

**Title:** Prognostic features of *STAT3* in an ER(+) breast cancer model system

## Supplementary Information

### Table of Contents (pages 1-8)

**Suppls.1-4** are summarized gene pools in four combined Tables starting with sets of their corresponding Venn diagrams. Some of gene symbols (~100) are left as “blank” within Tables based on annotation provided by Gene Spring GX7.3.1(March, 2011). The gene symbols high-lighted with light blue are either transcription factors or transcription factor subunits. Some of gene symbols high-lighted with light yellow are molecules of interest. We only mark “Y” in each table that means the gene symbol to be one of components listed in the analysis of interest.

**Suppl. 5** contains heatmaps for the common gene pools shared by both 11 signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A), respectively.

**Suppl. 6** has the prognostic value(s) of component(s) in the *STAT3* subnetworks (Figure S6.2) and not in the *STAT3* subnetworks (Figure S6.3) in 90A (Groups IE and IIE) and 72A (Luminal A and Luminal B), respectively. Two controls in survival analyses are made for pair-wise comparisons between subtypes used in this study (see Figure S6.1). Additional Survival analyses are provided (Table S6.1).

**Suppl. 7** contains ANOVA test results on genes of interest in 90A and 72A, respectively. They are presented by ten mean plots and a table.

### Suppls 1-4 (pages 9-261)

#### Suppl.1.

**Table S 1.1.** Clinically significant *STAT3* transcriptional regulatory network for mitotic count parameter in 90 ER(+) infiltrating ductal carcinomas (IDCs).

**Table S1.2.** The gene pool for the *MYC* transcriptional regulatory network relevant to clinical mitotic count parameter in ER(+)IDCs (i.e. 90A).

**Table S1.3.** The gene pool for the overlapped network of *MYC&STAT3* in ER(+)IDCs (i.e. 90A) relevant to clinical mitotic count parameter.

**Table S1.4.** The gene pool for the *STAT3* transcriptional regulatory network in ER(+) IDCs (i.e. 90A).

**Table S1.5.** The gene pool for the *MYC* transcriptional regulatory network in ER(+) IDCs (i.e. 90A).

**Table S1.6.** The common gene pool for the overlapped network of *MYC&STAT3* in ER(+)IDCs (i.e. 90A).

**Table S1.7.** The shared gene pool between the overlapped network of *MYC&STAT3* relevant to ER(+) IDCs and that significant to clinical mitotic count parameter.

**Table S1.8.** The gene pool for the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 72A).

**Table S1.9.** The shared gene pool between overlapped networks of *MYC&STAT3* in 90A and 72A and their non-overlapping gene pools.

**Table S1.10.** The *FOXCI* transcriptional regulatory network in 90 ER(+) infiltrating ductal carcinomas.

**Table S1.11.** The *FOXCI* transcriptional regulatory network in 72 ER(+) infiltrating ductal carcinomas.

## **Suppl. 2.**

**Table S2.1.** The shared gene pool between the overlapped networks of *MYC&STAT3* in 90A (for ER(+) IDCs) and 77A (for ER(-) IDCs) and their non-overlapping gene pools. The gene pool shared by two univariate overlapped networks of *MYC* and *STAT3* in ER(+) and ER(-) IDCs indicating different percentages of shared network activities for *MYC* and *STAT3* in two subtypes (81.93% (77A) versus 45.81% (90A)), respectively. A cohort (77A) consisting of 48 triple negatives (TN) and 29 ERBB2+ was used (see reference 9 in the main text).

**Table S2.2.** The shared gene pool between networks of *MYCnSTAT3* in 90A (for ER(+)

IDCs) and 77A (for ER(-) IDCs) and their non-overlapping gene pools.

The gene pool shared by two networks of *MYCnSTAT3* in ER(+) and ER(-) IDCs indicating different percentages of shared network activities for *MYC* and *STAT3* in two subtypes (73.14% (77A) versus 45.75% (90A)), respectively.

**Table S2.3.** The common gene pool overlapped by networks of *ESR1*, *MYC* and *STAT3* in ER(+) IDCs (i.e. 90A).

**Table S2.4.** The common gene pool overlapped by networks of *ESR1n STAT3* and *MYCnSTAT3* in ER(+) IDCs (i.e. 90A).

**Table S2.5.** The common gene pool overlapped by networks of *MYC & STAT3* and *MYCnSTAT3* in ER(+) IDCs (i.e. 90A).

**Table S2.6.** The common gene pool overlapped by networks of *ESR1 & STAT3* and *ESR1nSTAT3* in ER(+) IDCs (i.e. 90A).

**Table S2.7.** Venn diagrams and their overlapping and non-overlapping gene pools for both networks of *ESR1* and *MYC* in ER(+) IDCs (i.e. 90A). The common gene pool overlapped by networks of *ESR1 & MYC* (S2.7a), the potential *ESR1* regulated gene pool but not in *MYC* network (S2.7b) and the potential *MYC* regulated gene pool but not in *ESR1* network (S2.7c) are listed.

### **Suppl. 3.**

**Table S3.1.** The common gene pool between age-related gene pool in ER(+) IDCs (i.e.90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.2.** The common gene pool between grade-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.3.** The common gene pool between LNM-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.4.** The common gene pool between LVI-related gene pool in ER(+) IDCs

(i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.5.** The common gene pool between LYM-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.6.** The common gene pool between MC-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.7.** The common gene pool between NP-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.8.** The common gene pool between size-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.9.** The common gene pool between stage-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

**Table S3.10.** The common gene pool between TF-related gene pool in ER(+) IDCs (i.e. 90A) and the overlapped network of *MYC&STAT3* not only relevant to ER(+) IDCs but to clinical mitotic count parameter.

#### **Suppl. 4.**

**Table S4.1.** The common gene pool in both cell cycle signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.2.** The common gene pool in both VEGF signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.3.** The common gene pool in both BER signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).



**Table S4.4.** The common gene pool in both DRS signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.5.** The common gene pool in both HR signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.6.** The common gene pool in both MRP signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.7.** The common gene pool in both NER signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.8.** The common gene pool in p53 signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.9.** The common gene pool in both proteasome signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.10.** The common gene pool in both ribosome signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.11.** The common gene pool in both PDGFRB signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.12.** The common gene pool in both ERBB2 signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Table S4.13.** The common gene pool shared between the multivariate space of *STAT3* network (*MYCnSTAT3*) and ERBB2 signal transduction pathway in ER(+) IDCs (i.e. 90A).

**Table S4.14.** The common gene pool in both ERBB2 signal transduction pathway and the *MYC & STAT3* overlapped network in ER(-) IDCs (i.e. 77A). ER(-) IDCs contain 48 triple negatives (TN) and 29 ERBB2+.

**Table S4.15.** The common gene pool shared between the *MYCnSTAT3* network and ERBB2 signal transduction pathway in ER(-) IDCs (i.e. 77A).

## Suppl.5 (pages 262-267)

**Figure S5.1.** Heatmaps for the common gene pool in both cell cycle signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.2.** Heatmaps for the common gene pool in both VEGF signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.3.** Heatmaps for the common gene pool in both BER signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.4.** Heatmaps for the common gene pool in both DRS signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.5.** Heatmaps for the common gene pool in both HR signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.6.** Heatmaps for the common gene pool in both MRP signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.7.** Heatmaps for the common gene pool in both NER signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.8.** Heatmaps for the common gene pool in both p53 signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.9.** Heatmaps for the common gene pool in both proteasome signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A).

**Figure S5.10.** Heatmaps for the common gene pool in both ribosome signal transduction pathway and the overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. 90A). Grey color within the heatmap stands for data not available after data processing.

In this case, we do not use this piece of information for further discussion due to most data not available.

**Figure S5.11.** Heatmaps for the common gene pool in both PDGFRB signal transduction pathway and the overlapped network of *MYC*&*STAT3* in ER(+) IDCs (i.e. 90A).

#### **Suppl.6 (pages 268-278)**

**Figure S6.1.** The survival analyses on subtypes of ER(+) infiltrating ductal carcinoma in this study. Figure S6.1a. stands for survival analysis on Group IE vs. Group IIE; Figure S6.1b. stands for survival analysis on Luminal A vs. Luminal B. Both results are not significant.

**Figure S6.2.** The survival analyses on probes in the *STAT3* driven subnetworks in coupling with its major transcription factor partner - *MYC* for (a) proliferation; (b) sustained angiogenesis; (c) Warburg effect; (d) ES-like phenotype; (e) *FOXC1* transcriptional regulatory network and (f) ERBB2 signaling in ER(+) IDCs. We compared the probes in the subnetworks in both 90A and 72A for predicting clinical outcome that was evaluated by overall disease free survival. Their results are summarized in Figure 3F (please see main text).

**Figure S6.3.** The survival analyses on a probe –*FOXC1*(12715) not in the *STAT3* subnetwork. We compared this probe in the subnetworks in both 90A and 72A for predicting clinical outcome that was evaluated by overall disease free survival. It is a significant favorable prognostic factor in 90A cohort but not in 72A cohort.

**Table S6.1.** Univariate and multivariate analyses for survival on prognostic factors in 90A and 181A cohorts, respectively. The p values of tests in the Cox proportional hazard model to be less or equal to 0.05 are high-lighted with light blue. Both univariate and multivariate analyses are negative except univariate COXPH analysis on subcohort 1 versus non subcohort1.

#### **Suppl. 7 (279-284)**

**Figure S7.1** Mean plot analyses of *MELK* mRNA levels in eight clinical categories of two cohorts of infiltrating ductal carcinomas (IDCs).

Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis (LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90IDCs) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see

the main text for definitions).

**Figure S7.2.** Mean plot analyses of *METAP2* mRNA levels in eight clinical categories of two cohorts of infiltrating ductal carcinomas (IDCs).

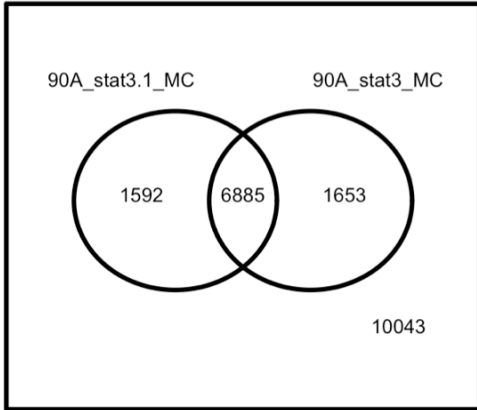
**Figure S7.3.** Mean plot analyses of *SRC* mRNA levels in eight clinical categories of two cohorts of infiltrating ductal carcinomas (IDCs).

**Figure S7.4.** Mean plot analyses of a transcript variant *SRC* mRNA levels in eight clinical categories of two cohorts of infiltrating ductal carcinomas (IDCs).

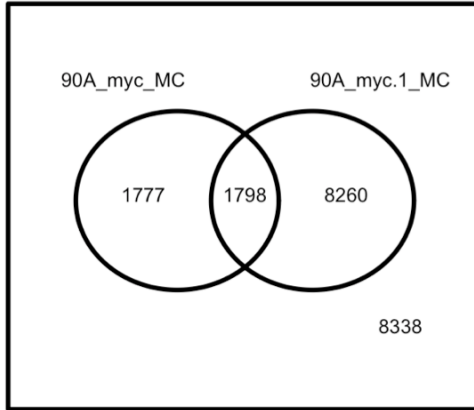
**Figure S7.5.** Mean plot analyses of *OIP5* mRNA levels in eight clinical categories of two cohorts of infiltrating ductal carcinomas (IDCs).

**Table S7.1.** ANOVA tests on *ARNT* mRNA levels in eight clinical categories of two cohorts of infiltrating ductal carcinomas (IDCs).

