

FOXA1 is Prognostic of Triple Negative Breast Cancers by Transcriptionally Suppressing *SOD2* and *IL6*

Xiaofeng Dai²✉, Hongye Cheng¹, Xiao Chen¹, Ting Li¹, Jia Zhang^{2,3}, Guoyin Jin^{2,3}, Dongyan Cai^{2,3}, Zhaohui Huang^{2,3}✉

1. School of Biotechnology, Jiangnan University, Wuxi, China
2. Wuxi School of Medicine, Jiangnan University, Wuxi, China
3. Wuxi Cancer Institute, Affiliated Hospital of Jiangnan University, Wuxi, China

Supplementary Tables

Supplementary Table S1. Breast cancer cell lines, subtypes and subtyping markers.

Cell lines		ER	PR	HER2	BRCA1 mutation	Culture media	Culture conditions
Luminal A	MCF7	+	+	-	-	DMEM, 10% FBS	37°C, 5% CO ₂
	T47D	+	+	-	-	DMEM, 10% FBS	37°C, 5% CO ₂
Luminal B	BT474	+	+	+	-	RPMI 1640, 20% FBS	37°C, 5% CO ₂
Her2 Positive	SKBR3	-	-	+++	-	RPMI 1640, 10% FBS	37°C, 5% CO ₂
	MDAMB453	-	-	+	-	L15, 10% FBS and 10 µg/mL insulin	37°C, 5% CO ₂
Triple Negative						F12, 10% FBS, 10 µg/mL insulin, 10 mmol/l HEPES, and 20 ng/mL epidermal growth factor	37°C, 5% CO ₂
	SUM159PT	-	-	-	-		
	MDAMB231	-	-	-	-	DMEM, 10% FBS	37°C, 5% CO ₂
	HCC1937	-	-	-	+	RPMI 1640, 10% FBS	37°C, 5% CO ₂
	SUM149PT	-	-	-	+	F12, 10% FBS, 10 µg/mL insulin, 10 mmol/l HEPES, and 20 ng/mL epidermal growth factor	37°C, 5% CO ₂
	MDAMB436	-	-	-	+	L15, 10% FBS and 10 µg/mL insulin	37°C, 5% CO ₂

Supplementary Table S2. The sequences of the primers used in Real time PCR and ChIP assays and sgRNAs used in CRISPR mediated gene up-regulation.

