. Wang Y, *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365(9460):671-679.
13. Buyse M, *et al.* (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 98(17):1183-1192.
14. Glas AM, *et al.* (2006) Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 7:278.
15. Paik S, *et al.* (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351(27):2817-2826.
16. van de Vijver MJ, *et al.* (2002) A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347(25):1999-2009.
17. Hatzis C, *et al.* (2011) A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* 305(18):1873-1881.
18. Kao KJ, Chang KM, Hsu HC, & Huang AT (2011) Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* 11:143.
19. Desmedt C, *et al.* (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 13(11):3207-3214.
20. Bos PD, *et al.* (2009) Genes that mediate breast cancer metastasis to the brain. *Nature* 459(7249):1005-1009.
21. Rody A, *et al.* (2011) A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast cancer research : BCR* 13(5):R97.