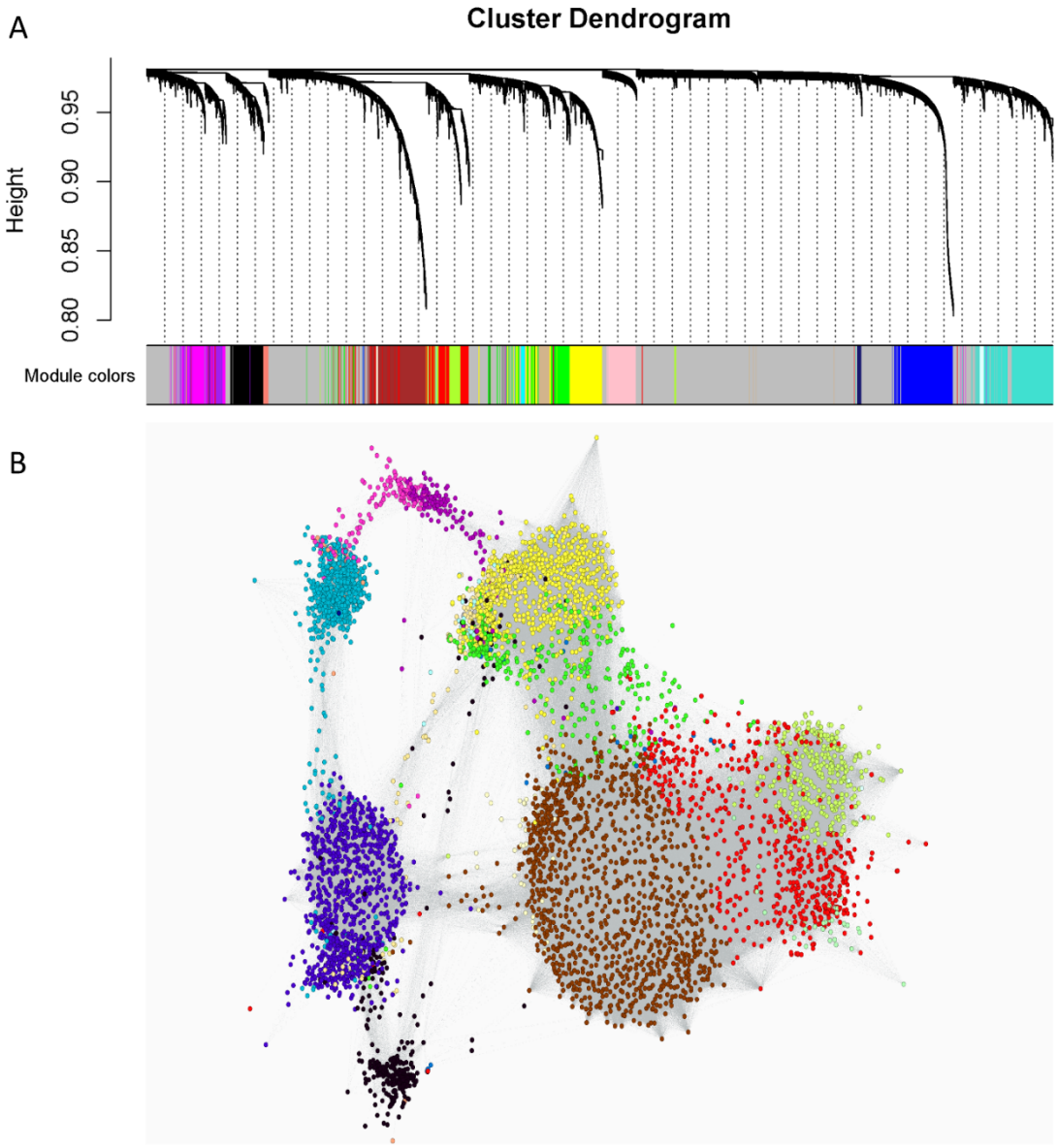
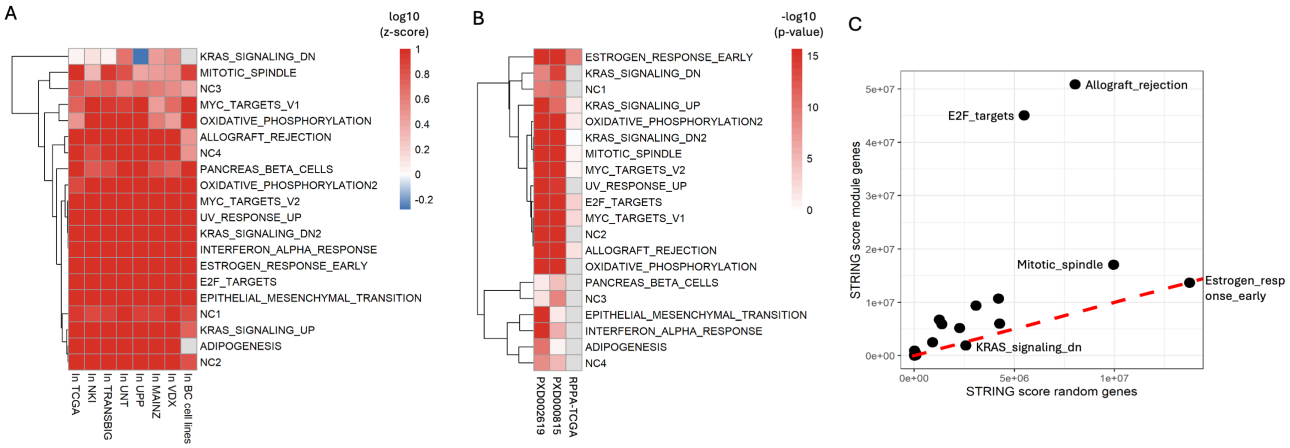
Supplementary Materials

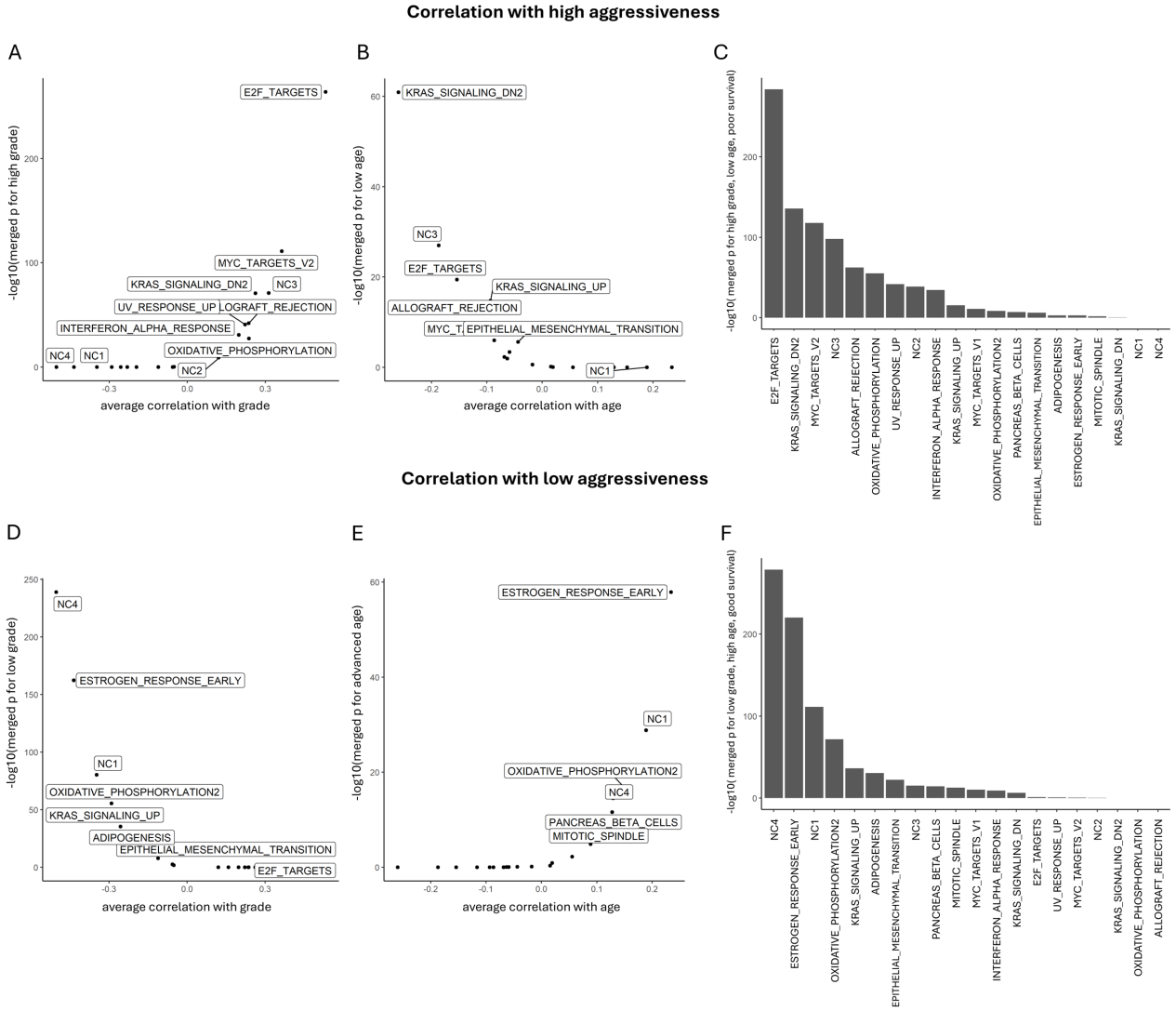


Supplementary Figure 1. METABRIC gene co-expression network. (A) Clustering of genes performed with WGCNA, each colour corresponding to a module. **(B)** Gephi representation of the network, where nodes are genes and edges are connections (only the strongest are represented, Topological Overlap ≥ 0.05. Different colours correspond to different modules.