Primer name	Sequences (5'→3')	Assay
-------------	-------------------	-------

<i>GAPDH-F</i>	CCCACTCCTCCACCTTTGAC	qRT-PCR
<i>GAPDH-R</i>	ATGAGGTCCACCACCCTGTT	qRT-PCR
<i>FOXA1-F</i>	GAAGATGGAAGGGCATGAAA	qRT-PCR
<i>FOXA1-R</i>	GCCTGAGTTCATGTTGCTGA	qRT-PCR
<i>SOD2-F</i>	GACAAACCTCAGCCCTAACG	qRT-PCR
<i>SOD2-R</i>	TTGGACACCAACAGATGCAG	qRT-PCR
<i>IL6-F</i>	ACTCACCTCTTCAGAACGAATTG	qRT-PCR
<i>IL6-R</i>	CCATCTTTGGAAGGTTTCAGGTTG	qRT-PCR
<i>MYC-F</i>	GTCAAGAGGGCGAACACACAAC	qRT-PCR
<i>MYC-R</i>	TTGGACGGACAGGATGTATGC	qRT-PCR
<i>HER2-F</i>	TGTGACTGCCTGTCCCTACAA	qRT-PCR
<i>HER2-R</i>	CCAGACCATAGCACACTCGG	qRT-PCR
<i>MLPH-F</i>	TGCCCATCTGAACGAGACC	qRT-PCR
<i>MLPH-R</i>	GAGCCGATCTTCACGACTCTG	qRT-PCR
<i>XBPI-F</i>	CCCTCCAGAACATCTCCCAT	qRT-PCR
<i>XBPI-R</i>	ACATGACTGGGTCCAAGTTGT	qRT-PCR
<i>GATA3-F</i>	GCCCCTCATTAAGCCCAAG	qRT-PCR
<i>GATA3-R</i>	TTGTGGTGGTCTGACAGTTCG	qRT-PCR
<i>KIAA1324-F</i>	GATGACTGCACGGTGTCTTTG	qRT-PCR
<i>KIAA1324-R</i>	TGAGAGCCCCATTCTCCATTG	qRT-PCR
<i>ESR-F</i>	CAGGCATTCGGTTTGATGAGT	qRT-PCR
<i>ESR-R</i>	TTGGACGAAGTACAGTTCCCG	qRT-PCR
<i>CA12-F</i>	AGTGAACGGTTCCAAGTGGAC	qRT-PCR
<i>CA12-R</i>	CCACACGACGGGTACTTCT	qRT-PCR
<i>PSAT-F</i>	TGCCGCACTCAGTGTGTTAG	qRT-PCR
<i>PSAT-R</i>	GCAATTCCCGCACAAGATTCT	qRT-PCR
<i>MYB-F</i>	GAGGTGGCATAACCACTTGAA	qRT-PCR
<i>MYB-R</i>	AGGCAGTAGCTTTGCGATTTC	qRT-PCR
<i>SCNNIA-F</i>	TCTGCACCTTTGGCATGATGT	qRT-PCR
<i>SCNNIA-R</i>	GAAGACGAGCTTGTCGGAGT	qRT-PCR
<i>AGR2-F</i>	GTCAGCATTCTTGCTCCTTGT	qRT-PCR
<i>AGR2-R</i>	GGGTGCGAGAGTCCTTTGTGTC	qRT-PCR
<i>FBP1-F</i>	GAACCGGAGAAAAGGGGTAAA	qRT-PCR
<i>FBP1-R</i>	GTTCCAACGGACACAAGGCA	qRT-PCR
<i>MYO5C-F</i>	TCGTGGGCGAGAATGACCT	qRT-PCR
<i>MYO5C-R</i>	GGCAACTGCTTGTAAGGATTCA	qRT-PCR
<i>siFOXA1-F</i>	GGACUUCAAGGCAUACGAATT	siRNA
<i>siFOXA1-R</i>	UUCGUAUGCCUUGAAGUCCAG	siRNA
<i>siSOD2-F</i>	UUCGUAUGCCUUGAAGUCCAG	siRNA
<i>siSOD2-R</i>	AGUGGAAUAAGGCCUGUUGTT	siRNA
<i>siMYC-F</i>	ACAGCCCACUGGUCCUCAATT	siRNA
<i>siMYC-R</i>	UUGAGGACCAGUGGGCUGUGA	siRNA
<i>siIL6-F</i>	GGAGUUUGAGGUUAUACCUATT	siRNA
<i>siIL6-R</i>	UUCGUAUGCCUUGAAGUCCAG	siRNA
<i>GAPDH siRNA</i>	UGACCUCAACUACAUGGUUT	siRNA
<i>NC siRNA</i>	UUCUCCGAACGUGUCACGUTT	siRNA

<i>SOD2-F</i>	GTTCCAGTGAGCCGACAT	ChIP
<i>SOD2-R</i>	AACAGTCAGGCGAAGAGG	ChIP
<i>IL6-F</i>	GGTCCTTGATGTAACAGCCAG	ChIP
<i>IL6-R</i>	AGGATTTCTGCACTTACTTGTG	ChIP
<i>FOXA1-sg1</i>	GCTCGCACCTACAAAGCCCG	CRISPR
<i>FOXA1-sg2</i>	AAGCCCGAGGTGCACCTGCA	CRISPR
<i>FOXA1-sg3</i>	ACACGCCACCTTCCGAGCGC	CRISPR

Supplementary Table S3. Phenotypic information of 82 collected samples.

Features	N
Subtypes	
Luminal A	20
Luminal B	12
Her2+	21
TNBC	29
Ages (years)	
< 56	41
≥ 56	41
Tumor size (cm)	
≤2	31
≥ 2	51
Grade	
I	3
II	41
III	38
Stage	
I + II	37
III + IV	45
Histology	
Invasive ductal carcinoma	76
Others	6

Supplementary Table S4. Evaluation method for the nuclear marker expression in breast cancer tissues.

Score	Staining intensity	Percentage of nuclear expression for marker protein
0	0-3	<5
1	1	<80
1	2	<40
2	1	≥80
2	2	40-80
2	3	<40
3	2	≥80
3	3	≥40

Supplementary Table S5. Public datasets used for statistical assessment of candidate markers for breast cancer subtyping.

	METABRIC	TCGA	GSE24450	E-MTAB-181
Data portal	cBioPortal for Cancer Genomics www.cbioportal.org	TCGA www.tcgacancer.gov /dataportal	Gene Expression Omnibus www.ncbi.nlm.nih.gov/g eo/	ArrayExpress www.ebi.ac.uk/arrayexpress/experimen ts/E-MTAB-181/
Gene number	24368	17814	48803	23886
Sample number	1904	502	183	56
Sample type	Clinical samples	Clinical samples	Clinical samples	Cell lines
Detection approach	Illumina Human v3 microarray	Agilent 244 K Custom Gene Expression G4502A-07-3	Illumina HumanHT-12 V3.0 expression beadchip	Affymetrix GeneChip Human Exon 1.0 ST Array

Supplementary Table S6. Clinical information of the 3951 breast cancer samples collected in KM plotter and used in this

study. 'NA' represents 'not available'.