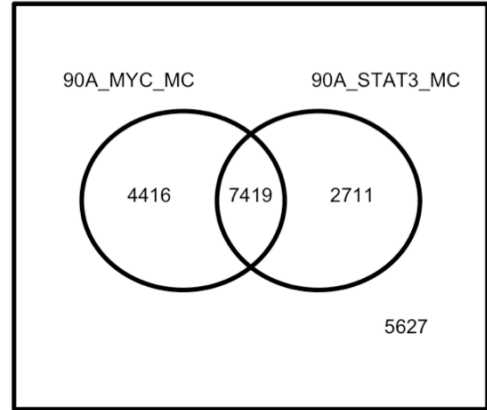
S1.1



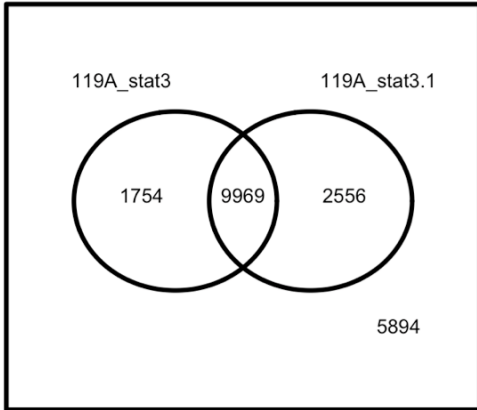
S1.2



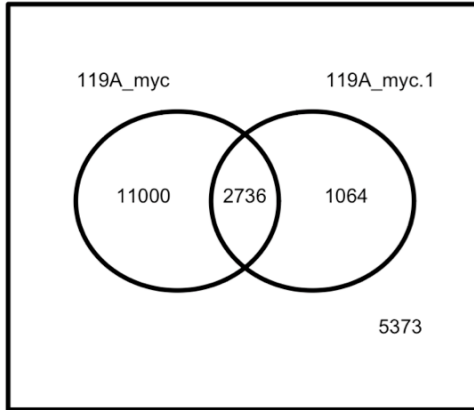
S1.3



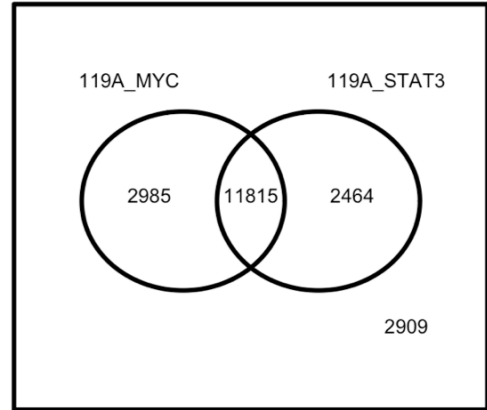
S1.4



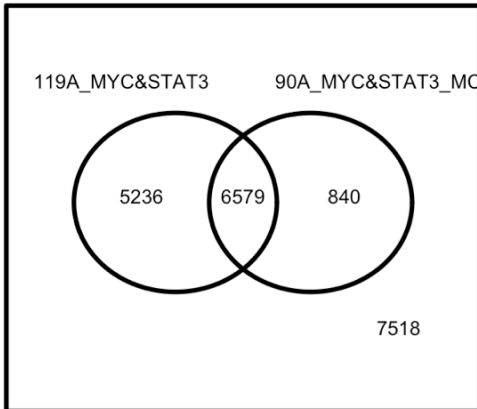
S1.5



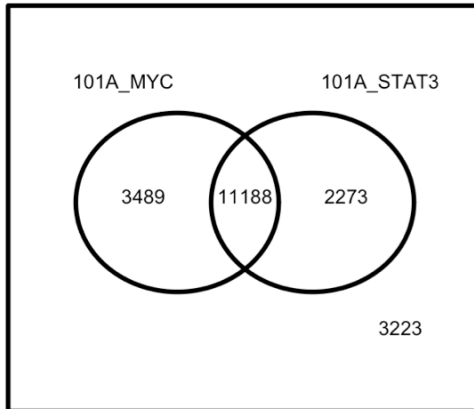
S1.6



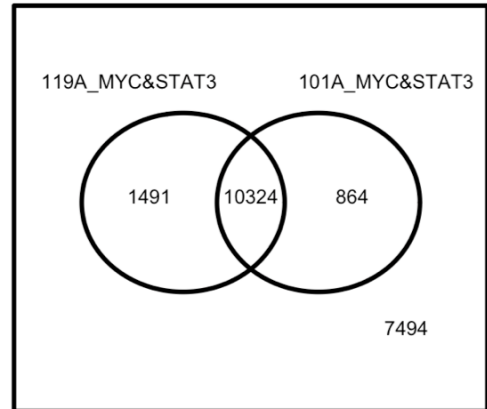
S1.7



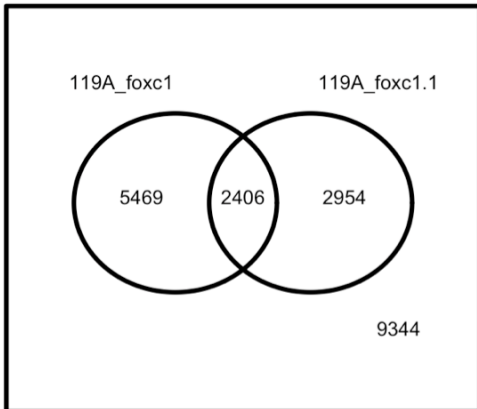
S1.8



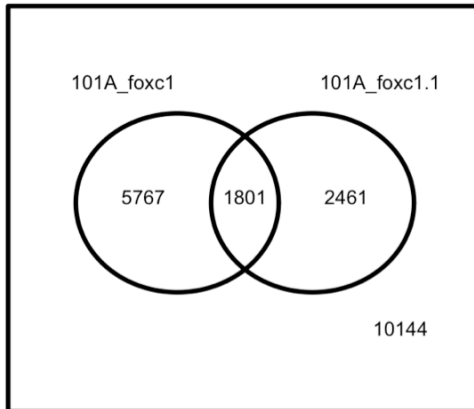
S1.9



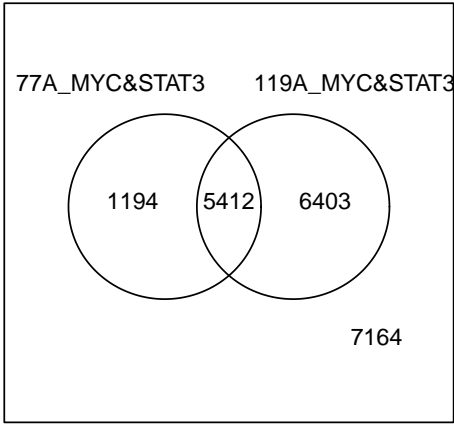
S1.10



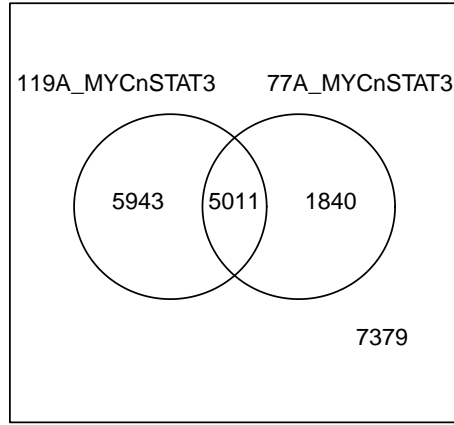
S1.11



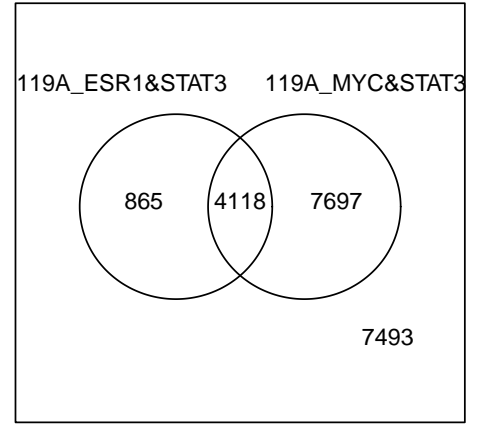
S2.1



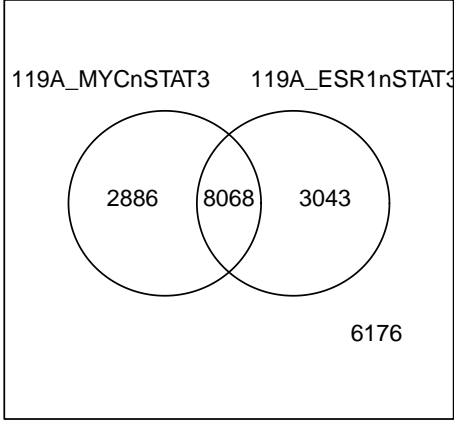
S2.2



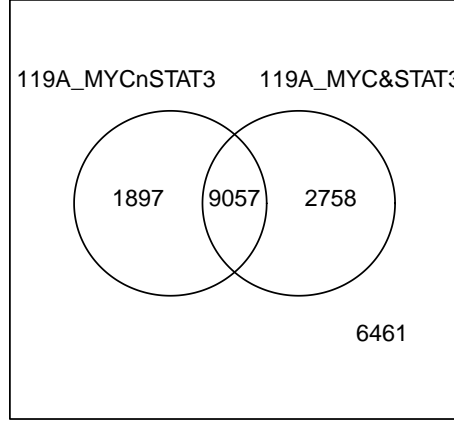
S2.3



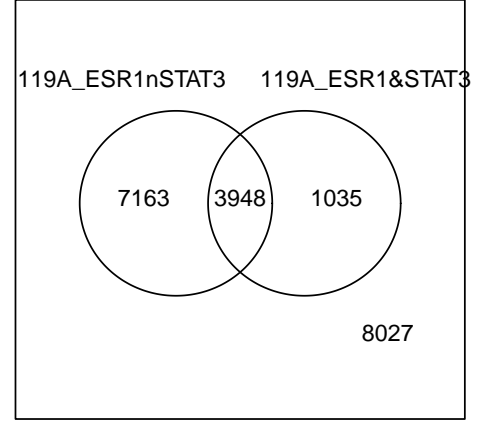
S2.4



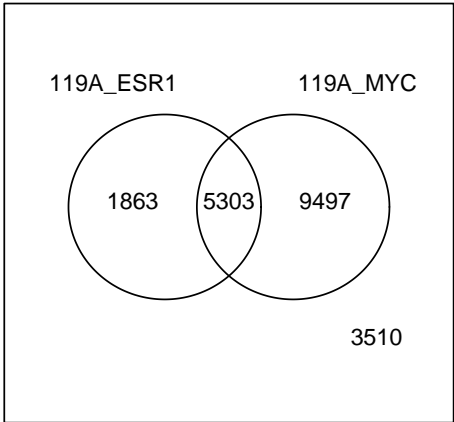
S2.5

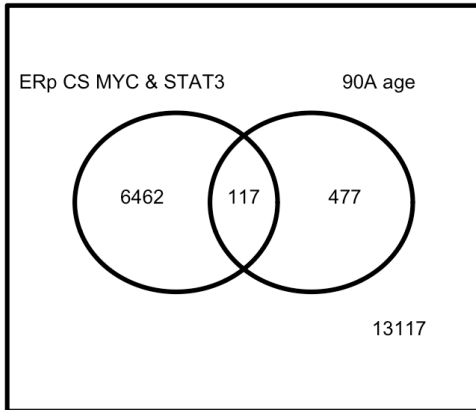
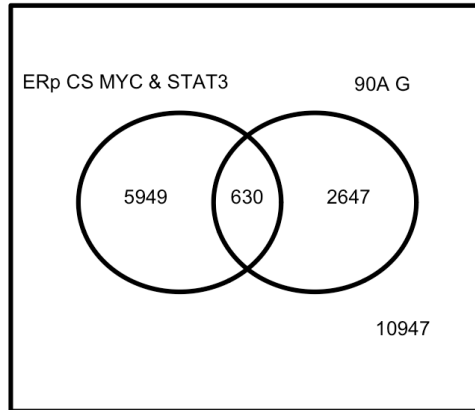
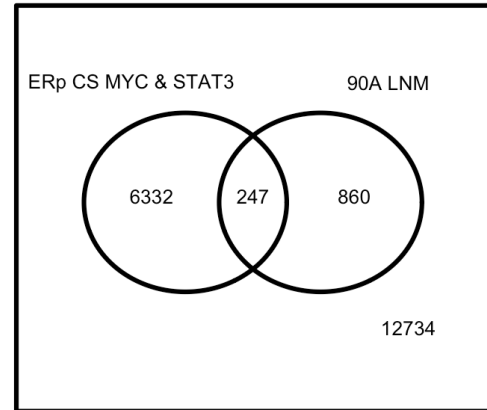
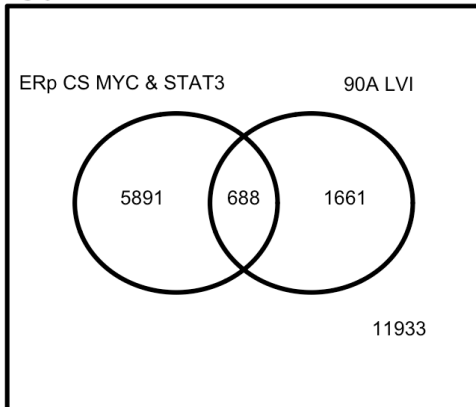
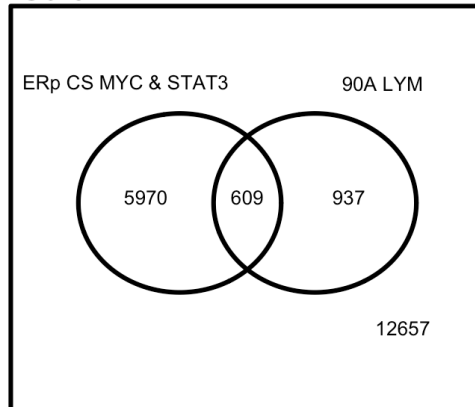
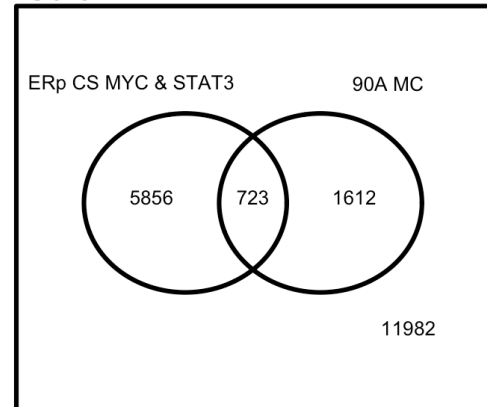
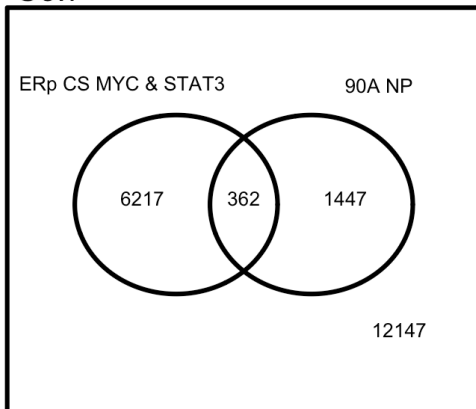
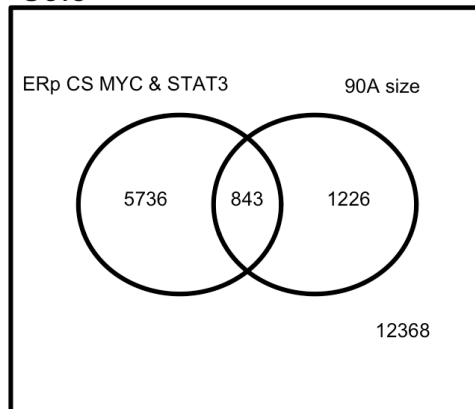
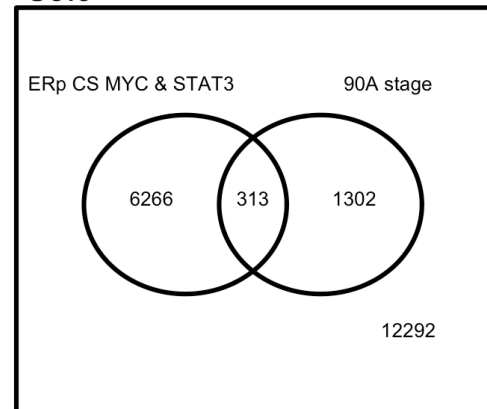
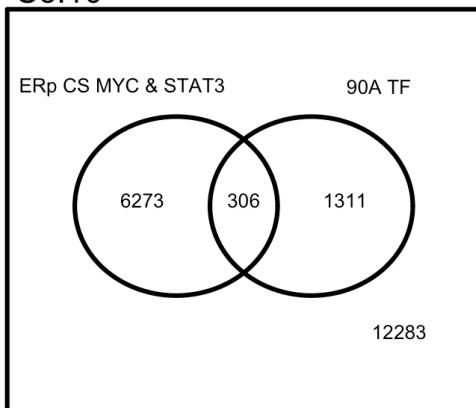