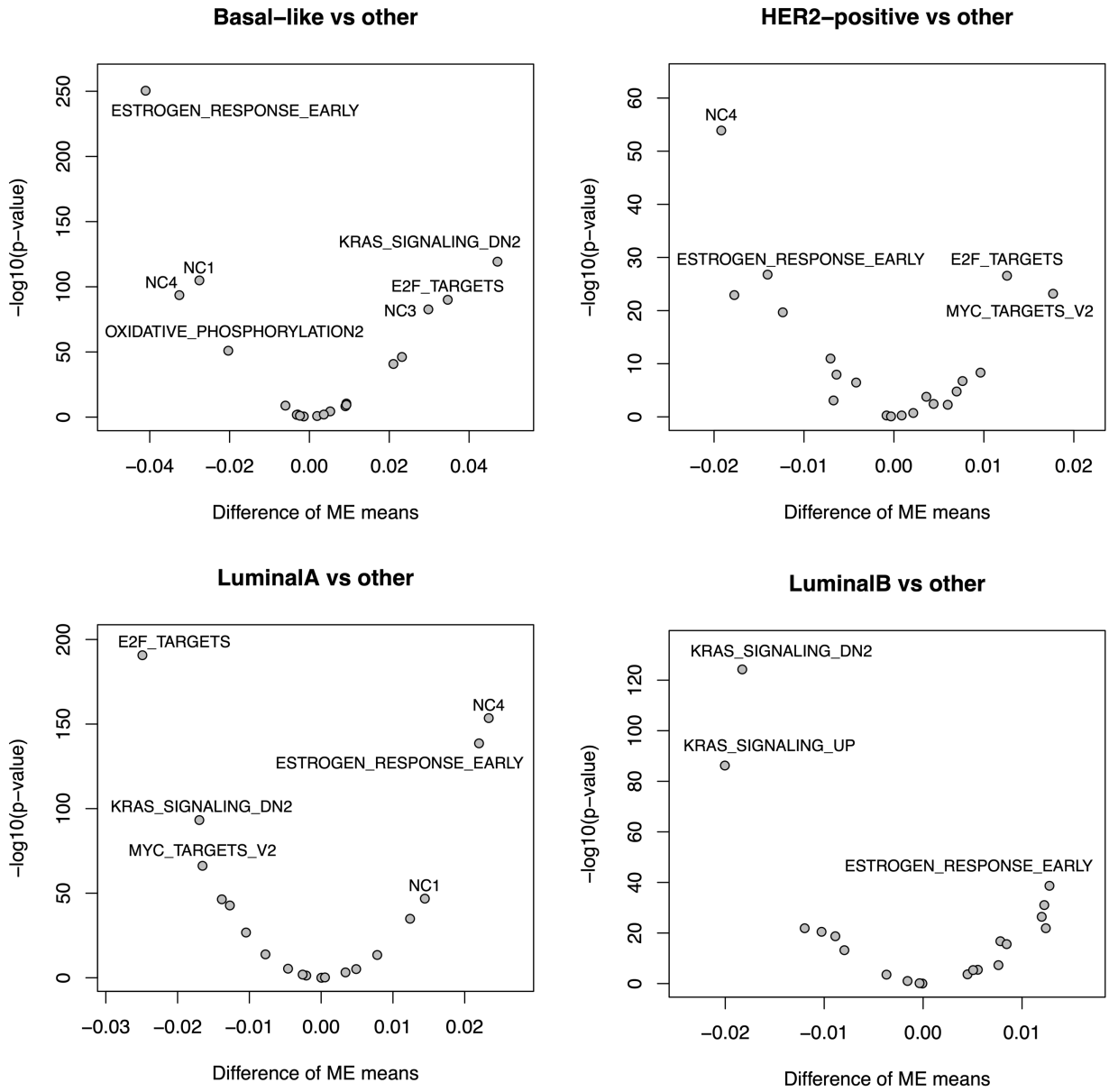


Supplementary Figure 2. Breast cancer co-expression networks are robust. (A) Heatmap summarizing the z-score of modules' preservation (rows) across 7 independent primary BC transcriptome datasets and one BC cell lines dataset (columns). The colour represents the log10 of z-scores, with red corresponding to higher preservation. **(B)** Heatmap showing the connectivity of modules (rows) in proteomic datasets (columns). The colour indicates the –log10 of the p-value for a Kolmogorov-Smirnov test comparing the correlations of modules' genes with those of random genes. **(C)** The proteins encoded by co-expressed genes are more tightly connected than random proteins in the STRING interaction database. Obtained by summing the scores of the STRING database for each pair of interacting genes in one module (y axis) and comparing it with the corresponding measure obtained with an equal number of random genes (x axis). Each dot represents a module. Some relevant modules' names have been included next to the corresponding dot.

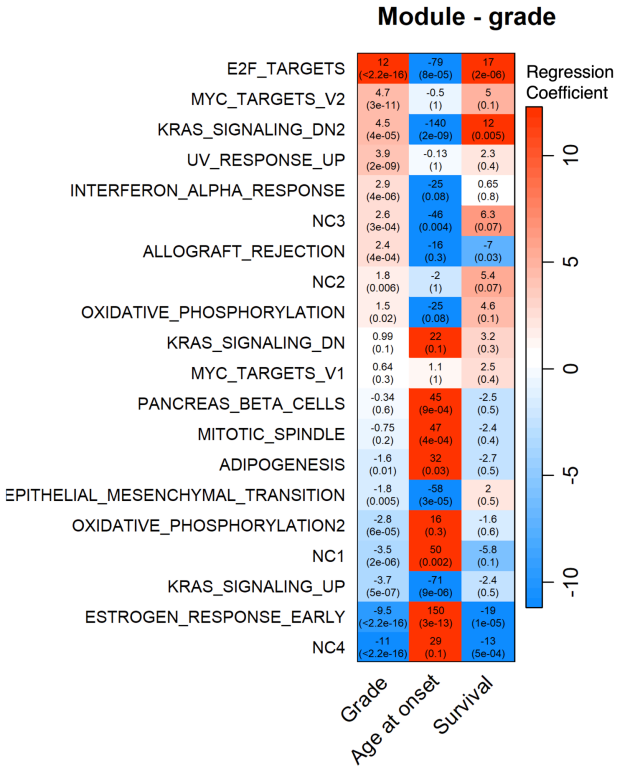
Coexpression networks are clinically informative



Supplementary Fig. 3. Summary of modules' relationship with clinical features. (A-C) Modules' correlation with features of high aggressiveness: high grade **(A)**, early onset **(B)**, merged score for high grade, early onset and poor prognosis **(C)**. **(D-F)** Modules' correlation with features of low aggressiveness: low grade **(D)**, late onset **(E)**, merged score for low grade, late onset and good prognosis **(F)**. **(A,B,D,E)** For each clinical feature, p-values were obtained by merging the p-values across all the individual tested datasets with the Fisher's method, considering the direction of the relationship between module's expression and the clinical feature. **(C, F)** for each module, the overall significance for the relationship with grade, age at diagnosis and survival was obtained by multiplying the individual dataset-merged p-values.