Features	N
ER status	
ER positive	2565
ER negative	1214
NA	172
PR status	
PR positive	954
PR negative	1028
NA	1969
HER2 status	
HER2 positive	416
HER2 negative	1456
NA	1979
Lymph node status	
Lymph node positive	1459
Lymph node negative	2259
NA	231
TP53 status	
TP53 positive	232
TP53 negative	363
NA	3356
Grade	
I	378
II	1077
III	1090
NA	1406

Pietenpol subtype

Basal-like 1	239
Basal-like 2	97
Immunomodulatory	290
Mesenchymal	229
Mesenchymal stem-like	115
Luminal androgen receptor	276
NA	2705

Supplementary Table S7. The 18 genes transcriptionally correlated with *FOXA1*. *METABRIC*, *TCGA* and GSE24450 were used to conduct the Pearson correlation analysis. KM plotter was used to compute the 5 years survival which includes 3951 patients.

Gene	Pearson correlation				KM plotter			
	cor_MET ABRIC	p_META BRIC	cor_GSE2 4450	p_GSE2 4450	cor_TCGA	p_TCGA	p(5year)	HR(5year)
FOXA1	1	0	1	0	1	0	8.80E-09	0.69
IL6DBP	-0.5815	1.12E-172	0.5549	3.63E-16	-0.5610	3.37E-92	8.60E-08	1.39
SOD2	-0.5747	8.45E-168	-0.6457	5.67E-23	-0.6104	2.16E-47	3.60E-08	1.43
ESR1	0.7241	<1E-300	0.7328	4.32E-32	0.7693	2.18E-89	<1E-16	0.58
CA12	0.7440	0	0.7029	1.40E-28	0.7673	1.15E-88	<1E-16	0.56
MYB	0.6799	1.53E-258	0.7550	5.32E-35	0.7051	5.04E-69	<1E-16	0.54
MLPH	0.8985	0.00E+00	0.8252	9.07E-47	0.8931	6.81E-158	1.20E-11	0.64
MYO5C	0.6461	1.82E-225	0.8143	1.25E-44	0.6190	4.67E-49	<1E-16	0.58
DNALI1	0.6086	2.45E-193	0.6433	9.26E-23	0.7132	2.68E-71	1.3E-13	0.52
KIAA1324	0.7017	2.43E-282	0.8335	1.61E-48	0.6959	1.44E-66	4.90E-14	0.61
XBP1	0.7969	0	0.8394	8.28E-50	0.8287	3.13E-115	2.20E-16	0.58
GATA3	0.7810	0	0.8344	1.05E-48	0.7989	3.47E-101	7.00E-12	0.64
SLC22A5	0.6297	8.05E-21	0.7066	5.36E-29	0.6806	1.24E-62	<1E-16	0.54
WWP1	0.5751	4.27E-168	0.6298	1.30E-21	0.6009	1.34E-45	0.02	1.16
FBP1	0.6765	6.00E-255	0.7626	4.45E-36	0.7478	6.65E-82	2.80E-09	0.68
LRBA	0.6336	3.34E-214	0.7161	4.46E-30	0.6084	5.18E-47	7.40E-08	0.71
AGR2	0.6161	1.98E-199	0.6659	8.23E-25	0.8297	9.16E-116	8.00E-03	0.84
CRIP1	0.6033	3.69E-189	0.6946	1.10E-27	0.6063	1.28E-46	2.90E-03	0.82
PSAT1	-0.7091	6.43E-291	-0.7758	4.87E-38	-0.7154	6.43E-72	4.30E-07	1.56

Supplementary Table S8. IHC staining results of SOD2 in breast cancer tissues.