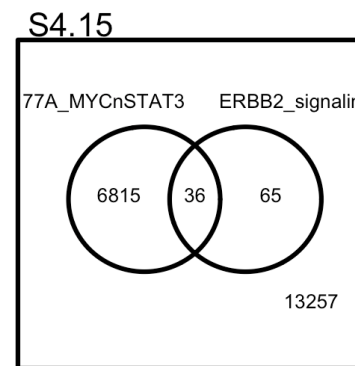
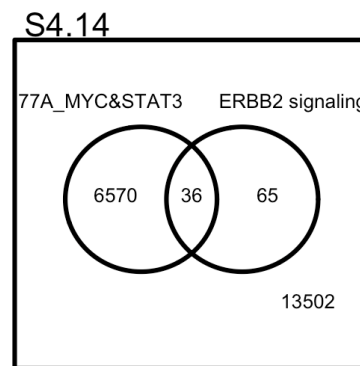
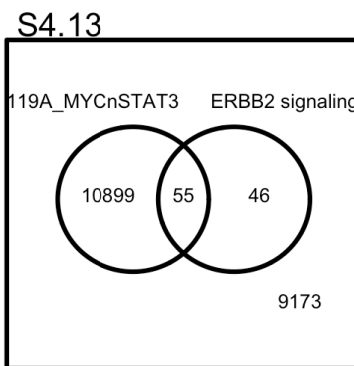
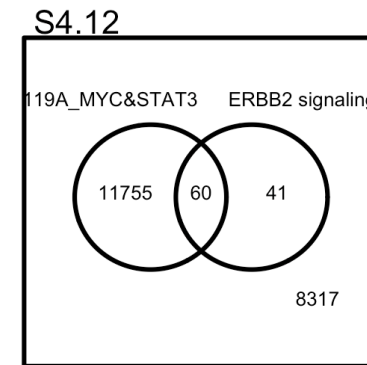
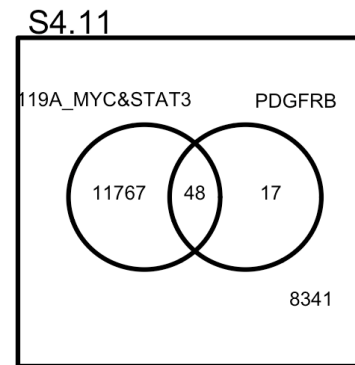
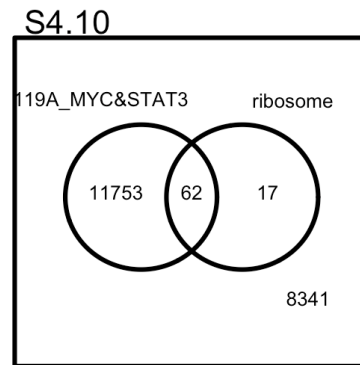
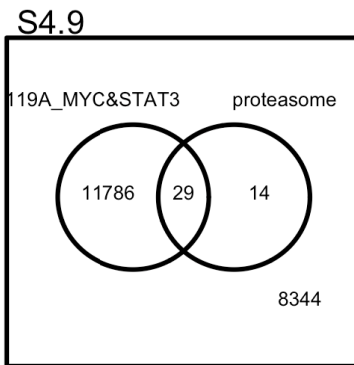
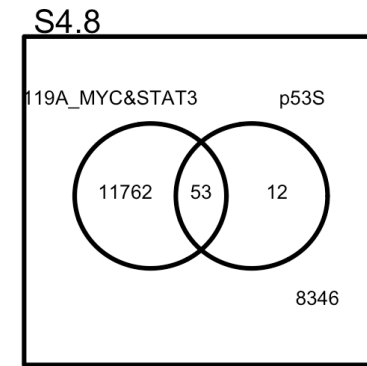
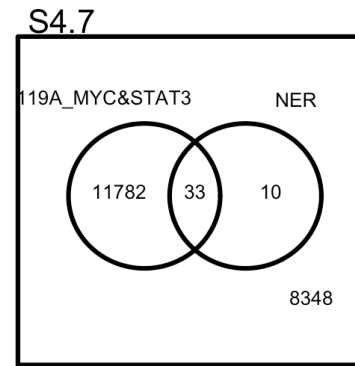
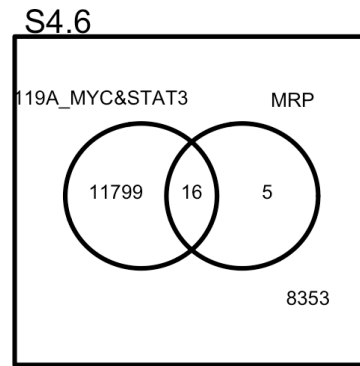
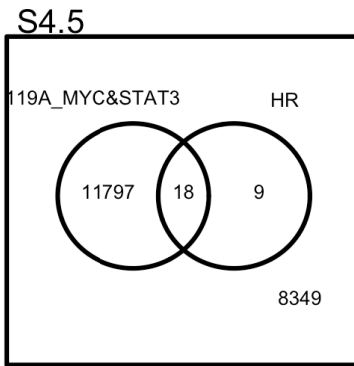
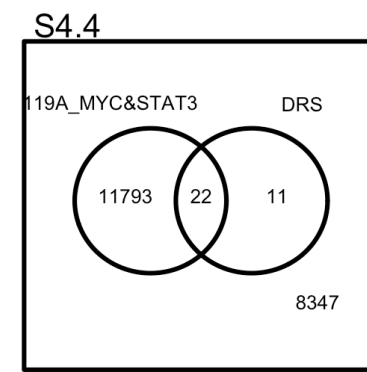
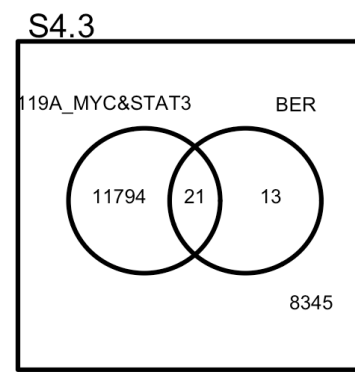
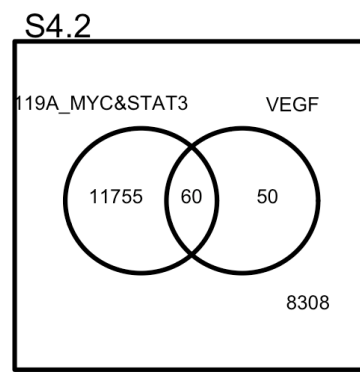
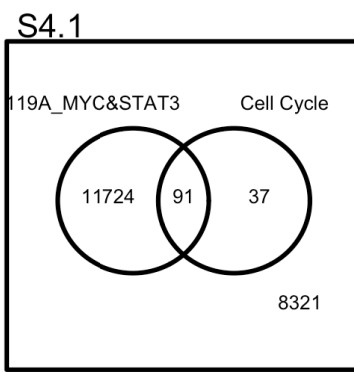
S2.6