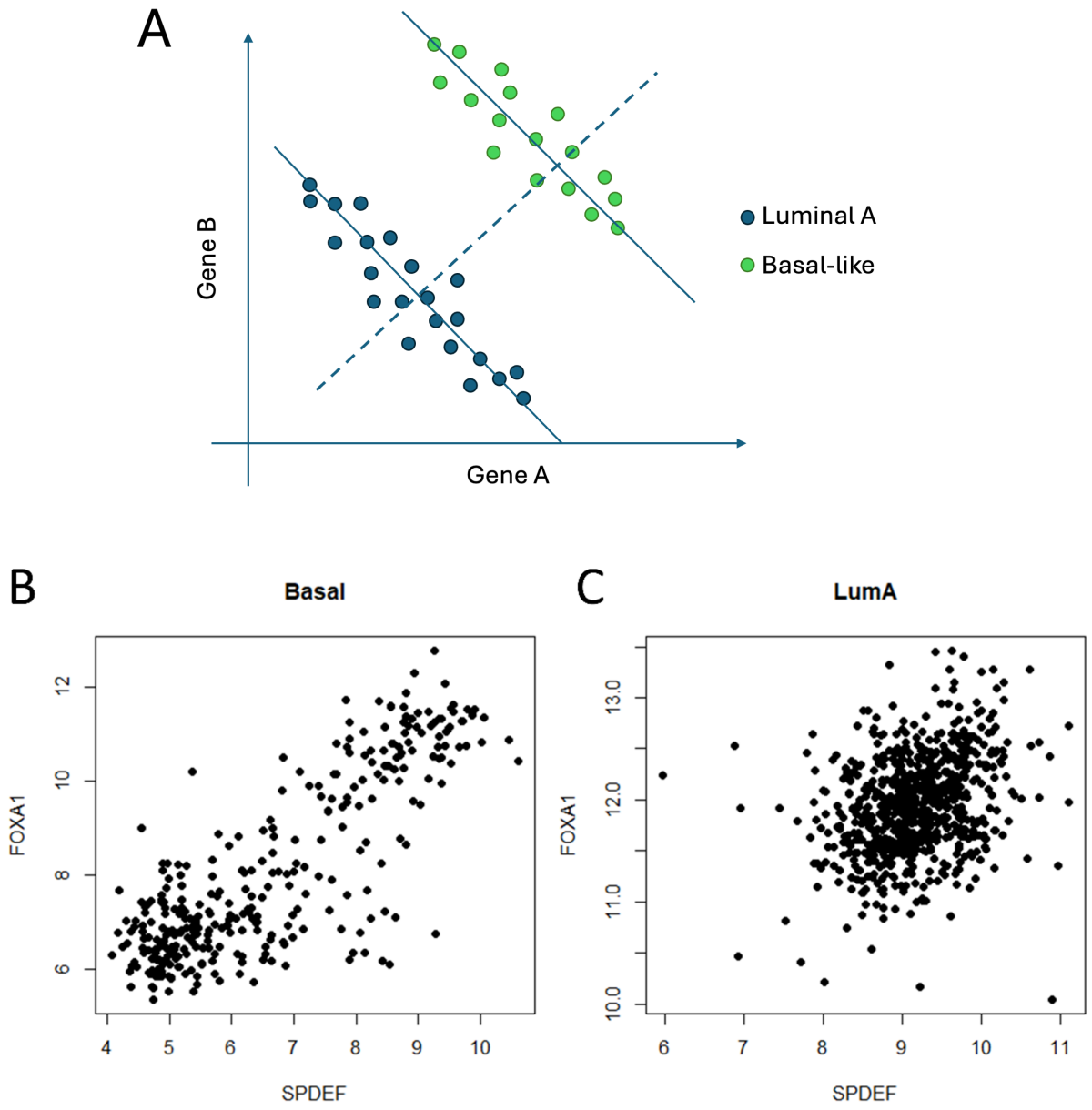


Supplementary Figure 4. Expression of the modules across BC subtypes. The expression of each module in a subtype (specified in the plot title) was compared with the expression of the same module in all other subtypes pooled together. The x-axis shows the difference between the module eigengene (ME) between the two groups, the y-axis represents the significance for the comparison. Note the overexpression of the modules *KRAS_signaling_dn2* and *E2F_targets* in the basal-like tumors, of the *Estrogen_response_early* module in both luminal subtypes, and of the *E2F_targets* in the HER2-positive tumors. p-value $<2.2*10^{-16}$ in all cases.

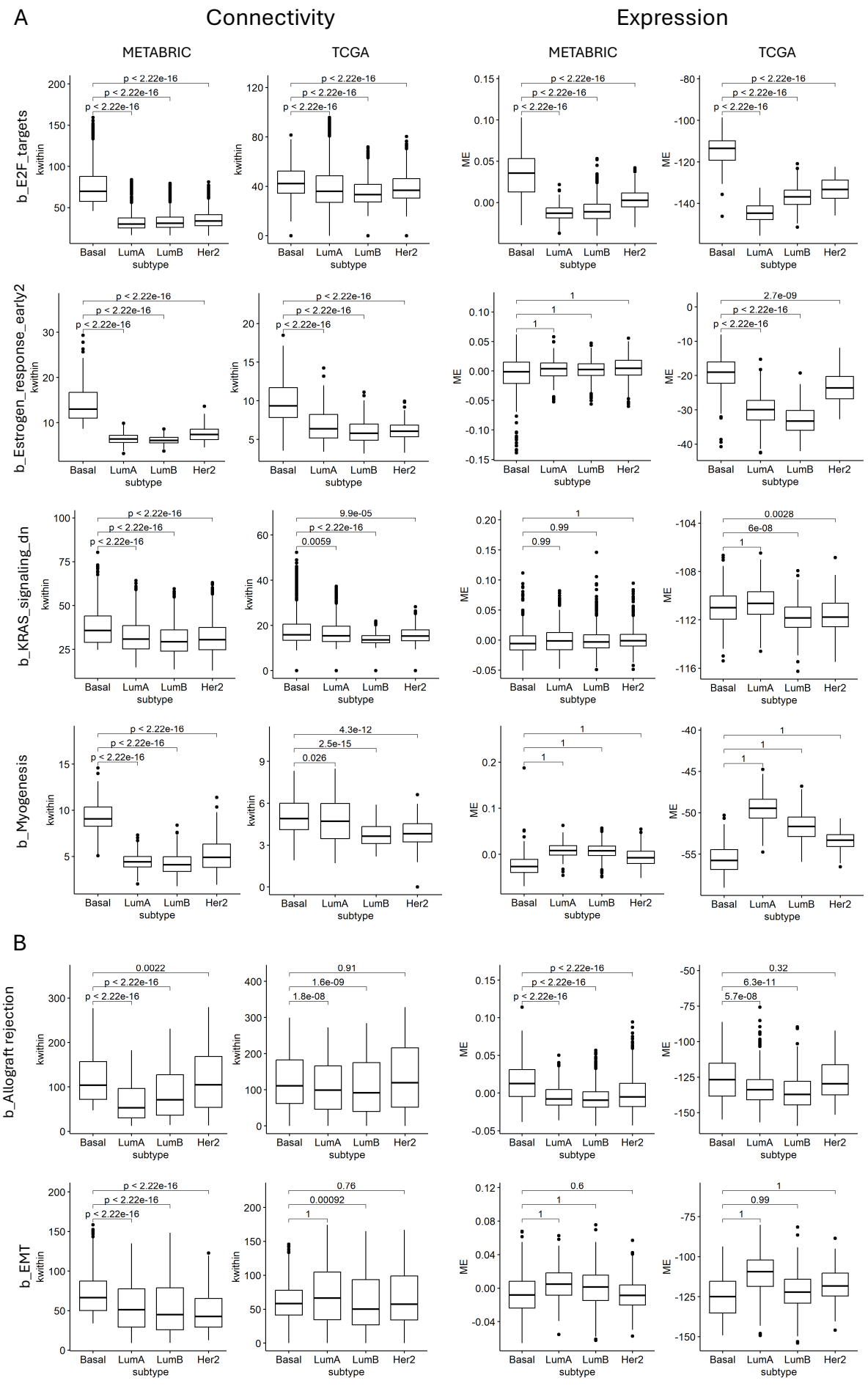


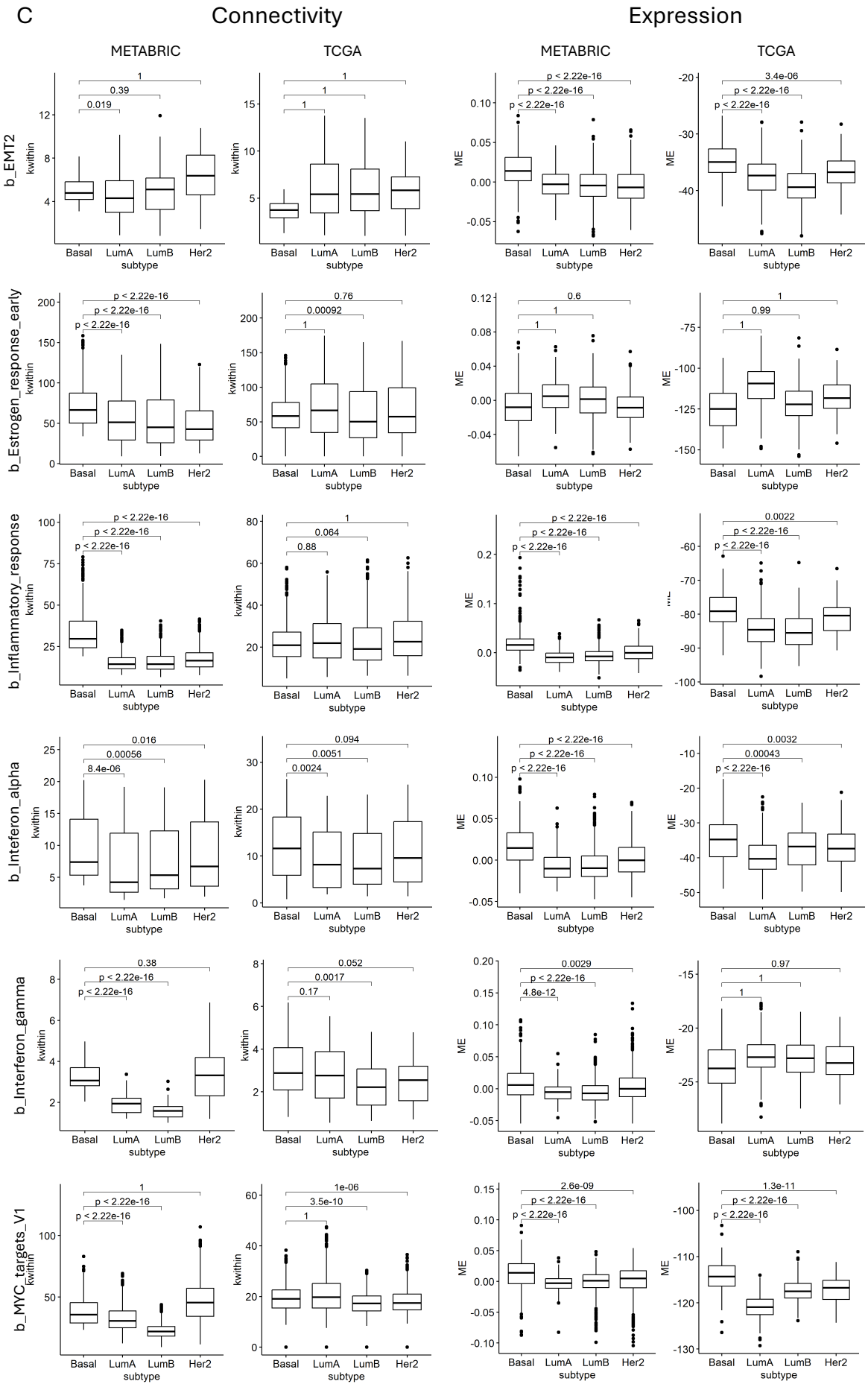
Supplementary Figure 5. Modules correlate with clinical features independently of tumor subtype. The heatmap and the enclosed numbers shows regression coefficients and corresponding p-values for module eigengenes (rows) in models with the indicated clinical features (columns) as dependent variables, and tumor subtype as a categorical covariate.