Subtype	N	SOD2 expression		p value
		Score 0-1	Score 2-3	

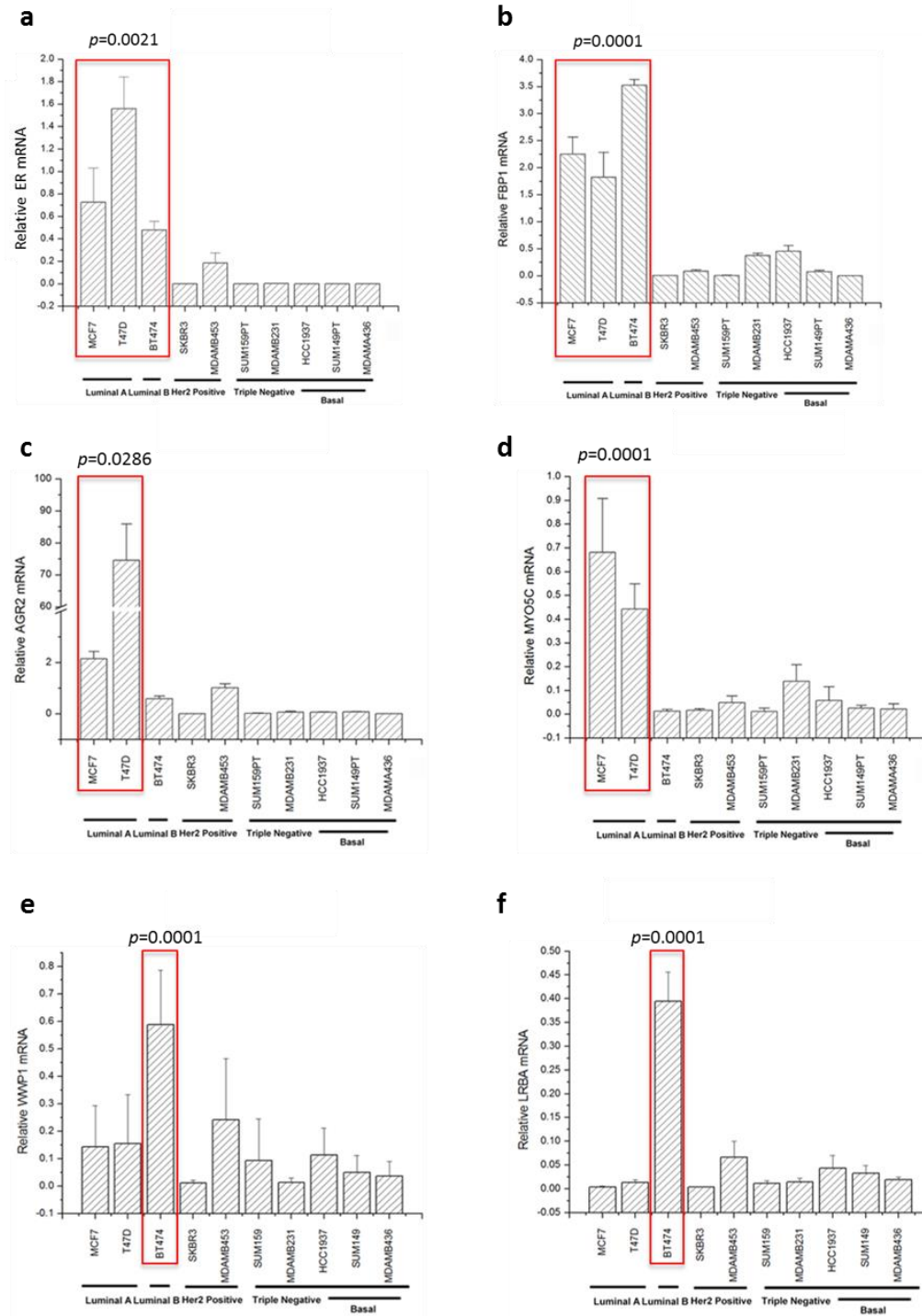
TNBC	29	5	24	1.35E-05
non-TNBC	53	28	25	
Luminal	32	14	18	9.28E-03
non-Luminal	50	19	31	

Supplementary Table S9. Correlation analysis among FOXA1, SOD2 and IL6 at the transcription level. Gene expression from METABRIC, TCGA and GSE24450 were used.