S2.7



**S3.1****S3.2****S3.3****S3.4****S3.5****S3.6****S3.7****S3.8****S3.9****S3.10**













































































































































































































































































































































































































































































































































































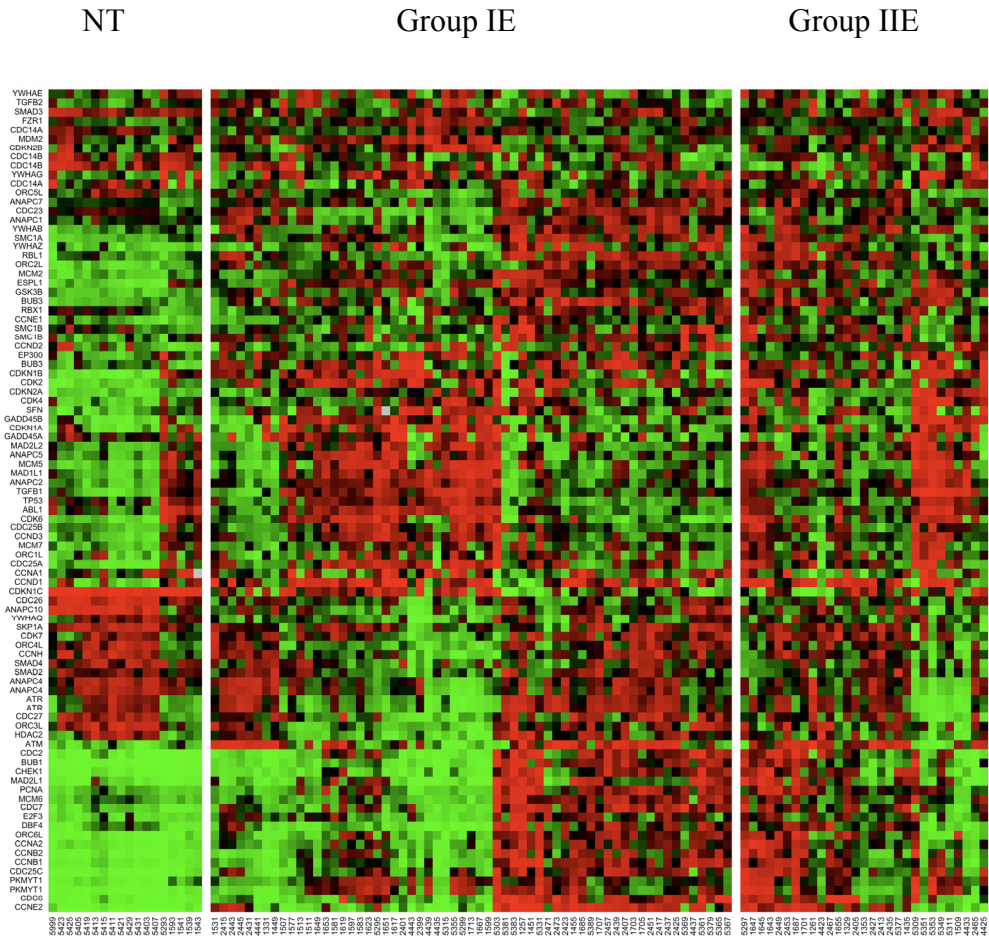




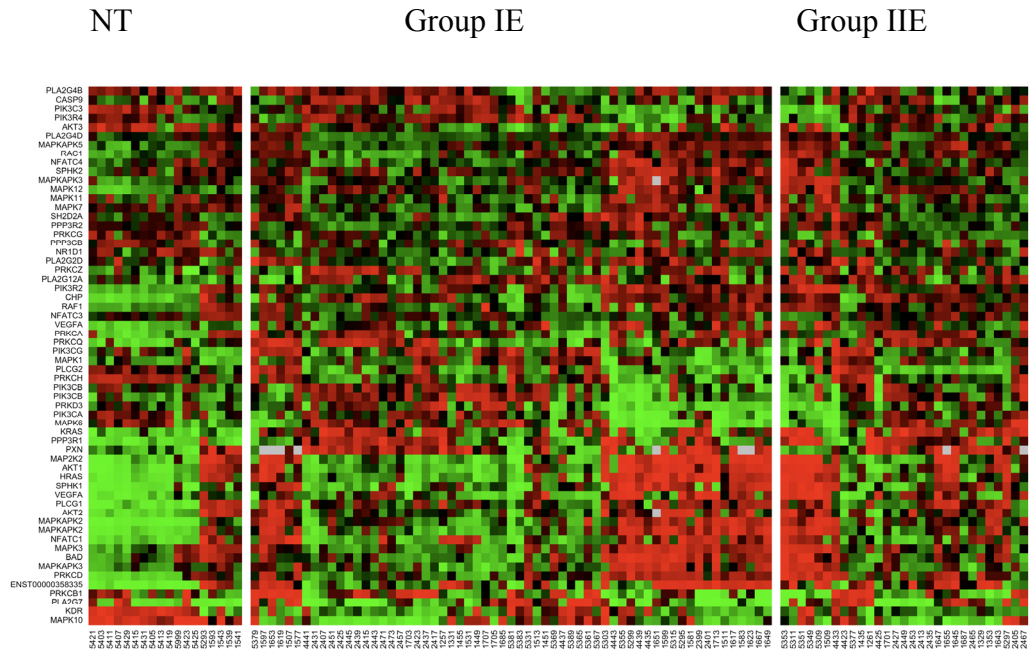




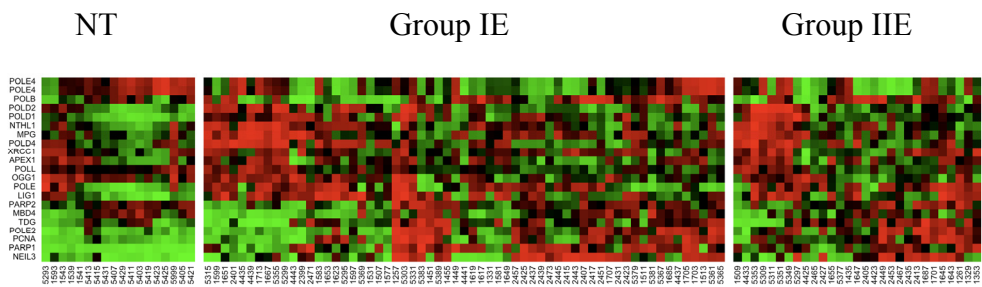
**Figure S5.1.** Heatmaps for the common gene pool in both cell cycle signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).



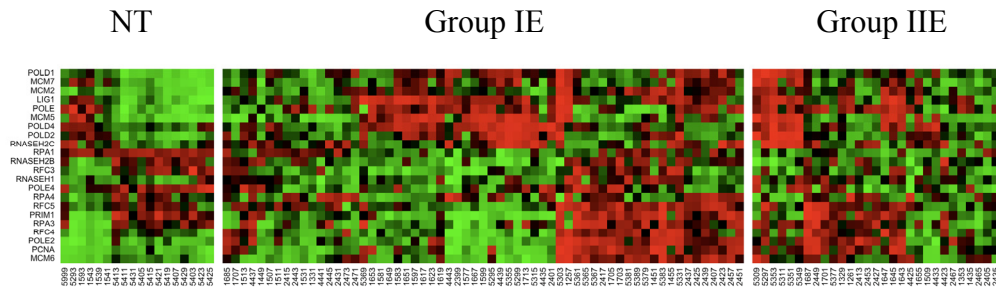
**Figure S5.2.** Heatmaps for the common gene pool in both VEGF signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).



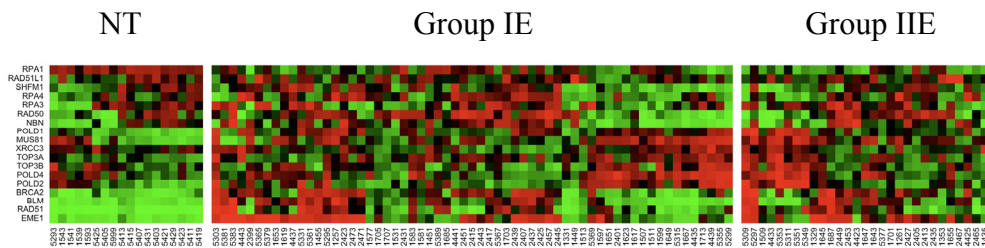
**Figure S5.3.** Heatmaps for the common gene pool in both BER signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).



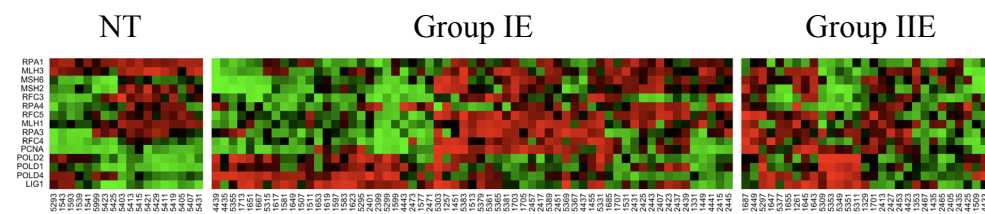
**Figure S5.4.** Heatmaps for the common gene pool in both DRS signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).



**Figure S5.5.** Heatmaps for the common gene pool in both HR signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).

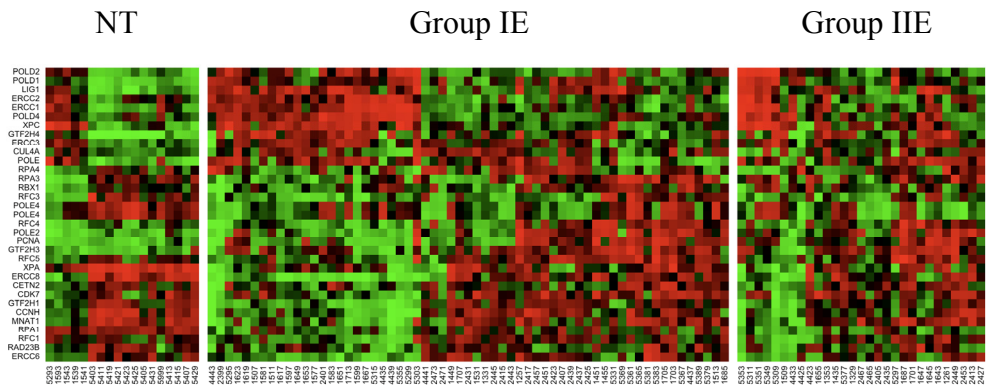


**Figure S5.6.** Heatmaps for the common gene pool in both MRP signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).

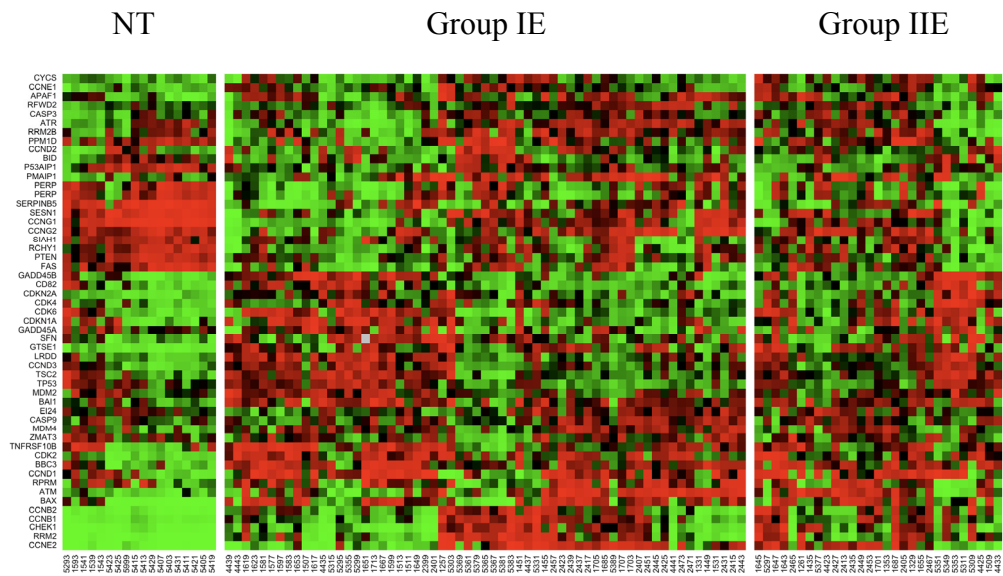


**Figure S5.7.** Heatmaps for the common gene pool in both NER signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from

90A cohort).

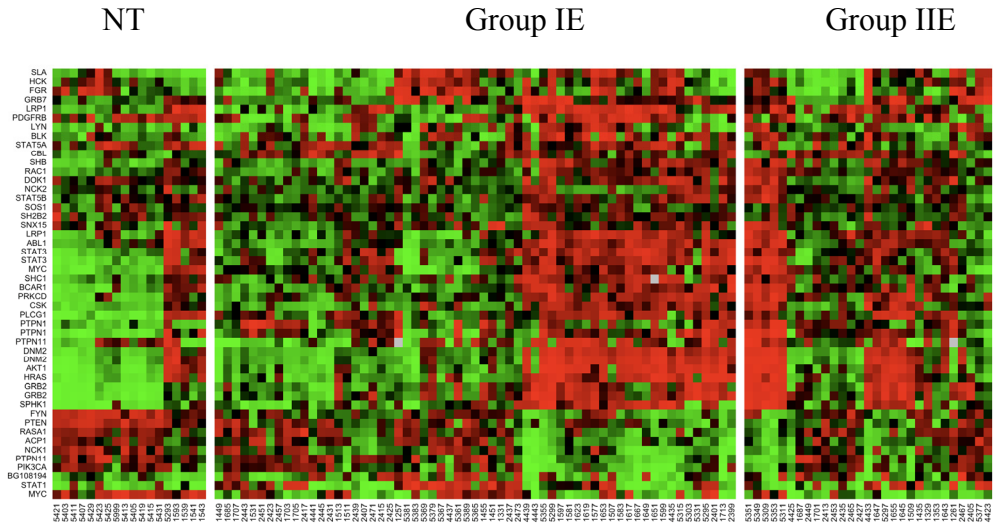


**Figure S5.8.** Heatmaps for the common gene pool in both p53 signal transduction pathway and overlapped network of *MYC&STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).





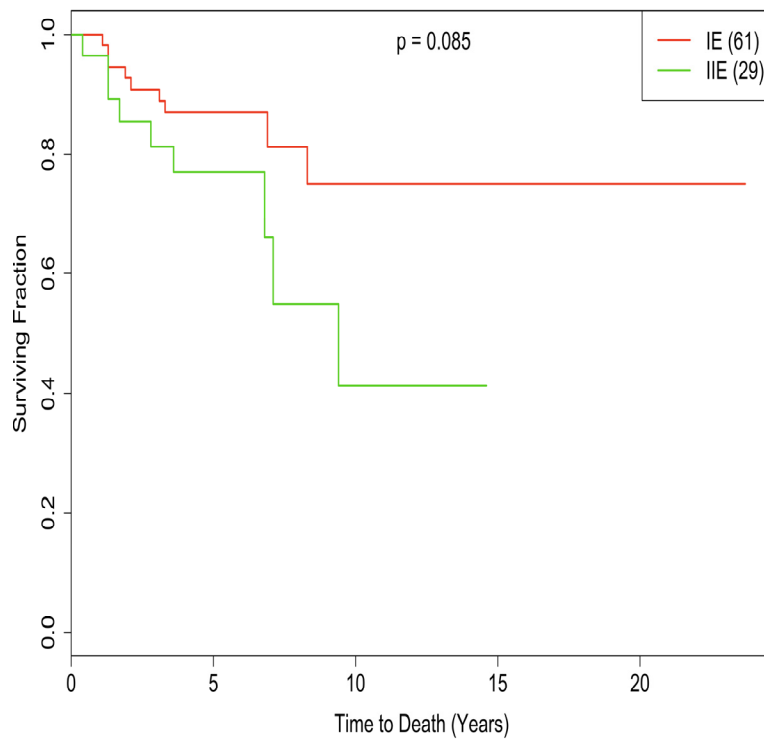
**Figure S5.11.** Heatmaps for the common gene pool in both PDGFRB signal transduction pathway and overlapped network of *MYC* & *STAT3* in ER(+) IDCs (i.e. IDCs from 90A cohort).