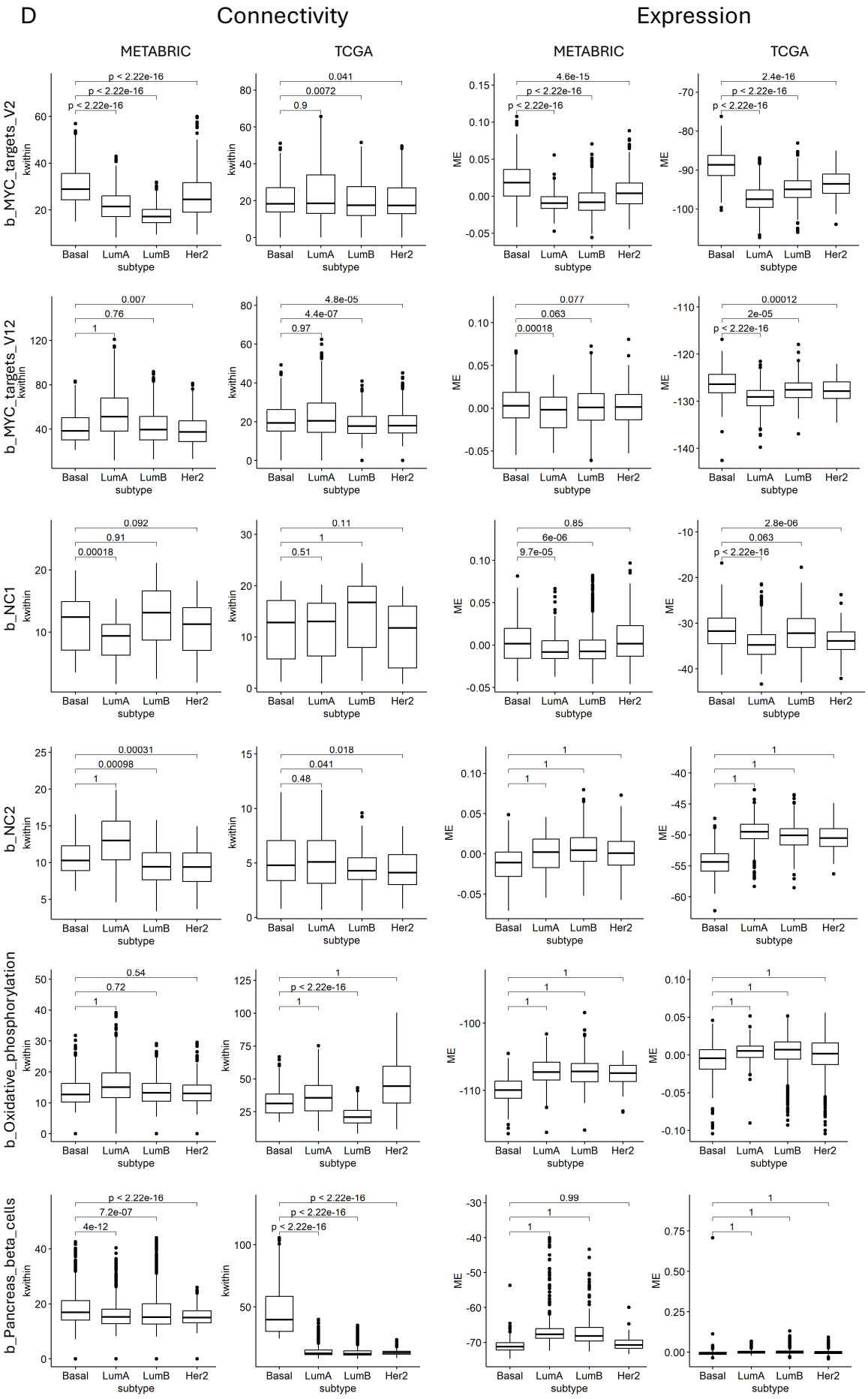
Disentangling subtype effects reveals robust and clinically informative basal-like coexpression modules

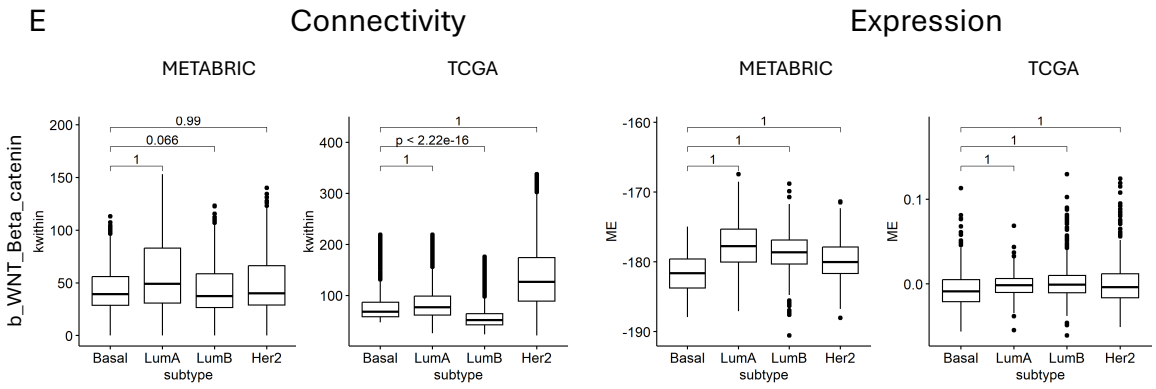


Supplementary Figure 6. Exemplification of the Simpson's Paradox. The Simpson's Paradox is a statistical phenomenon where an association between two variables in a dataset emerges, disappears or reverses when the data are divided into more homogeneous subgroups. **(A)** Two ideal genes, known to display overall positive correlation (not shown), are instead negatively correlated in Basal-like or Luminal A subtypes. **(B, C)** The correlation between FOXA1 and SPDEF is strikingly different between Basal-like tumors **(B)** and Luminal A tumors **(C)** in the METABRIC cohort.

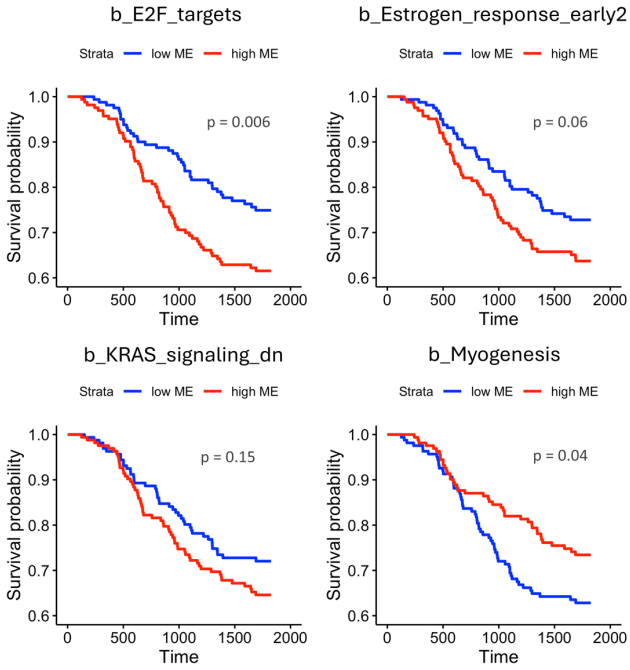

A
Connectivity
Expression
METABRIC
TCGA
METABRIC
TCGA
b_E2F_targets
b_Estrogen_response_early2
b_KRAS_signaling_dn
b_Myogenesis
B
b_Allograft rejection
b_EMT
kwithin
ME
subtype
Basal
LumA
LumB
Her2
p < 2.22e-16
9.9e-05
0.0059
4.3e-12
2.5e-15
0.026
0.0022
0.91
1.6e-09
1.8e-08
0.32
6.3e-11
5.7e-08
0.76
0.00092
2.7e-09
0.0028
6e-08
0.99
0.6


C
Connectivity
Expression
METABRIC
TCGA
METABRIC
TCGA
b_EMT2
b_Estrogen_response_early
b_Inflammatory_response
b_Inteferon_alpha
b_Interferon_gamma
b_MYC_targets_V1
kwithin
ME
subtype
Basal
LumA
LumB
Her2