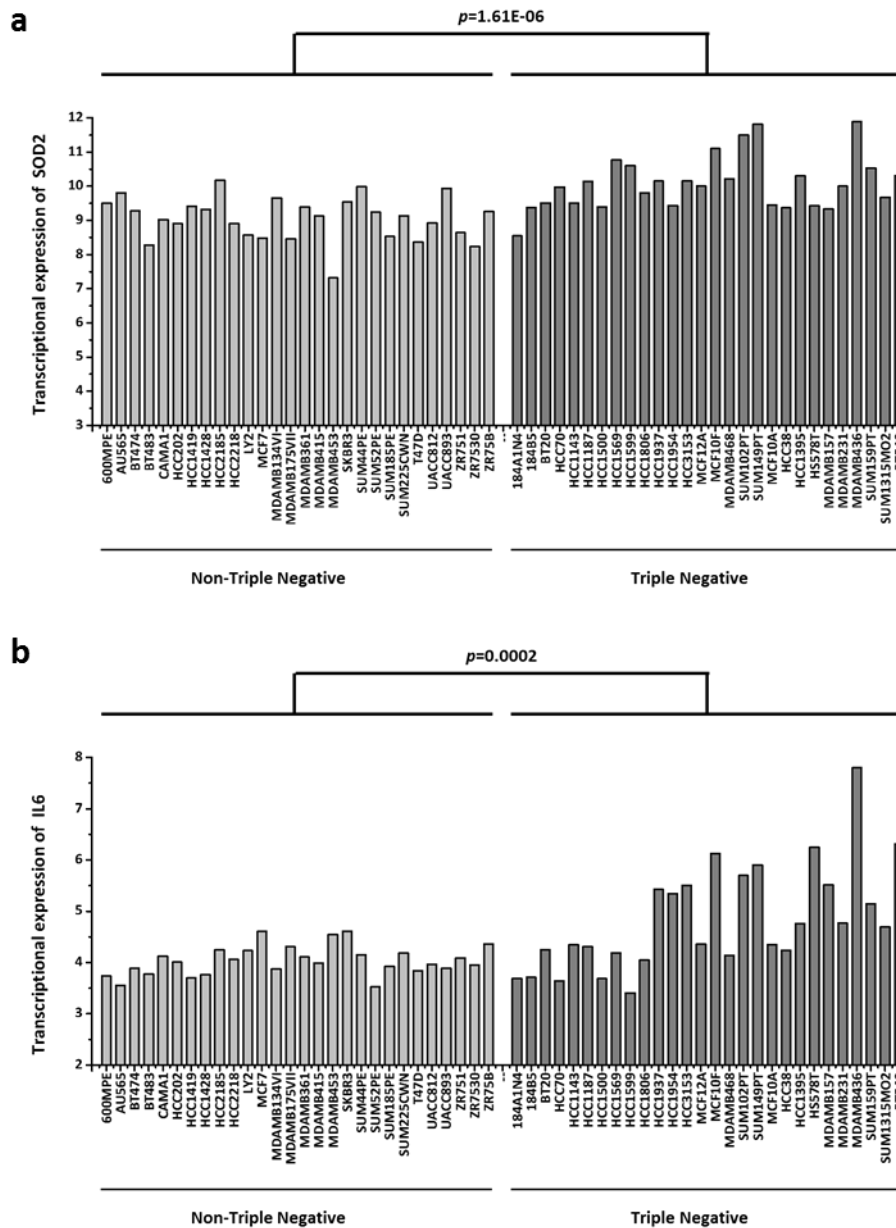
Data set	FOXA1, SOD2		FOXA1, IL6		SOD2, IL6	
	cor	p	cor	p	cor	p
METABRIC	-0.575	8.45E-168	-0.245	1.73E-27	0.304	1.73E-27
TCGA	-0.61	2.16E-47	-0.17	1.30E-08	0.186	4.80E-10
GSE24450	-0.646	5.67E-23	-0.55	1.08E-15	0.419	3.42E-09