**Figure S6.1.** The survival analyses on subtypes of ER(+) infiltrating ductal carcinoma.

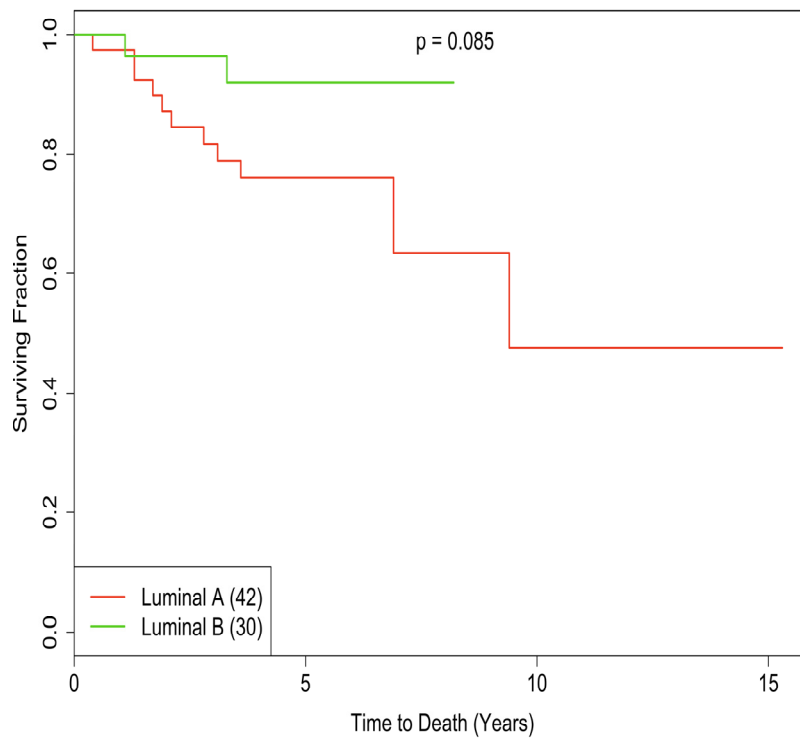
Fig. S6.1a demonstrates survival analysis on Group IE vs. Group IIE; Fig. S6.1b demonstrates survival analysis on Luminal A vs. Luminal B.

a. Group IE vs. Group IIE



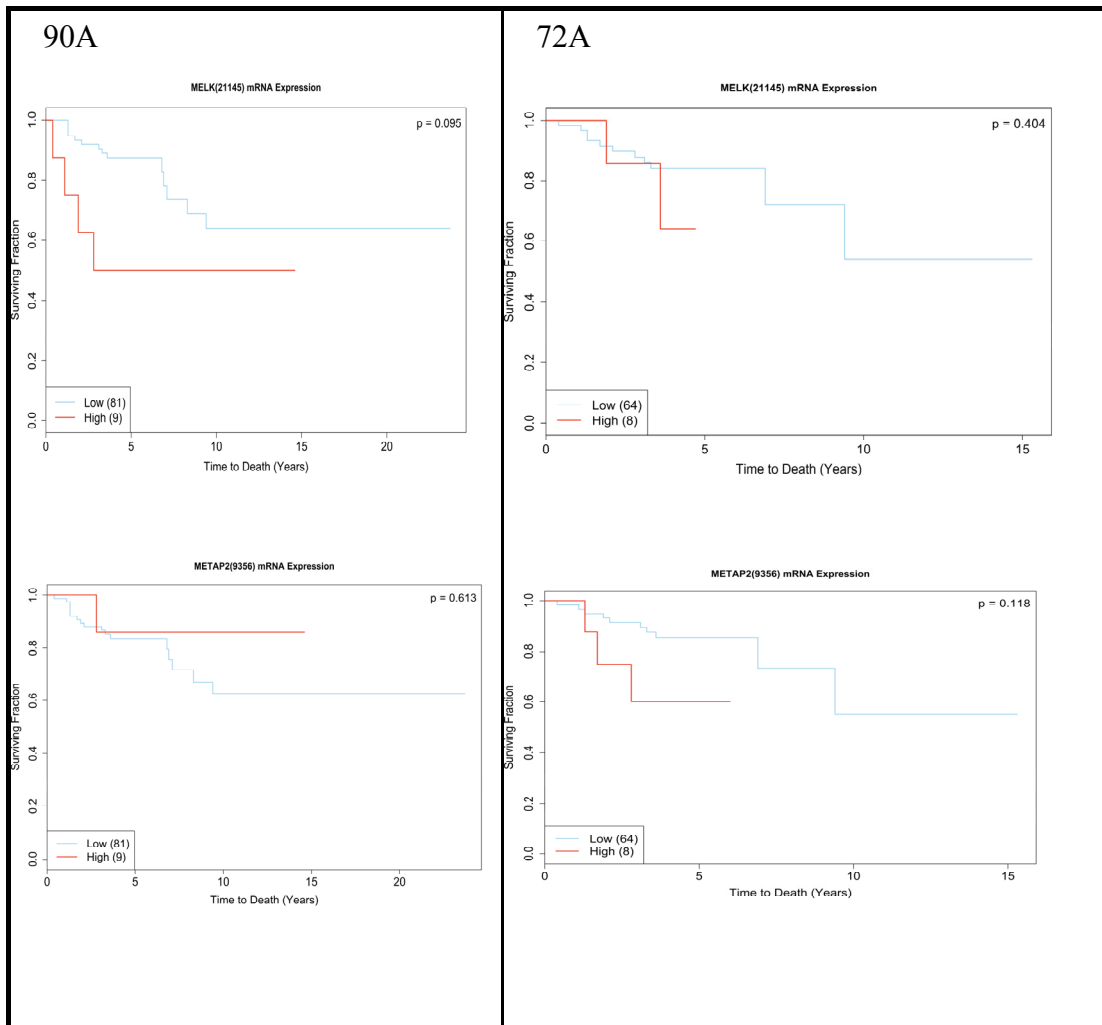


b. Luminal A vs. Luminal B

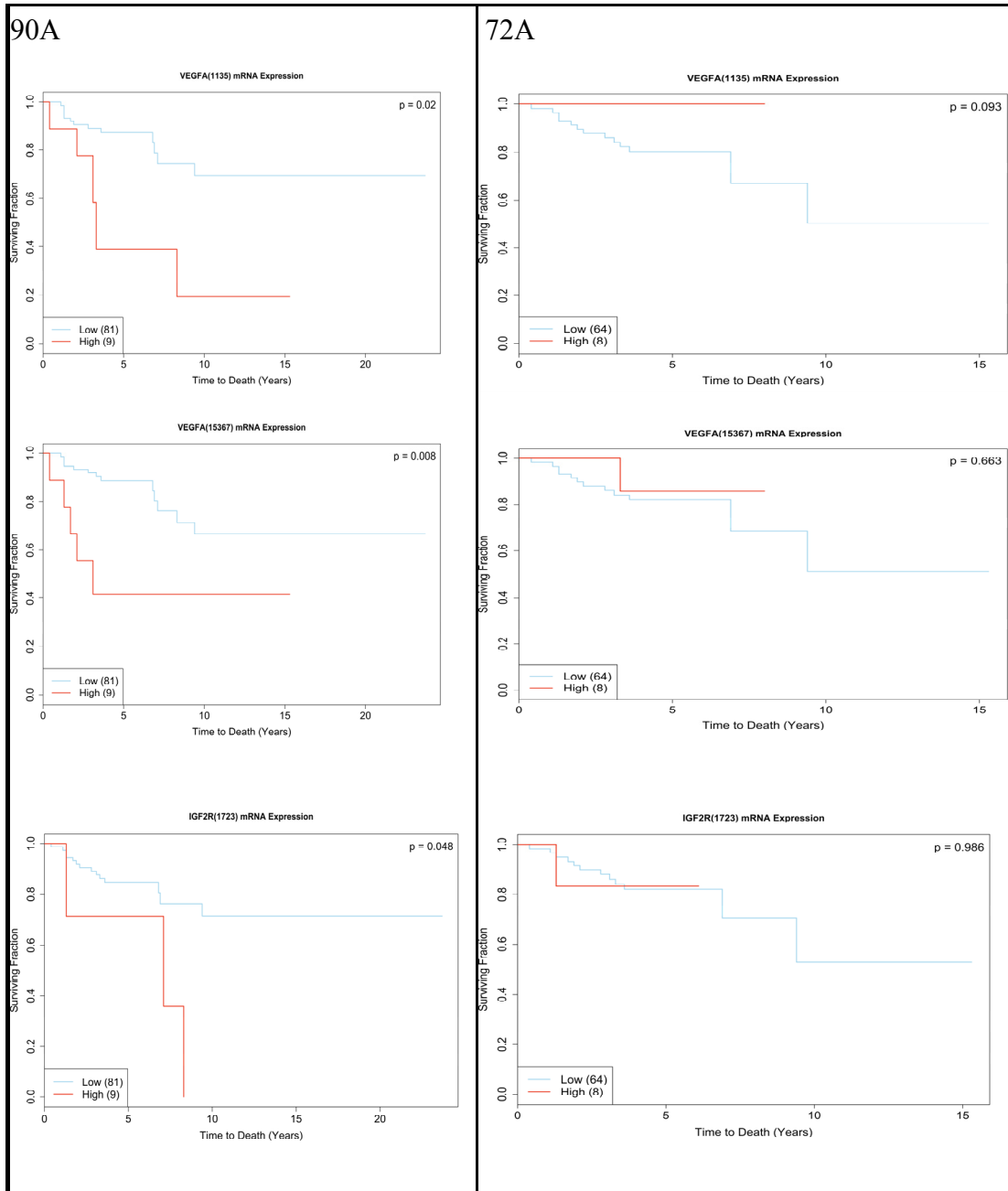


**Figure S6.2.** The survival analyses on probes in *STAT3* driven subnetworks in coupling with its major transcription factor partner - *MYC* for (a) proliferation; (b) sustained angiogenesis; (c) Warburg effect; (d) ES like phenotype; (e) *FOXCI* transcriptional regulatory network and (f) ERBB2 signaling in ER(+) IDCs. We compared the probes in those subnetworks (a - f) in cohorts 90A and 72A for predicting clinical outcome that was evaluated by overall disease free survival.

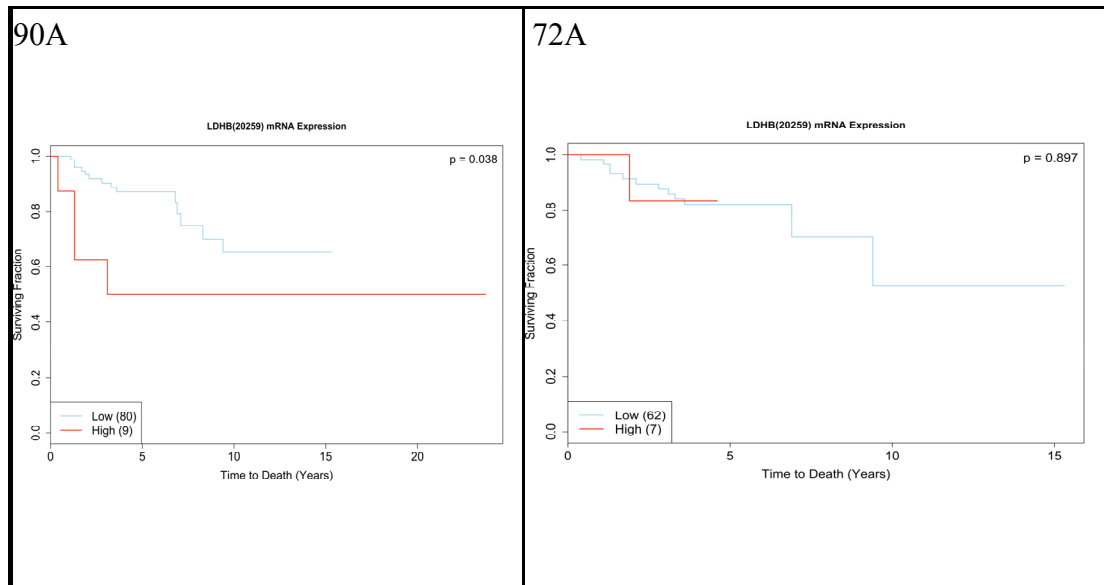
a. *STAT3* subnetwork components for proliferation