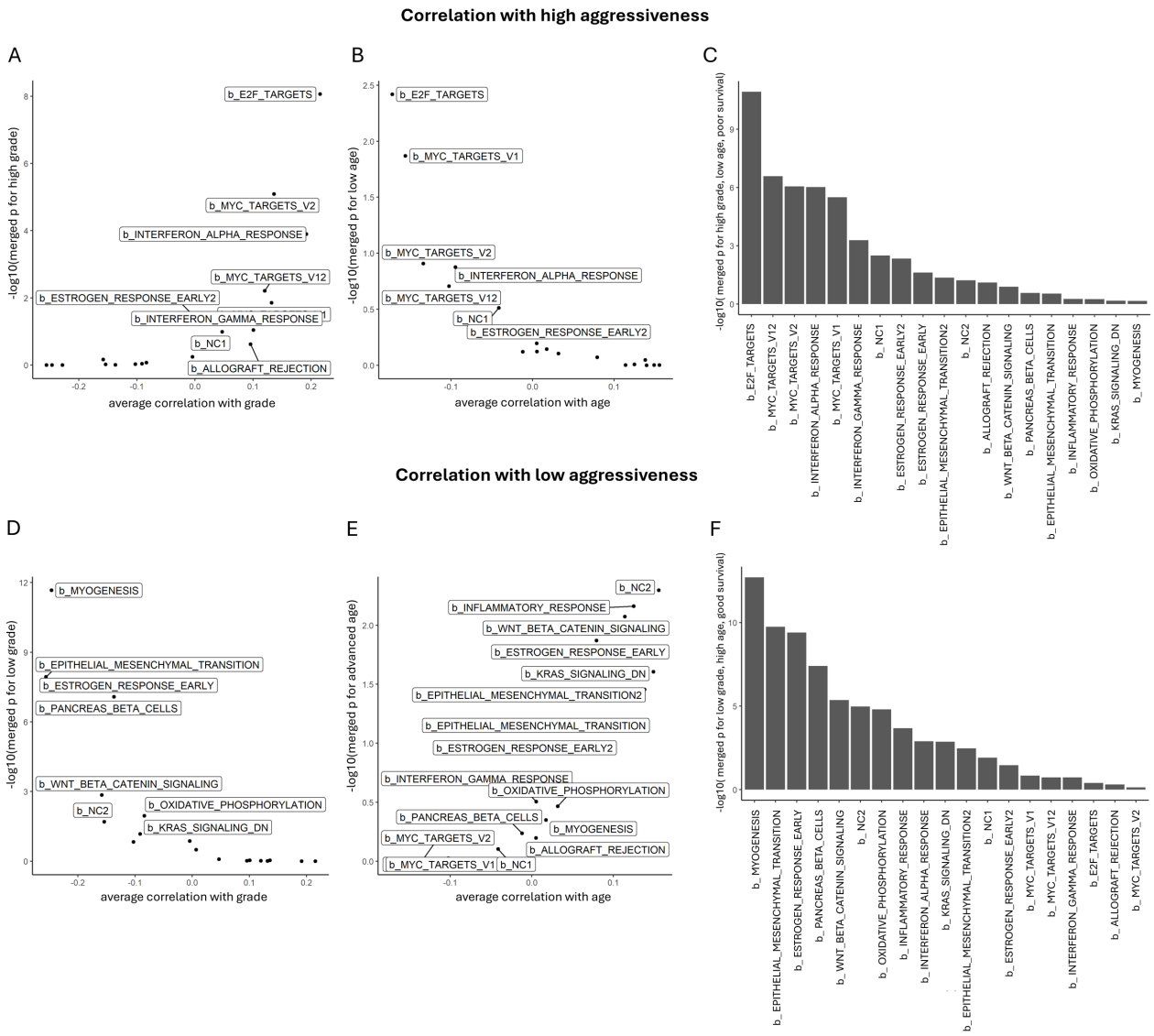

D
Connectivity
Expression
METABRIC
TCGA
METABRIC
TCGA
b_MYC_targets_V2
b_MYC_targets_V12
b_NC1
b_NC2
b_Oxidative_phosphorylation
b_Pancreas_beta_cells
kwithin
ME
subtype
Basal
LumA
LumB
Her2
p < 2.22e-16
0.041
0.0072
0.9
4.6e-15
2.4e-16
0.007
0.76
1
4.8e-05
4.4e-07
0.97
0.077
0.063
0.00018
0.00012
2e-05
0.092
0.91
0.00018
0.11
0.51
0.85
6e-06
9.7e-05
2.8e-06
0.063
0.00031
0.00098
0.018
0.041
0.48
0.54
0.72
0.99
7.2e-07
4e-12



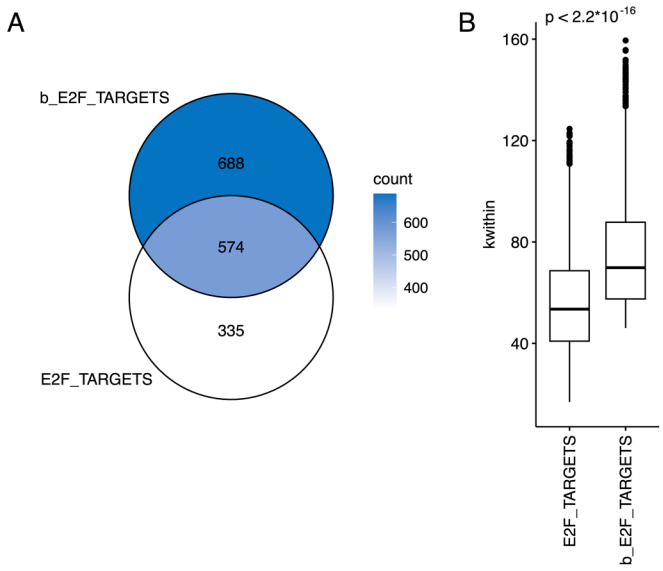
Supplementary Figure 7. Connectivity and expression of basal-like modules across subtypes. For each module (rows), intramodular connectivity (kWithin) and Module Eigengene (ME) were computed in both the METABRIC and TCGA databases across subtypes. P-values were obtained with a Wilcoxon rank-sum test. **(A)** shows the four modules displaying significantly higher connectivity/expression in the basal-like subtype. **(B,C,D,E)** all other modules.



Supplementary Figure 8. Kaplan-Meier survival curves for basal-like patients stratified by module expression. Patients were grouped based on the module eigengene (ME) values of modules that show higher intramodular connectivity in basal-like tumours compared to other subtypes.



Supplementary Figure 9. Summary of basal modules' relationship with clinical features. (A-C) Modules' correlation with features of high aggressiveness: high grade **(A)**, early onset **(B)**, merged score for high grade, early onset and poor prognosis **(C)**. **(D-F)** Modules' correlation with features of low aggressiveness: low grade **(D)**, late onset **(E)**, merged score for low grade, late onset and good prognosis **(F)**. **(A,B,D,E)** For each clinical feature, p-values were obtained by merging the p-values across all the individual tested datasets with the Fisher's method, considering the direction of the relationship between module's expression and the clinical feature. **(C, F)** for each module, the overall significance for the relationship with grade, age at diagnosis and survival was obtained by multiplying the individual dataset-merged p-values. Only basal-like samples were employed in this analysis.

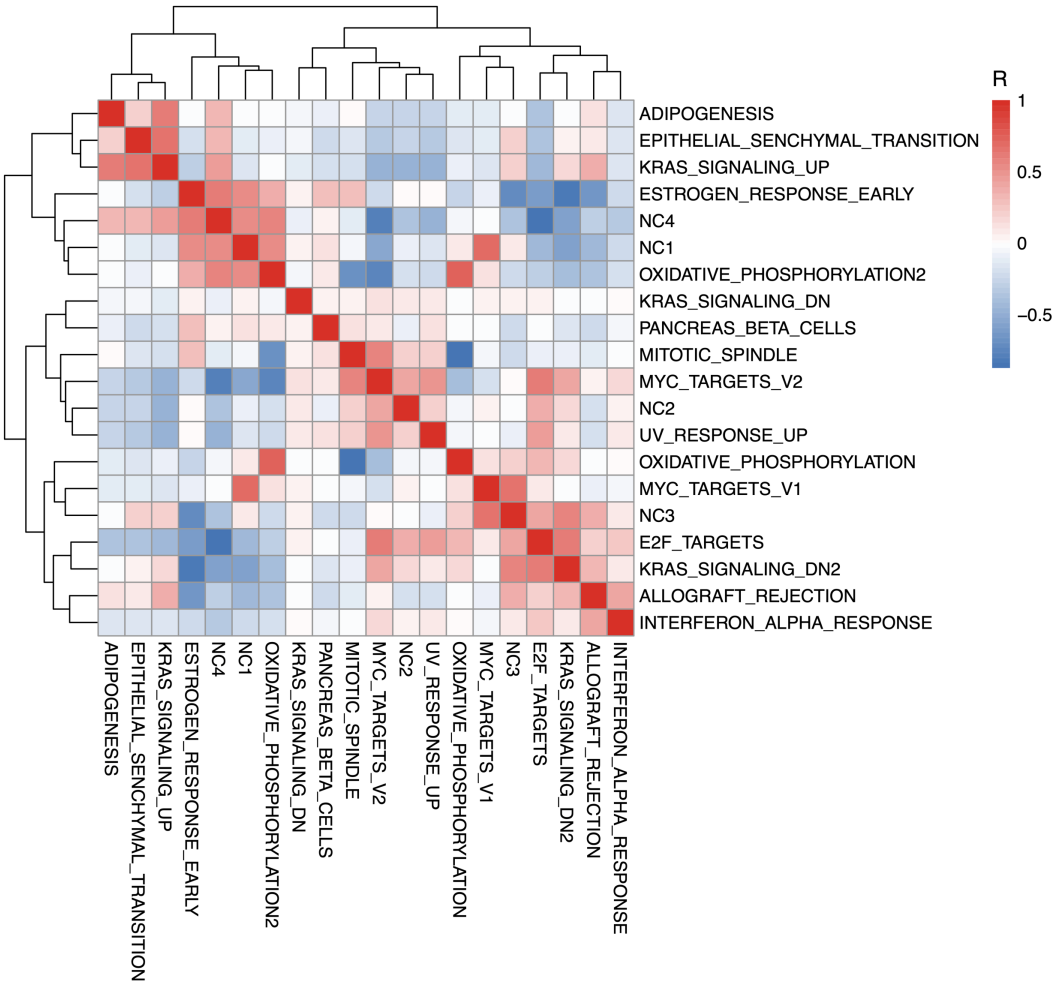


Supplementary Figure 10. Comparison between global and basal-like-specific *E2F_targets* modules. (A) Venn diagram showing the overlap between genes in the *E2F_targets* module (defined using all breast cancer samples) and the *b_E2F_targets* module (defined using only basal-like samples). **(B)** Boxplot comparing the intramodular connectivity (kWithin) of genes in each module, computed in basal-like samples. Statistical significance was assessed using a two-sided Wilcoxon test.

Highly connected b_E2F_targets transcription factors significantly regulate the expression of module's genes.