Supplementary Figures

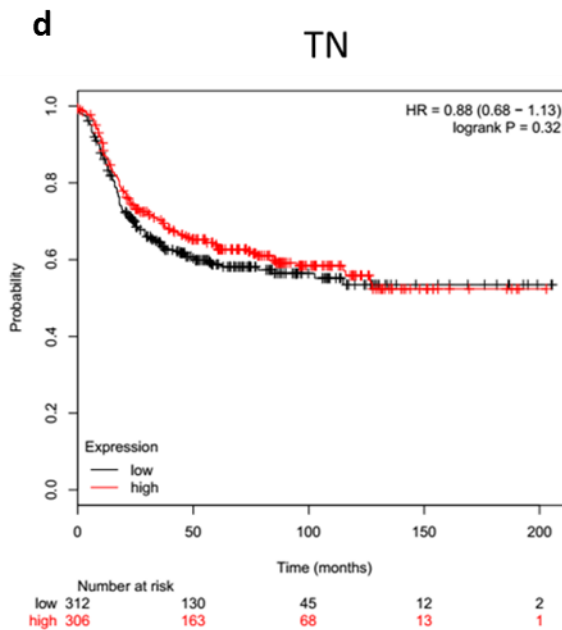
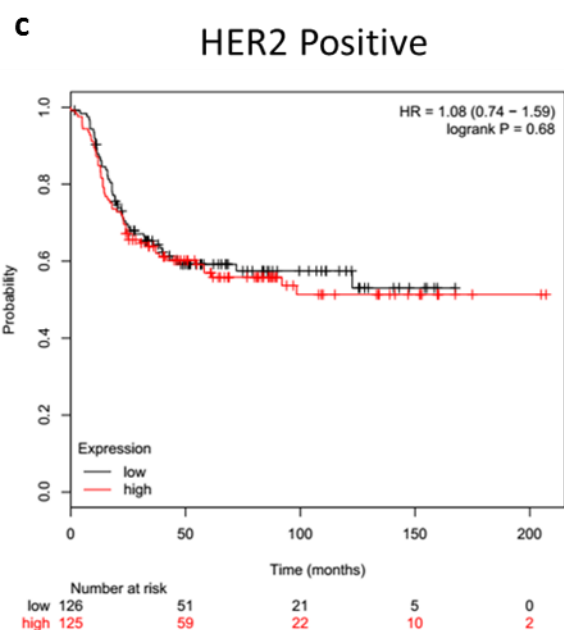
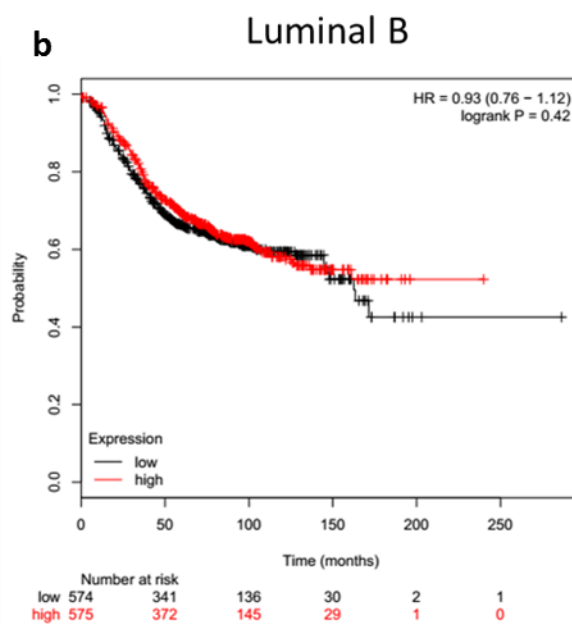
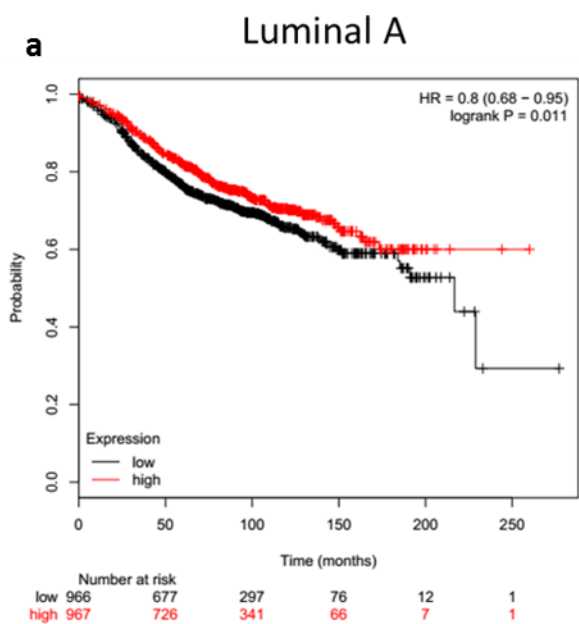
Supplementary Fig S1. *ER*, *FBP1*, *AGR2*, *MYO5C*, *WWP1* and *LRBA* expression profiles across breast cancer cell lines of different subtypes. (A) *ER* and (B) *FBP1* gene expression can differentiate luminal from non-luminal cell lines, (C) *AGR2* and (D) *MYO5C* gene expression can differentiate the luminal A from the rest cells, (E) *WWP1* and (F) *LRBA* gene expression can differentiate the luminal B cells at the transcriptional level.