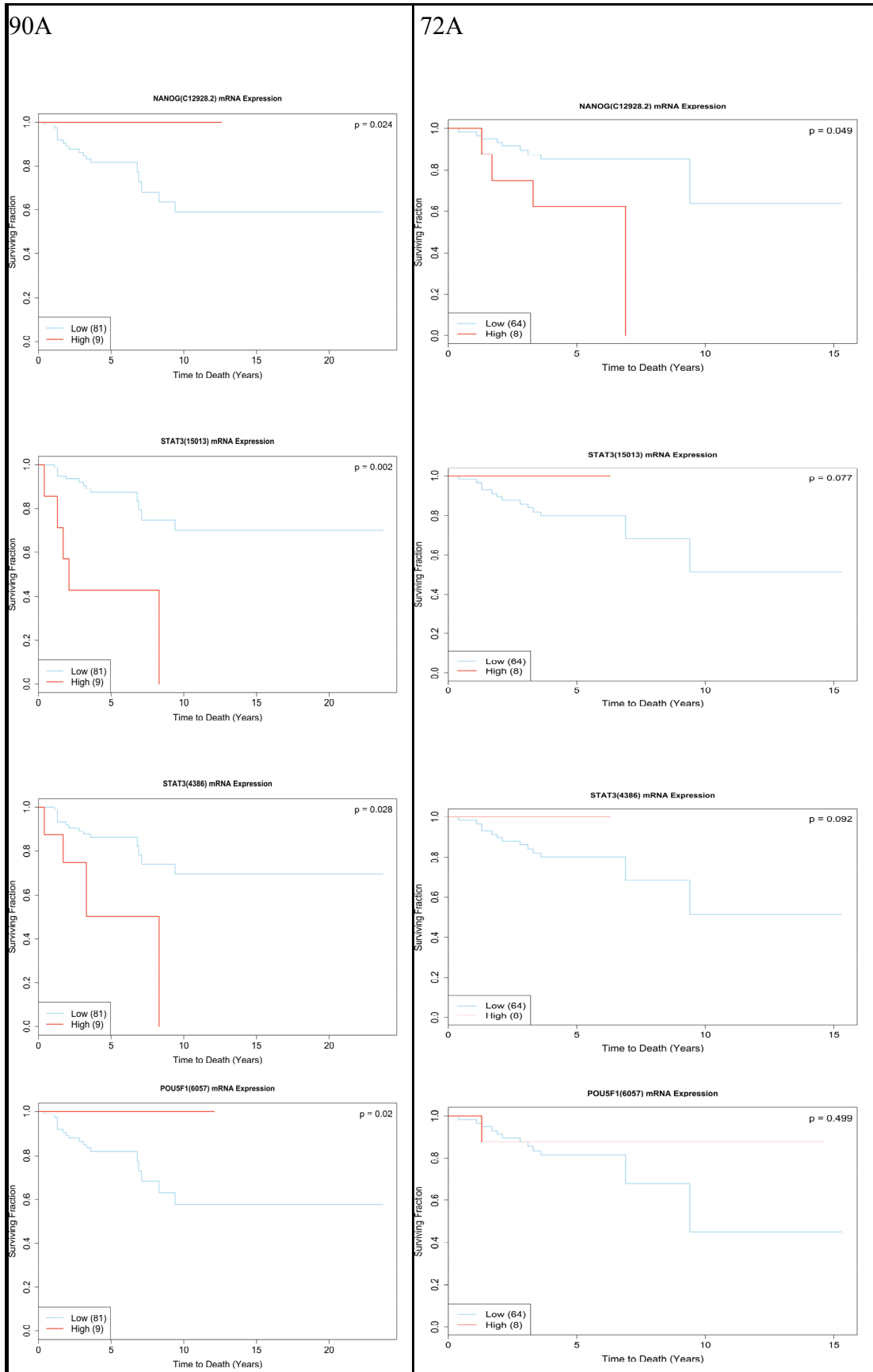
b. *STAT3* subnetwork components for sustained angiogenesis



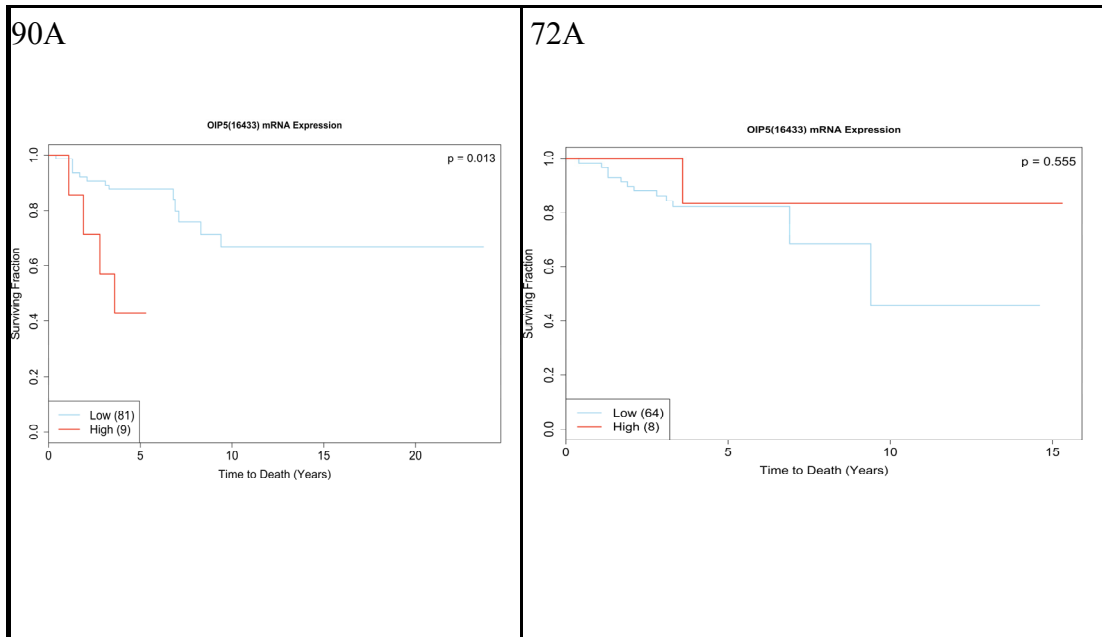
c. *STAT3* subnetwork components for Warburg effect



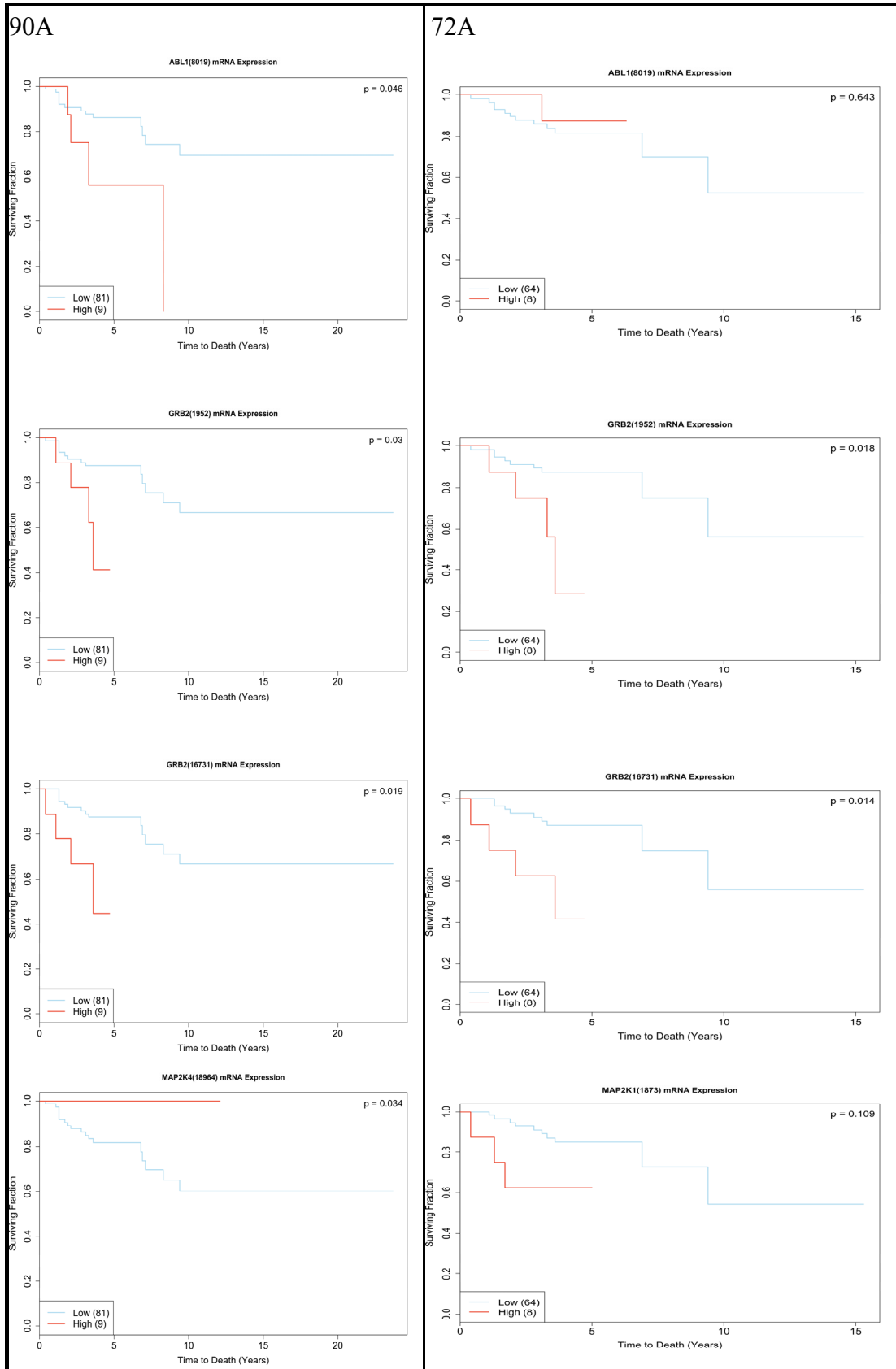
d. *STAT3* subnetwork components for ES like phenotype



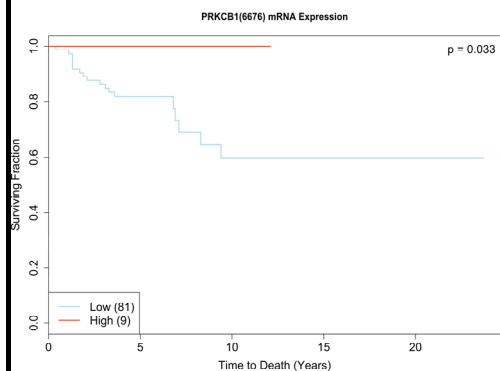
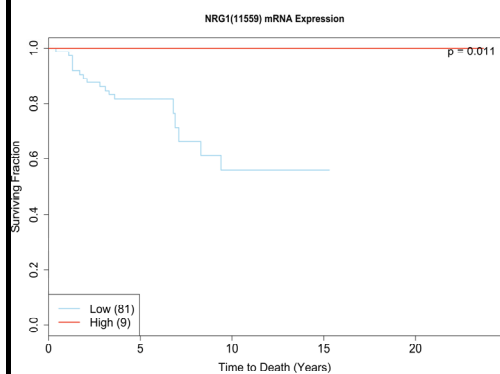
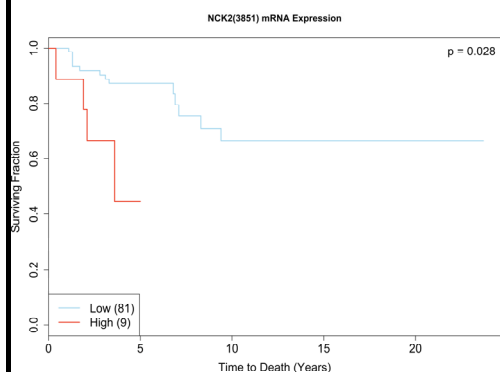
e. *STAT3* subnetwork components in *FOXCI* driven subnetwork



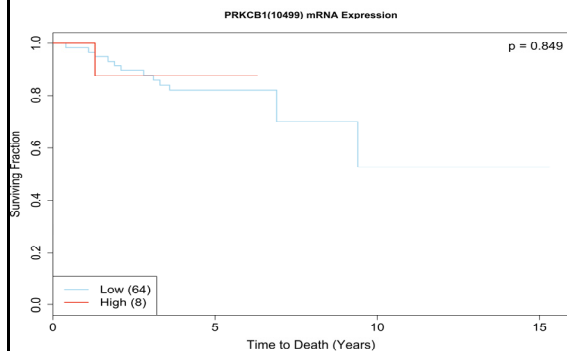
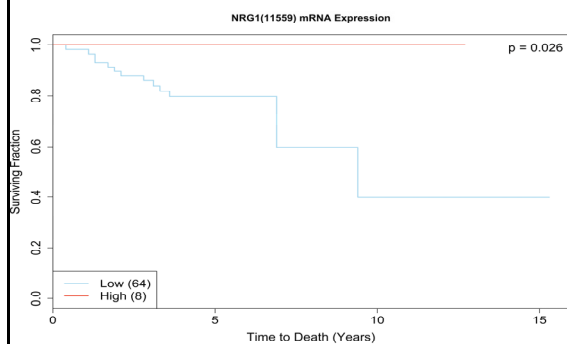
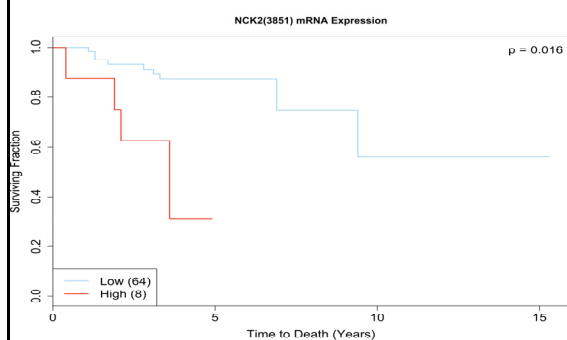
f. *STAT3* subnetwork components for ERBB2 signal transduction pathway



90A

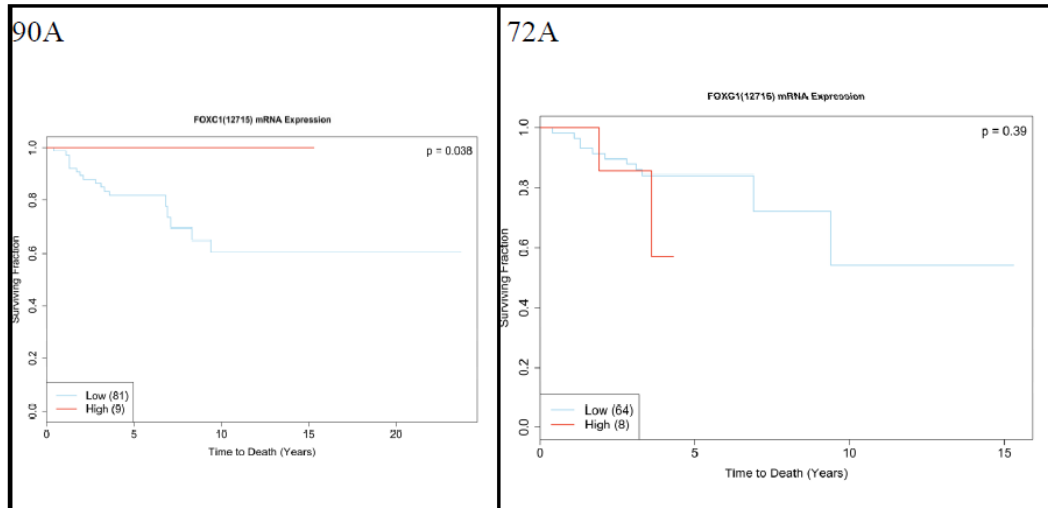


72A





**Figure S6.3.** The survival analysis on a probe - FOXC1(12715) not in the *STAT3* subnetwork. We compared this probe in cohorts 90A and 72A for predicting clinical outcome that was evaluated by overall disease free survival.