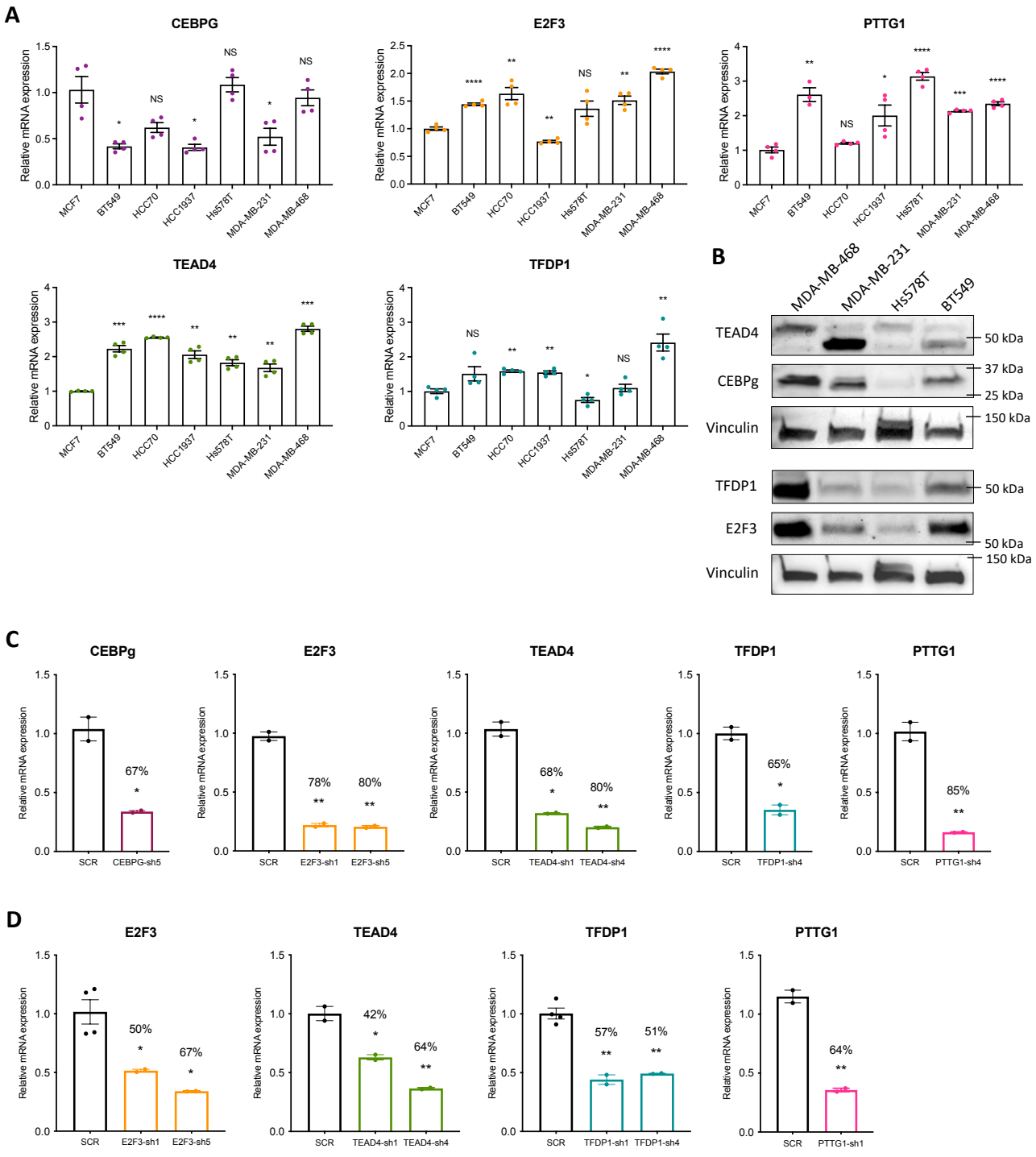
Supplementary Table 1.
b_E2F_targets genes belonging to the GO:0003700 - DNA-binding transcription factor activity - Gene Ontology term*, ranked by kWithin*

TF	kWithin	rank	dataset
FOXM1	151,849192	4	GSE2222, GSE55204, GSE25741
PTTG1	113,661926	97	GSE48928
TEAD4	111,890624	104	
EZH2	111,377135	106	GSE36939, GSE48979, GSE103242
E2F3	105,938864	137	
FOXC1	104,640658	146	GSE73234
PARP1	101,120351	161	GSE34817
TFDP1	96,5986486	200	
CEBPG	96,2457574	202	
CHCHD3	95,9415737	208	
NOLC1	93,6657614	241	
HMGA1	91,4973555	265	GSE45483, GSE35525
UHRF1	90,0384401	286	
TAF5	90,0092093	287	
SOX9	89,7260918	289	
BOLA1	88,419527	305	
TCF7L1	88,3648149	306	
SSRP1	88,2140726	310	GSE92281
ZNF232	87,4847614	320	
ZNF165	86,9895521	326	GSE63984
ELF5	86,6682148	330	GSE30405
RCOR2	84,5795255	357	
NR2C2	82,2037687	390	
ZBTB5	81,7372596	398	
DNMT1	81,3654839	404	



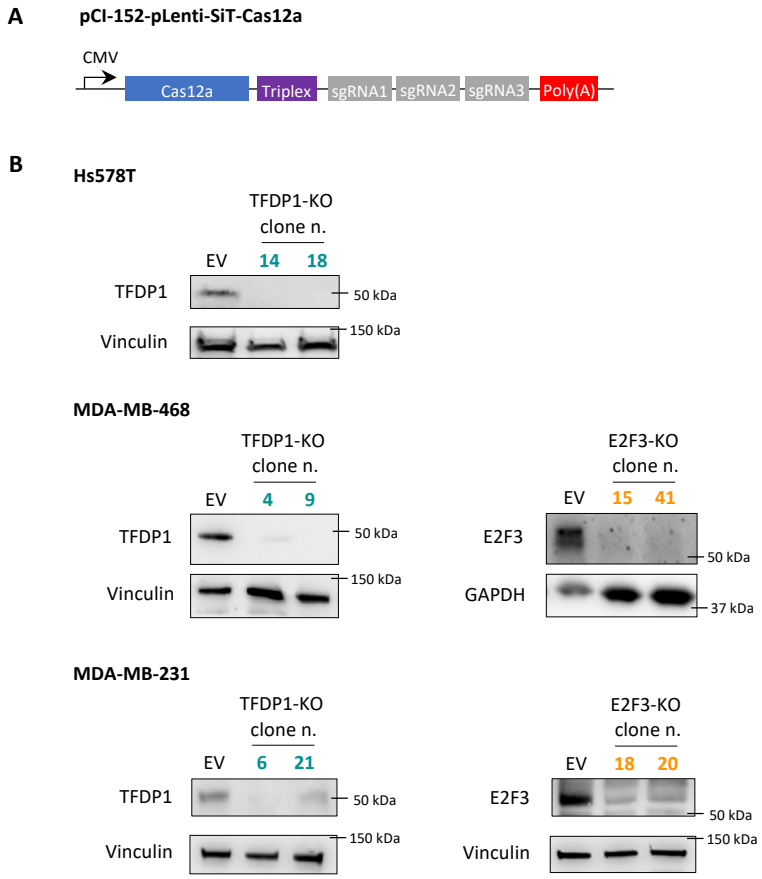
Supplementary Figure 11. Correlations between basal-like modules. Heatmap showing Pearson's correlation between module eigengenes, computed across basal-like patients. Hierarchical clustering was based on their eigengene correlation.

Silencing of five TF hubs significantly affects proliferation of BLBC cell lines.