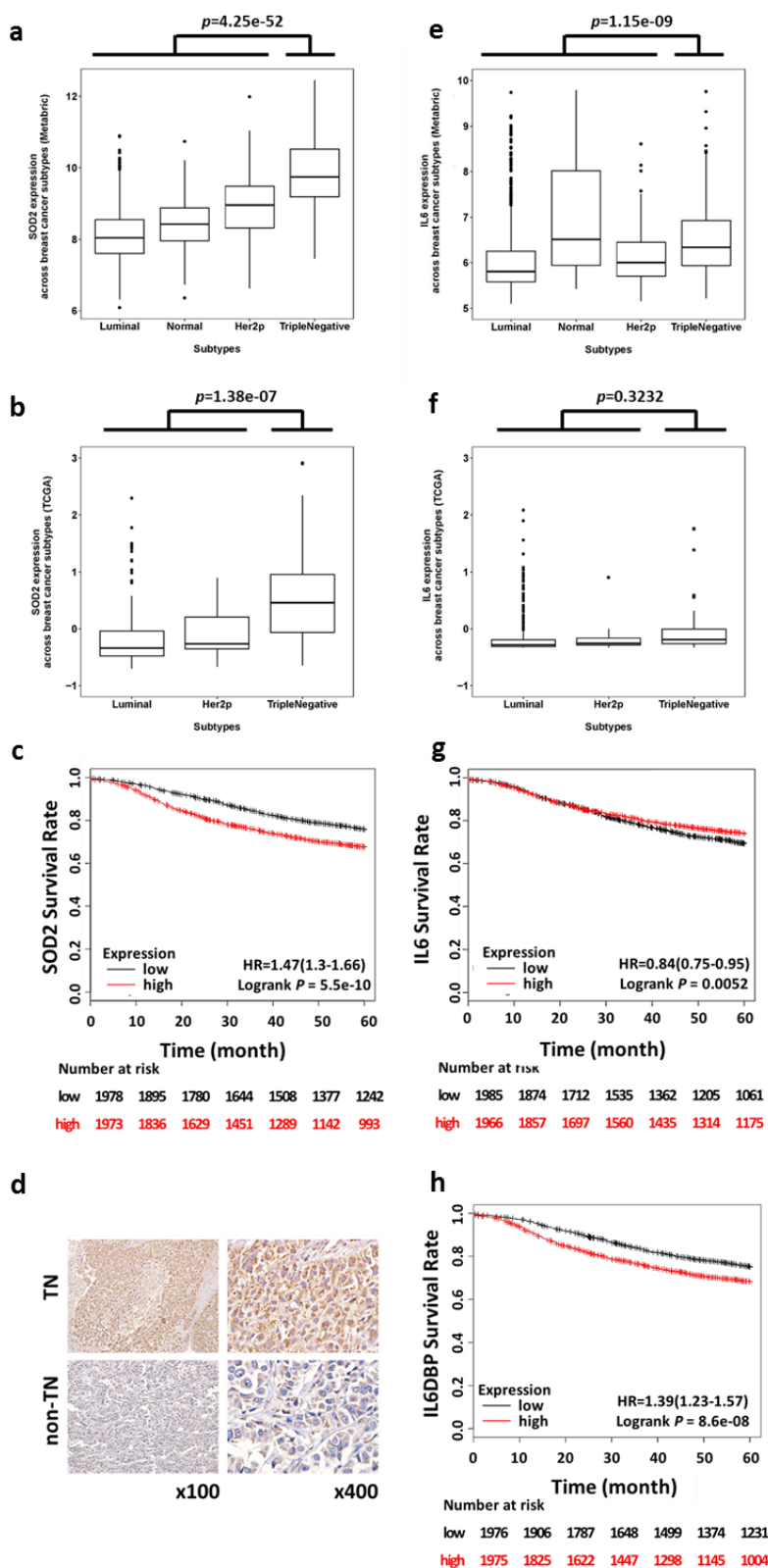
Supplementary Fig S2. Gene expression profile of (A) SOD2 and (B) IL6 across 56 breast cancer cell lines using E-MTAB-181 data. The 56 cell lines are stratified into Triple Negative and Non-Triple Negative subgroups, with the significance being evaluated using student's T test.