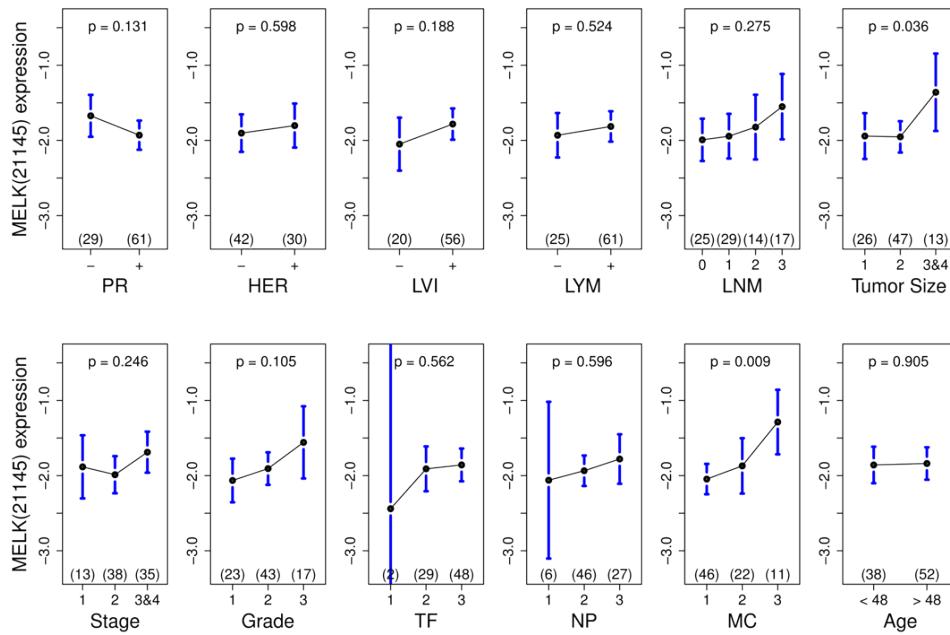


**Table S6.1.** Univariate and multivariate analyses for survival on prognostic factors in 90A cohorts. The p values of tests in the Cox proportional hazard model are provided.

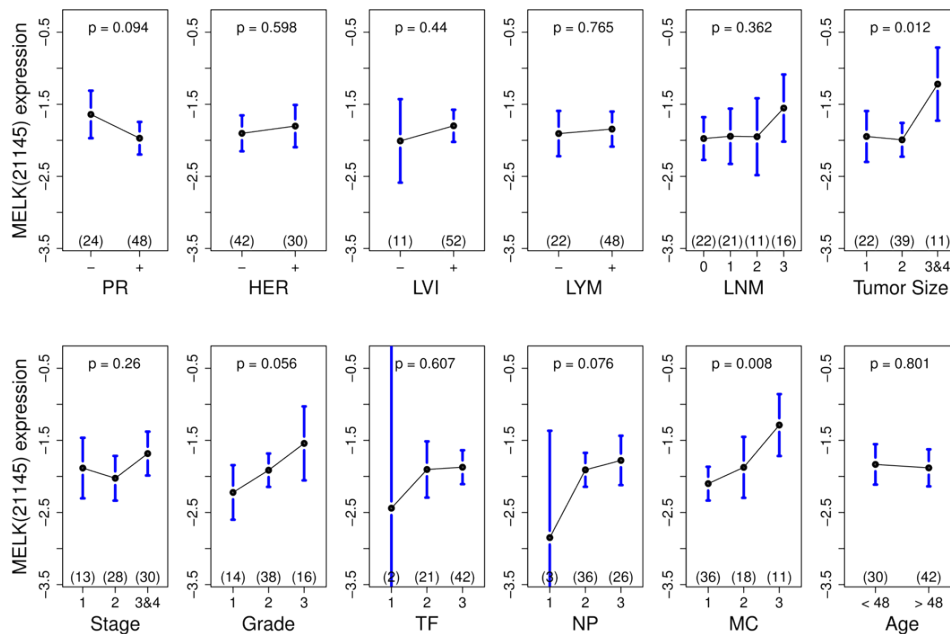
Prognostic factor	90A													
	Univariate Analysis		Multivariate Analysis		Multivariate Analysis		Multivariate Analysis		Multivariate Analysis		Multivariate Analysis		Multivariate Analysis	
	Hazard Ratio	P value	Hazard Ratio	P value	Hazard Ratio	P value	Hazard Ratio	P value	Hazard Ratio	P value	Hazard Ratio	P value	Hazard Ratio	P value
Grade1 vs. 2&3	1.92	0.32	1.28	1.00	0.37	0.39	0.62	0.69	0.41	0.41	0.39	0.39	0.45	0.42
LVInegative vs. positive	3.30	0.13	0.18	0.99	4.81	0.34	58593665.52	1.00	10.15	0.11	2.32	0.51	3.67	0.35
Size 1 vs. 2&3 &4	1.76	0.38	1.39	1.00	0.69	0.67	0.71	0.70	0.87	0.88	0.76	0.75	1.21	0.82
NP 1 vs. 2&3	80907452.95	1.00	0.22	1.00	87730350.72	1.00	430020716.88	1.00	307737605.19	1.00	130070686.52	1.00	288537097.52	1.00
TF 1 vs. 2&3	25794474.04	1.00	0.00	0.99	1.00	1.00	0.00	1.00	3.24	1.00	0.00	1.00	2.70	1.00
LN0 vs. 1& 2&3	6.82	0.06	409376117590.53	0.65	567995703.41	1.00	689162786.60	1.00	457946953.13	1.00	2804267118.40	1.00	3974365499.84	1.00
MC 1 vs.2&3	1.39	0.59	26154.48	0.95	0.68	0.62	0.44	0.32	0.85	0.84	0.62	0.53	0.73	0.68
Stage 1vs, 2&3&4	2.32	0.42	8560.57	0.97	2.01	1.00	0.53	1.00	2.27	1.00	1.60	1.00	0.86	1.00
LVM positive vs. negative	6.49	0.07	0.00	0.56	0.25	0.35	1.00	1.00	0.38	0.48	0.11	0.07	0.07	0.07
15 gene signature (subcohort 1/2)	0.00	1.00	0.00	0.93	-	-	-	-	-	-	-	-	-	-
15 gene signature (subcohort 1/3)	0.47	0.16	-	-	0.40	0.37	-	-	-	-	-	-	-	-
15 gene signature (subcohort 2/3)	311133550.19	1.00	-	-	-	-	444898936.60	1.00	-	-	-	-	-	-
15 gene signature (subcohort 1/non1)	0.34	0.04	-	-	-	-	-	-	0.23	0.11	-	-	-	-
15 gene signature (subcohort 2/non2)	291402675.98	1.00	-	-	-	-	-	-	-	-	353542959.48	1.00	-	-
15 gene signature (subcohort 3/non3)	0.67	0.44	-	-	-	-	-	-	-	-	-	-	0.55	0.55

**Figure S7.1.** Mean plot analyses of mRNA levels for *MELK* in eight clinical categories and in two cohorts of infiltrating ductal carcinoma (IDCs), respectively. Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis(LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90A) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see main text for definitions).

A. Cohort 1

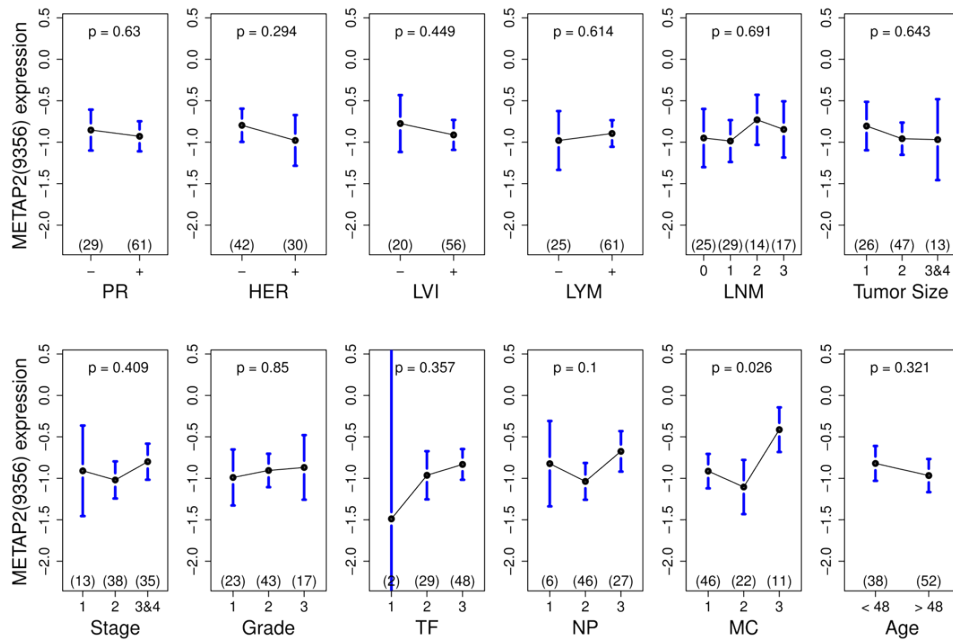


B. Cohort 2.

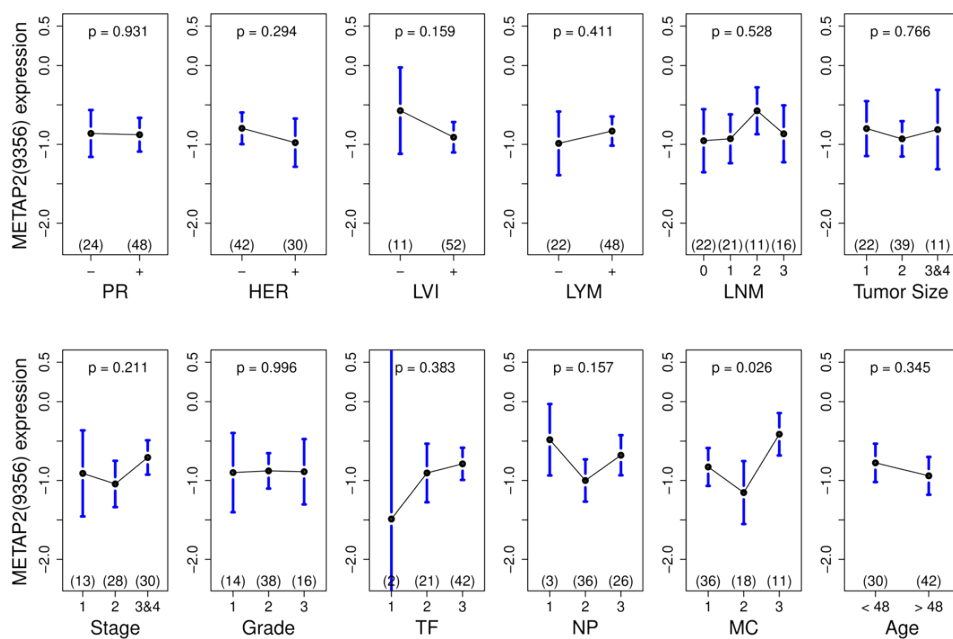


**Figure S7.2.** Mean plot analyses of mRNA levels for *METAP2* in eight clinical categories and in two cohorts of infiltrating ductal carcinoma (IDCs), respectively. Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis(LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90A) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see main text for definitions).