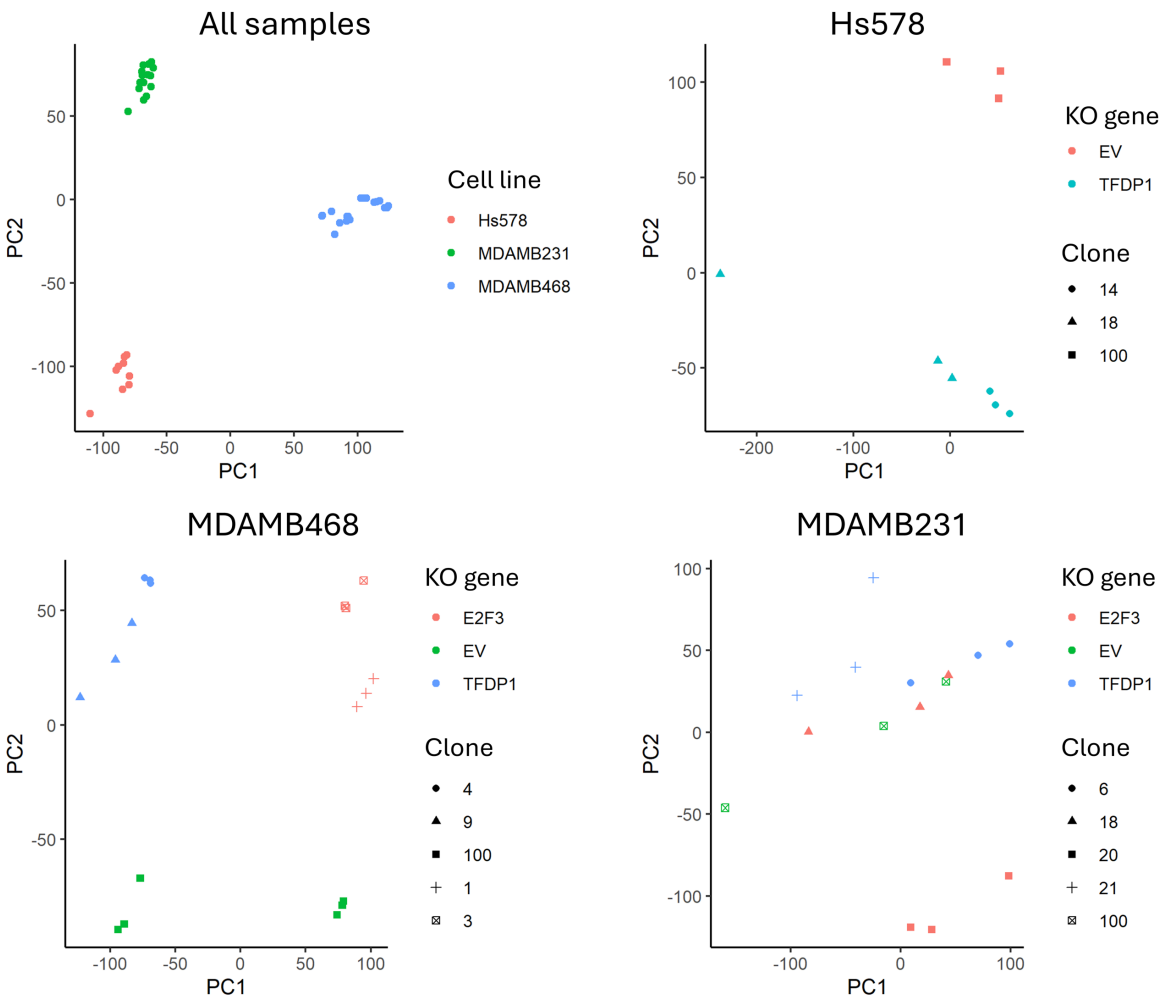


Supplementary Figure 12. Expression and silencing of *b_E2F_targets* TF hubs. The expression of the indicated genes was measured in the indicated cell lines by RT-qPCR **(A)**, shown as mean±S.E.M relative to the value of MCF7 cells upon normalization to Actin B (n=4 independent experiments), one way ANOVA. *, p<0.05; **, p < 0.01, ***, p < 0.001, ****, p < 0.0001, or by Western blot on whole cell extracts (n=2 independent experiments) **(B)**. TF hubs were silenced by lentiviral-mediated shRNA expression in MDA-MB-468 **(C)** and -231 **(D)** cells. Silencing was measured by qRT-PCR and is shown as mean±S.E.M relative to the scrambled control (SCR). Only the best performing shRNAs are shown. (n=4 independent experiments). Unpaired t-test, *, p<0.05; **, p < 0.01.

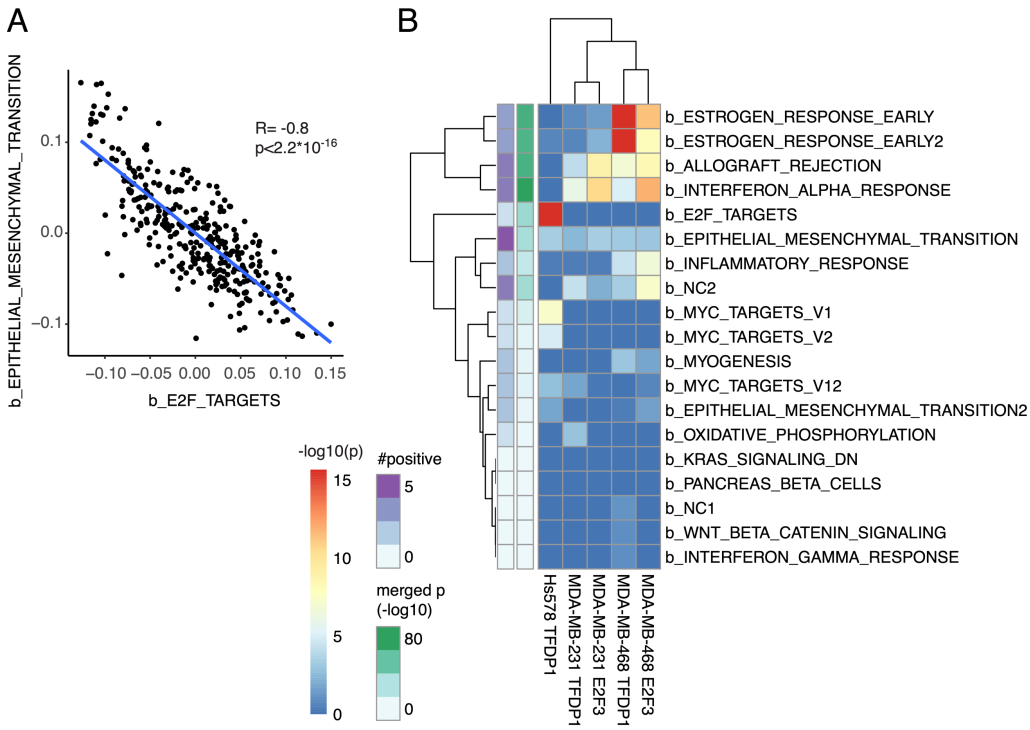
Phenotypic effects of TF hub genes knockout correlate with transcriptional module disruption.



Supplementary Figure 13. Generation of Hs578T, MDA-MB-231 and MDA-MB-468 cells KO for E2F3 and TFDP1. (A) Schematic representation of vectors encoding for Cas12a (AsCfp1) and carrying a CRISPR array with 3 sgRNAs directed against the target genes. **(B)** Residual protein expression was assessed by Western blot analysis of different clones derived from the indicated cell lines, with the indicated antibodies.



Supplementary Figure 14. Principal Component Analysis (PCA) of RNA-seq data. Each point represents a single sample. In the first plot, colors indicate the cell line. In the other plots, samples from each cell line are shown separately, and colors indicate the specific gene knockout or the empty vector control (EV). Clones are indicated by different shapes. Note that MDA-MB-468 E2F3 KO clones and their empty vector were sequenced in a different facility, explaining the PC1 separation from the other samples.



Supplementary Figure 15. (A) Scatter plot of *b_E2F_targets* and *b_Epithelial mesenchimal_transition* modules' eigengenes computed from METABRIC basal-like samples. Each point represents a patient, with the x- and y-axes showing the module eigengene of the *b_E2F_targets* and *b_Epithelial_Mesenchymal_Transition* modules, respectively. **(B)** Enrichment significance of up-regulated genes upon TF hubs' KO in basal-like BC lines, in-house data.

Suppl. Table 2. Comparison of all the TF hubs in the E2F_targets and b_E2F_targets modules.

The top 25% TF hubs in each module are shown in bold. Asterisks indicate genes completely absent from the general module.

Note that 7 of the validated genes (FOXC1, PARP1, SSRP, ELF5, TEAD4, E2F3, TFDP1) would have been missed without analysing subtype-specific networks.

E2F_targets

Gene	kTotal	kWithin	rank
FOXM1	545,8059478	157,2385805	13
PTTG1	518,1612739	146,7355187	27
E2F2	518,25252	131,6658127	53
EZH2	469,5950475	125,0671861	64
UHRF1	447,6648191	117,7752092	82
HMGA1	471,904967	93,27173723	165
E2F7	417,0809894	88,19651424	195
DNMT1	399,0062414	75,66846466	294
NOLC1	398,2846515	75,06320196	301
CHCHD3	401,8862114	71,3488528	337
RCOR2	421,053976	70,59843607	343
CEBPG	427,9626136	66,98986737	385
HNRNPAB	414,5493711	66,32517897	391
TEAD4	414,3963194	66,04319689	396
SSRP1	441,1457563	65,37224264	405
BOLA3	409,4217696	64,1340471	426
GABPB1	389,0360498	63,83230828	433
E2F5	388,8092443	62,5480216	456
TFDP1	394,1785418	62,42070088	457
NR2C2	395,1968325	60,94682183	476
ZBED4	391,1925901	59,95355777	495
WDHD1	367,8458545	59,83978917	497
TAF2	377,9299012	59,55837054	501
E2F8	357,7575194	58,91275579	512
TAF5	390,7865121	58,45965623	523
OTX1	392,2978003	57,78689338	541
PHF5A	384,158976	57,16442653	552
TGIF2	390,9910685	55,71181488	582
E2F4	448,1823073	54,14917889	606
SOX11	414,7745595	54,01127145	610
TCF19	364,8410881	53,96747972	613
NKRF	415,7650847	53,14778944	631
C9orf40	376,7558987	51,13765336	676
ZSCAN20	401,6309613	49,88284887	701
ZNF280C	373,1364283	48,37535118	723
RC3H2	368,9169567	47,44852961	736
BUD31	385,9782168	44,79921922	770
NONO	349,6588008	44,22706657	776
TAF4	368,5962118	43,44195666	788
NME2	345,396577	42,40858874	801
CARHSP1	391,870682	40,96340859	821
ZNF207	353,4693083	39,36430153	843
NFXL1	352,4212957	38,72882846	855
SMARCC1	358,1545195	38,01927747	867
HLTF	360,1994037	37,5891988	877
ONECUT2	346,4645296	37,102465	886
TFDP2	346,2991594	36,29408547	903