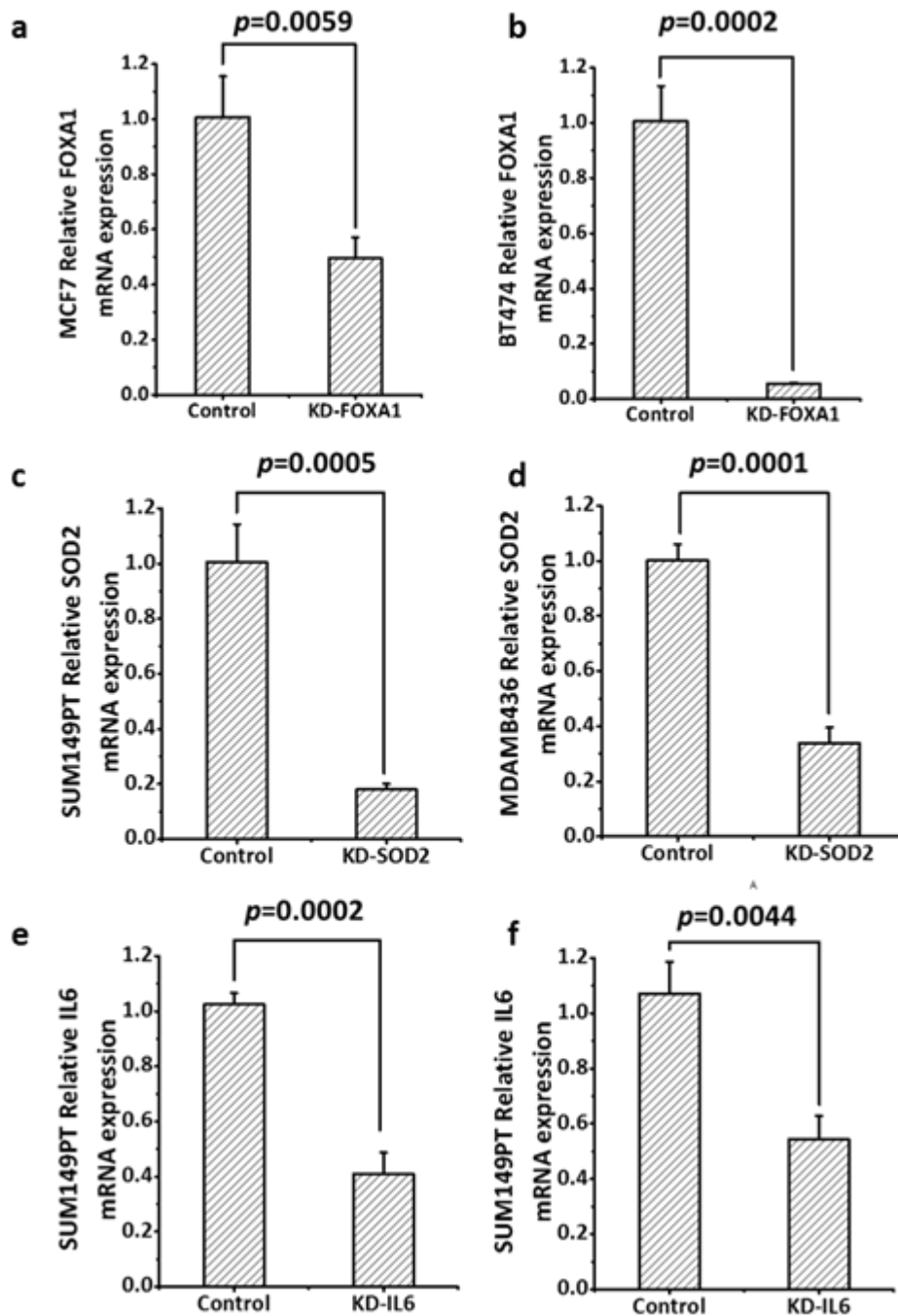
Supplementary Fig S3. Kaplan–Meier survival curves on the association between FOXA1 (probe is 204667_at) expression and breast cancer patient RFS in each molecular subtype using KM plotter. The subtypes are (A) luminal A, (B) luminal B, (C) HER2 positive, and (D) TNBCs.



Supplementary Fig S4. Clinical evidence evaluating the breast cancer subtyping potential of SOD2 and IL6. *SOD2* mRNA expression stratified by breast cancer subtypes using (A) METABRIC data and (B) TCGA data. (C) Patients 5 years' relapse free survival Kaplan–Meier curves of *SOD2* expression (probe is 215223_s_at) produced using KM plotter. (D) IHC staining of SOD2 in TNBC and non-TNBC tumors. *IL6* mRNA expression stratified by breast cancer subtypes using (E) METABRIC data and (F) TCGA data. Patients 5 years' relapse free survival Kaplan–Meier curves of (G) *IL6* (probe is 205207_at) and (H) *IL6DBP* (probe is 212501_at) produced using KM plotter.



Supplementary Fig S5. The knocking down efficiencies. The knocking down efficiencies of *FOXA1* in (A) MCF7 and (B) BT474 cells, that of *SOD2* in (C) SUM149PT and (D) MDAMB436 cells, and that of *IL6* in (E) SUM149PT and (F) MDAMB436 cells.



Supplementary Fig S6. Regulatory relationship between MYC and FOXA1 in HER2 positive cells. **(A)** Hypothesis of the triangle negative feedback loop involving FOXA1, HER2 and MYC. **(B)** The expression of MYC and HER2 in SKBR3 and MDAMB453 cells detected by q-PCR and western blotting. MYC shows distinct expression in these two HER2 positive cell lines. **(C)** The expression of FOXA1 and HER2 in SKBR3 cells transfected with siMYC that is detected by q-PCR and western blot. **(D)** The expression of MYC and HER2 in SKBR3 cells transfected with siFOXA1 and detected by q-PCR and western blot.