**A. Cohort 1**

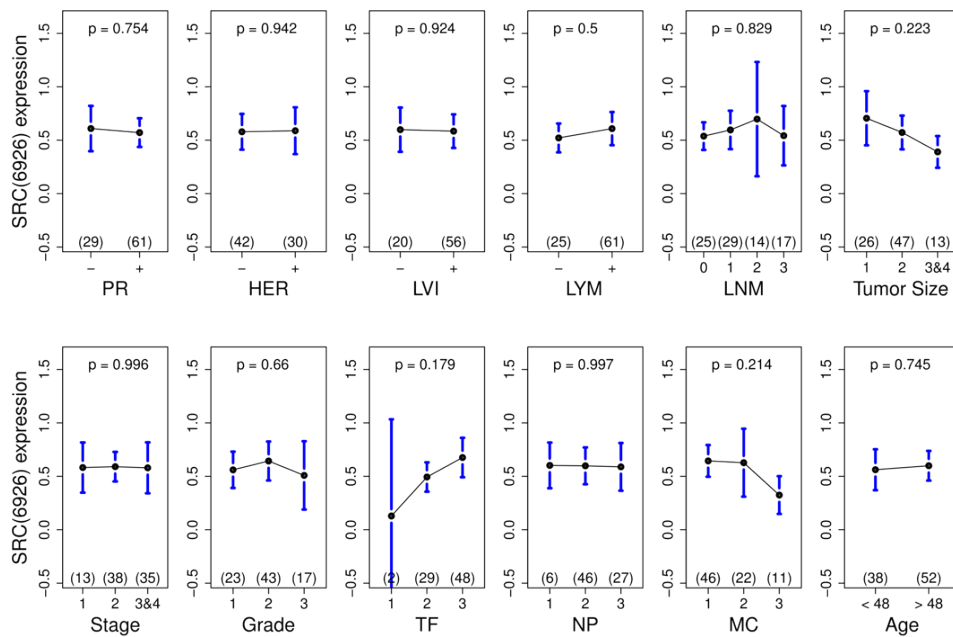


**B. Cohort 2.**

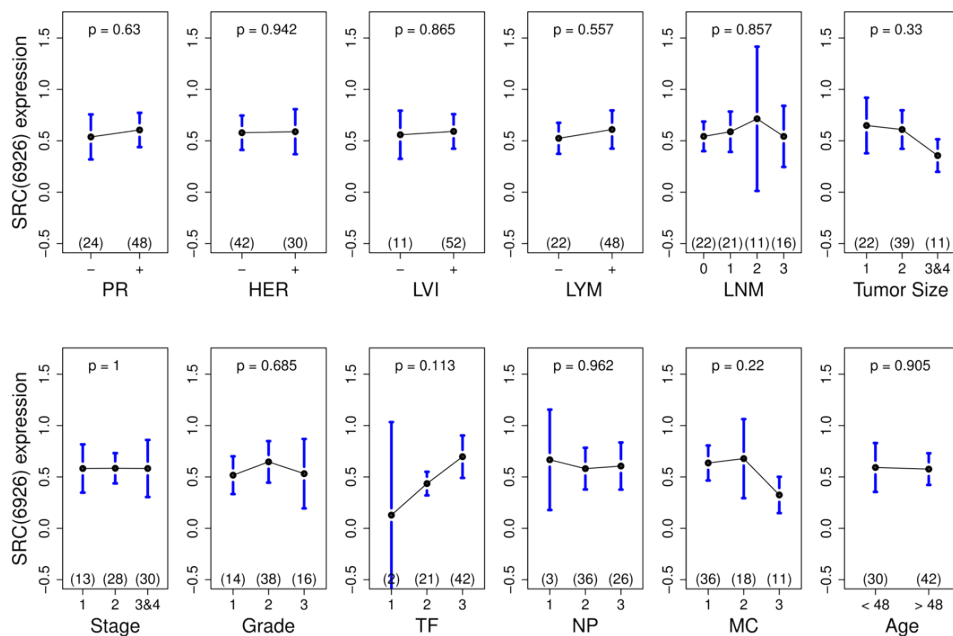


**Figure S7.3.** Mean plot analyses of mRNA levels for *SRC(6926)* in eight clinical categories and in two cohorts of infiltrating ductal carcinoma (IDCs), respectively. Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis(LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90A) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see main text for definitions).

A. Cohort 1.

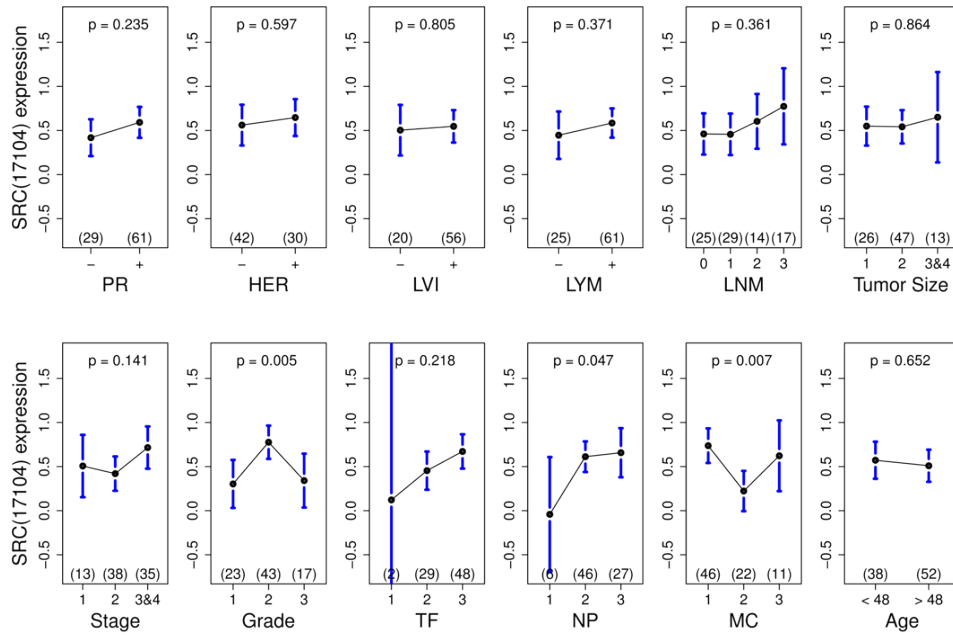


B. Cohort 2.

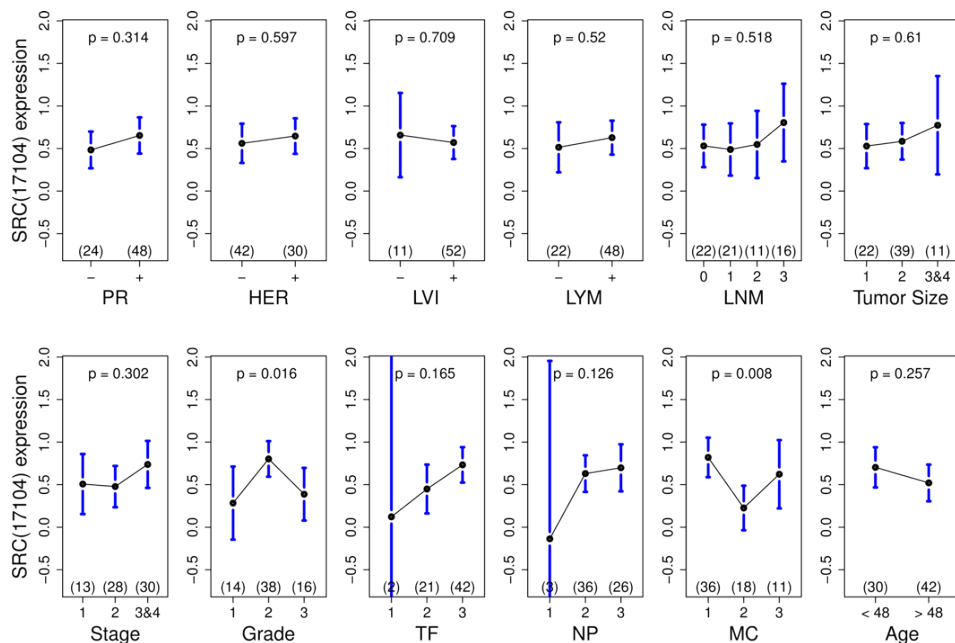


**Figure S7.4.** Mean plot analyses of mRNA levels for *SRC(17104)* in eight clinical categories and in two cohorts of infiltrating ductal carcinoma (IDCs), respectively. Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis(LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90A) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see main text for definitions).

A. Cohort 1.

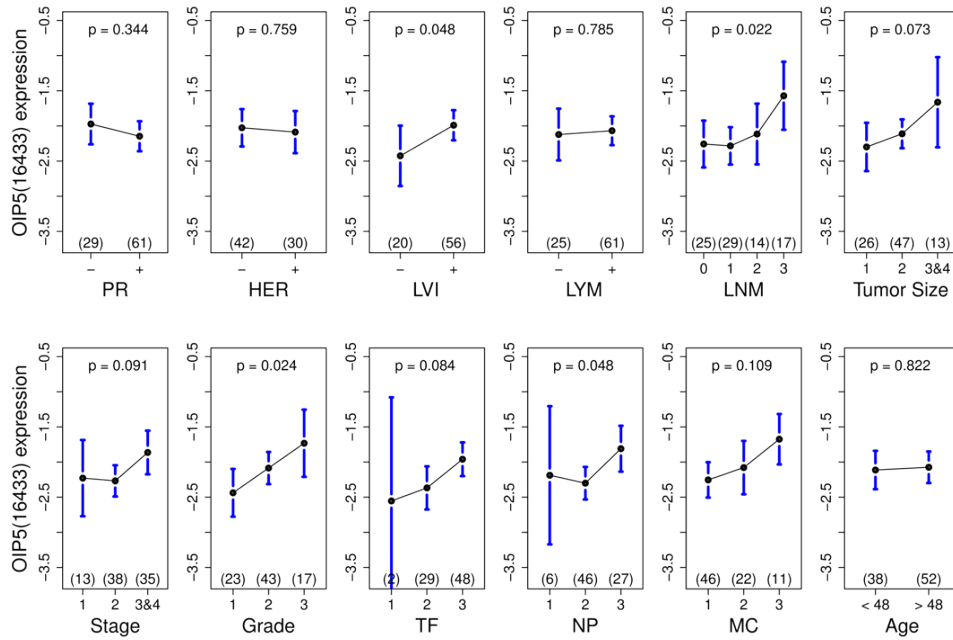


B. Cohort 2.

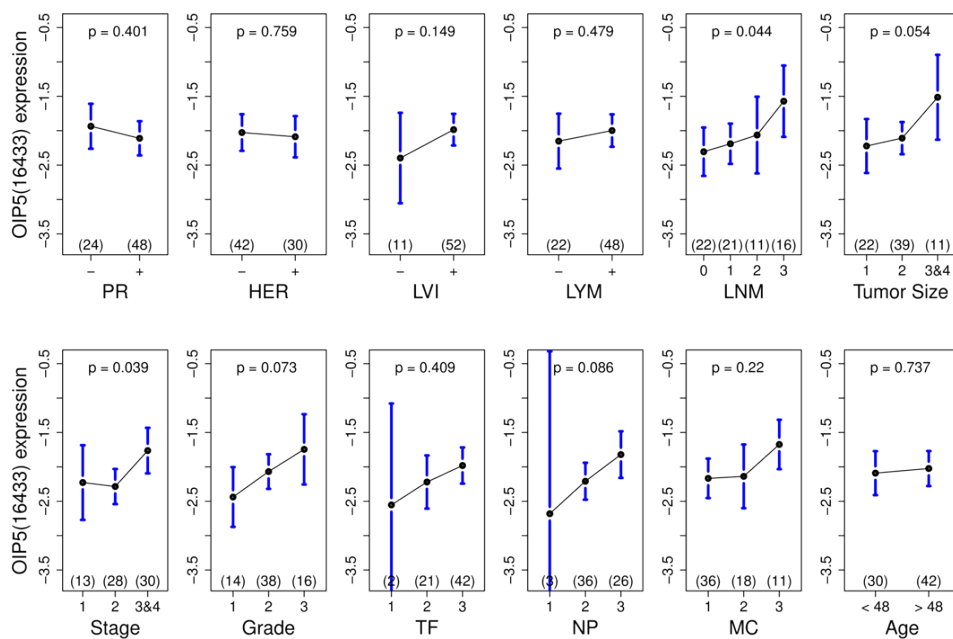


**Figure S7.5.** Mean plot analyses of mRNA levels for *OIP5* in eight clinical categories and in two cohorts of infiltrating ductal carcinoma (IDCs), respectively. Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis(LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90A) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see main text for definitions).

A. Cohort 1.



B. Cohort 2.



**Table S7.1.** ANOVA test on mRNA levels for *ARNT* in eight clinical categories and in two cohorts of infiltrating ductal carcinoma (IDCs), respectively. Lymphovascular invasion (LVI), nodal category (lymph node metastasis (LYM), number of nodal metastasis(LNM)), histological grade (Grade) category (nuclear pleomorphism (NP) and tubule formation (TF)) and stage were analyzed. Cohort 1 (90A) has Groups IE and IIE. Cohort 2 (72A) has Luminal A and Luminal B (see main text for definitions).

clinical index	ARNT(14595)	
	p value	
	90A	72A
PR	0.413	0.305
HER	0.820	0.820
stage	0.574	0.359
LYM	0.318	0.544
LVI	<b>0.010</b>	0.265
Age	0.786	0.584
Grade	0.821	0.131
TF	0.209	0.085
NP	0.130	0.116
MC	<b>0.025</b>	<b>0.008</b>
Size	0.576	0.733
LNM	0.818	0.599