b_E2F_targets

Gene	kTotal	kWithin	rank
FOXM1	517,5179103	151,8491924	4
PTTG1	446,7140888	113,6619256	97
TEAD4	464,990407	111,8906237	104
EZH2	432,6551732	111,3771352	106
E2F3*	473,9321194	105,9388642	137
FOXC1*	440,7405836	104,6406577	146
PARP1*	479,4244058	101,1203506	161
TFDP1	412,1732918	96,59864862	200
CEBPG	448,8300016	96,24575739	202
CHCHD3	426,7689524	95,94157373	208
NOLC1	422,1877497	93,6657614	241
HMGA1	433,6303252	91,49735551	265
UHRF1	404,7782849	90,03844009	286
TAF5	416,5724709	90,00920929	287
SOX9	421,4073231	89,72609181	289
BOLA1	422,449619	88,41952699	305
TCF7L1	414,9359793	88,36481492	306
SSRP1	460,6987668	88,21407263	310
ZNF232	436,6607204	87,48476136	320
ZNF165	427,9250669	86,9895521	326
ELF5*	393,182212	86,66821477	330
RCOR2	429,3935035	84,57952549	357
NR2C2	399,5668967	82,20376874	390
ZBTB5	391,6089104	81,73725958	398
DNMT1	394,6842731	81,36548393	404
E2F7	400,8491659	80,10593558	422
ZSCAN20	427,7471509	79,86075991	425
PHF5A	395,2935163	79,29558779	433
YEATS2	406,0701829	78,21482604	461
SOX10	389,7648041	77,9770118	467
ZNF239	403,9111869	77,78196325	469
FUBP1	396,8369653	77,60745143	473
SMARCC1	397,8883862	77,5632953	474
ZNF384	405,9870069	76,93547543	487
WDHD1	385,5216104	76,75213613	494
ZNF670	382,3450882	75,2118989	519
NFYA	382,4027691	75,06933825	520
TGIF2	417,1442289	74,44344408	532
TFCP2L1	394,6230583	74,05010485	541
YBX3	384,5219229	72,35824483	576
VGLL1	387,5594175	72,34639726	578
MAZ	401,513753	72,28084519	580
EN1	399,3617986	71,75996199	588
YBX1	382,3245053	71,23871339	594
TMSB15B	380,9568685	70,92572076	606
OTX1	393,3594442	70,89675465	607
SOX4	401,6418641	69,51332771	644
ZBED4	376,3568408	69,22634404	653
GATAD2B	409,5072559	69,21675457	654
FOXP4	402,3270348	68,90484656	662
NFE2L3	399,5758507	67,69332685	683
TFDP2	382,6879624	67,29559159	697
ZFP82	400,0982005	67,28340886	698
NFXL1	373,4736797	66,88047719	703
TBX19	389,1013931	66,78783836	705
C9orf40	366,3663232	66,11517465	732
AEBP2	380,2555544	65,89111461	740
CEBPB	403,5518432	65,73412403	742
HLTF	377,9463084	65,17978805	755
ZNF286A	378,4517372	64,75248909	766
E2F8	354,9233308	64,68845453	768
ADNP2	373,5173621	64,42152296	773
NONO	387,8471739	63,97066484	781
HCFC1	371,2887376	63,88609397	785
ETV3	368,2194684	62,83174142	811
MLLT10	363,2272514	62,70973976	817
BUD31	411,0265793	61,54055655	850
NKX2-5	361,3111176	61,24424751	855
TCF19	381,1940255	61,19142204	857
NFIB	378,5458887	61,07810058	859
ZNF280C	364,8530416	60,89722564	863
ETV4	380,0937957	59,76683398	895
PLAGL2	349,1219779	59,46128338	901
LMO4	405,0865999	59,06760596	913
TEAD2	391,1747808	58,70543135	920
HR	373,5257837	58,55145997	922
BARX2	355,5920909	57,94898961	942
TAF4	377,8939906	57,33074599	953
STAG1	366,1406773	57,29681045	954
MYNN	367,215472	57,21276879	958
ALX1	362,9987381	56,62261547	977
SCML2	376,7797932	56,48103523	980
SRF	380,4723943	56,23521831	987
ENO1	422,2986428	56,12750106	990
ZNF576	370,1216093	53,1790406	1060
HNRNPAB	365,8886302	52,9682541	1065
ZNF121	353,8446039	52,74373801	1071
NOCT	392,767661	52,34512882	1080
BMPR1A	353,8991549	52,10447718	1086
POU4F1	352,3619428	51,23862725	1118
LRRFIP2	374,0518174	50,95025584	1124
DMRT1	356,3936143	50,8452056	1127
FOXK2	353,0962496	50,39816632	1142
ZNF519	363,9779497	50,31956502	1147
RLF	359,5718594	50,31559673	1148
KIAA1549	369,3737717	50,07255753	1158
ZNF280B	362,9154775	49,7708106	1169
PPARA	357,3288836	49,46446769	1177
ZNF285	366,2030602	49,44221832	1178
ZSCAN16	371,1125928	48,99338275	1195
TET1	361,6461064	48,38719262	1216
ZNF200	351,8604136	47,61080732	1240

Supplementary Material and Methods

Supplementary Table 3. shRNA, sgRNA, qRT-PCR assay sequences and IDs.

shRNA	clone ID	sequence
CEBPG-sh1	TRCN0000003836	CCGGTGGCGACAATGCAGGACAGTACTCGAGTACTGTCCTGCATTGTCGCCATTTTT
CEBPG-sh2	TRCN0000003837	CCGGGCAACGCCGAGAGAGGAACAACTCGAGTTGTTCCTCTCTCGGCGTTGCTTTTT
CEBPG-sh3	TRCN0000003838	CCGGCCTCACCCTTTCCAGACTTTACTCGAGTAAAGTCTGGAAAGGGTGAGGTTTTT
CEBPG-sh4	TRCN0000003839	CCGGCTGACCAAGGAATTAAGTGTACTCGAGTACACTTAATTCCTTGGTCAGTTTTT
CEBPG-sh5	TRCN0000010821	CCGGGATTTGTTTCTTGAGCATGCACTCGAGTGCATGCTCAAGAAACAAATCTTTTT
E2F3-sh1	TRCN0000013803	CCGGCCCGCTTTACTCTTCAGGAATCTCGAGATTCCTGAAGAGTAAAGCGGGTTTTT
E2F3-sh2	TRCN0000013804	CCGGCCTCATTAAGAAGAAGTCTAACTCGAGTTAGACTTCTTCTTAATGAGGTTTTT
E2F3-sh3	TRCN0000013805	CCGGCCAAACTGTTATAGTTGTGAACTCGAGTTCACAACTATAACAGTTTGGTTTTT
E2F3-sh4	TRCN0000013806	CCGGCCTGACTCAATAGAGAGCCTACTCGAGTAGGCTCTCTATTGAGTCAGGTTTTT
E2F3-sh5	TRCN0000013807	CCGGCCAACTCAGGACATAGCGATTCTCGAGAATCGCTATGTCCTGAGTTGGTTTTT
TEAD4-sh1	TRCN0000015873	CCGGGAAGAGACGTGTGTGCAGGAACTCGAGTTCCTGCACACACGTCTCTTCTTTTT
TEAD4-sh2	TRCN0000015874	CCGGCCCGGATATTGAGCAGAGTTTCTCGAGAAACTCTGCTCAATATCCGGGTTTTT
TEAD4-sh3	TRCN0000015875	CCGGGAGACAGAGTATGCTCGCTATCTCGAGATAGCGAGCATACTCTGTCTCTTTTT
TEAD4-sh4	TRCN0000015876	CCGGCCTTTCTCTCAGCAAACCTATCTCGAGATAGGTTTGCTGAGAGAAAGGTTTTT
TEAD4-sh5	TRCN0000015877	CCGGGCTGTGCATTGCCTATGTCTTCTCGAGAAGACATAGGCAATGCACAGCTTTTT
PTTG1_sh1	TRCN0000015103	CCGGCCTTCAATCAAAGCCTTAGATCTCGAGATCTAAGGCTTTGATTGAAGGTTTTT
PTTG1_sh2	TRCN0000015104	CCGGCCTTACCTAAAGCTACTAGAACTCGAGTTCTAGTAGCTTTAGGTAAGGTTTTT
PTTG1_sh3	TRCN0000015105	CCGGCCAATCTGTTGCAGTCTCCTTCTCGAGAAGGAGACTGCAACAGATTGGTTTTT
PTTG1_sh4	TRCN0000015106	CCGGCCCTTCAATCCTCTAGACTTTCTCGAGAAAGTCTAGAGGATTGAAGGGTTTTT
PTTG1_sh5	TRCN0000015107	CCGGTGTGGTTGCTAAGGATGGGCTCTCGAGAGCCCATCCTTAGCAACCACATTTTT
TFDP1-sh1	TRCN0000019880	CCGGCCTACGGCATTTCTCCATGAACTCGAGTTCATGGAGAAATGCCGTAGGTTTTT
TFDP1-sh2	TRCN0000019881	CCGGGTCTTCATAGACCAGAACCTTCTCGAGAAGGTTCTGGTCTATGAAGACTTTTT
TFDP1-sh3	TRCN0000019882	CCGGCGACAACCACATCTTACCAAACTCGAGTTTGGTAAGATGTGGTTGTCGTTTTT
TFDP1-sh4	TRCN0000019883	CCGGGACGATGACTTCAACGAGAATCTCGAGATTCTCGTTGAAGTCATCGTCTTTTT
TFDP1-sh5	TRCN0000019879	CCGGCCGCTTATTTGCCGTGAGTTTCTCGAGAAACTCACGGCAAATAAGCGGTTTTT
SCR		CCGGCCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGGTTTTT

sgRNA		sequence
hsE2F3		AGATGAATATCTTGATATGTAACATAAAATTTCTACTCTTGTAGATTGATATCACCAACGT TCTGGAAGAATTTCTACTCTTGTAGATGACTATCTGGACTTCGTAGTGCA
hsTFDP1		AGATAAGGAAGCGCAACAGGAAAGGAGAATTTCTACTCTTGTAGATAAGGTGGAAAG ACAGAGGAGACTAATTTCTACTCTTGTAGATGCAAATTGCCTTCAAGAACCTGG

gene	cat. n.	qRT-PCR assay ID
CEBPg	4331182	Hs00156454_m1
E2F3	4331182	Hs00605457_m1
FoxM1	4453320	Hs01073586_m1
PTTG1	4331182	Hs00851754_u1
TEAD4	4331182	Hs01125032_m1
TFDP1	4331182	Hs00955491_gH
TFDP1	4331182	Hs00955488_g1
GAPDH	4331182	Hs99999905_m1
B-act	4448484	Hs01060665_g1