Supplementary methods

aCGH data processing. Raw log-ratios were first processed using *DNAcopy* R package which implements the circular binary segmentation (CBS) algorithm [1] to obtain regions of equal copy number. The high sensitivity of the algorithm (default parameters) gave rise to many segments containing less than five probes which mostly represented outliers or known Copy Number Variant (CNV) regions. To remove the excessive breakpoints then generated, we merged all segments containing five or less probes with the longest adjacent segment within the same chromosome arm.

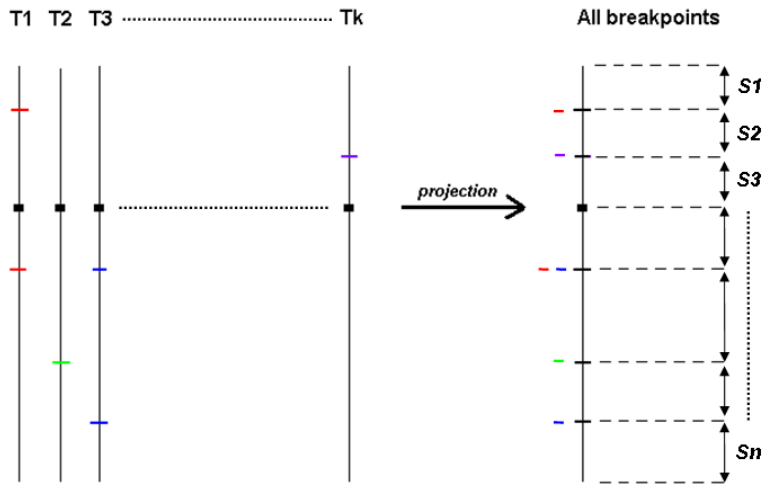
Calling gains, losses and amplifications: To call gains and losses, we applied a tumor-dependent threshold similar to that proposed by Chin et al [2]. For each tumor, we computed the standard deviation, *sigma*, of the middle 50% of the log-ratios (only considering autosomal chromosomes). Then, for status assignment, a segment of log-ratios mean *m* was declared:

- *amplified* if m>=7*sigma
- *gained* if m>=1.75*sigma
- *lost* if m<=1.75*sigma
- or *no change* status otherwise.

Management of CNVs: We removed known CNV regions from the analysis. To that purpose, we made use of the regions published by McCarroll et al [3], which are stored in the Database of Genomic Variants [4]. Based on hg17 annotations, for each CNV region identified by McCarroll et al we determined which probes of the Hu-244A

microarray fell into this region. Next, each segment represented by at least 70 percent of CNV in terms of genomic length (not in terms of probes) had its status reassigned. If the segment was either centromeric or telomeric, its status was set to the one of its only neighbor. If the segment was between two segments of the same status then this status was assigned. If the two adjacent segments did not have the same status then the no change status was assigned.

Dimension reduction: We performed a dimension reduction technique to avoid duplicating statistical tests along the 225,388 probes when not necessary. We considered all the distinct breakpoints generated by the whole set of tumors and ordered them along the genome. Between two consecutives breakpoints, possibly emerging from 2 different tumors, it was clear that each tumor had a constant status. Then, for each tumor, all the probes in between these two breakpoints shared the same status information and could be shrunk to a single segment information. In our dataset, the 225,388 probes were thus reduced to 8,737 homogeneous segments. A schematic representation of this methodology is presented below.



Let T1, T2, ..Tk be k tumors of a dataset for which segmentation and calling gains and losses steps have already been done. The figure shows here the reduction principle on a single chromosome as the reduction is done independently for each autosomal chromosome. The colored dashes (red, green, blue, purple) represent breakpoints in the different tumors (T1, T2, T3, Tk respectively). Next, all the distinct breakpoints are aligned together on the chromosome thus defining *n* contiguous regions S1, S2, …, Sn. On the projection, one can identify from which tumor a breakpoint come by the correponding colored dash(es) next to it. Each tumor has a constant status in regions S1, S2, …, Sn. We can then reduce the dataset to the knowledge of status of these regions for all tumors without losing any information. Between two regions S*i* and S*i+1*, at least one tumor has a change of status. The number of statistical tests done is considerably reduced.

In our dataset : k=100 (59 LuminalA + 41 LuminalB), n=8737 (instead of 225388 probes)

Data Analysis: To test whether or not there was an association of a segment status with the Luminal A or Luminal B phenotype, the Fisher's exact test was used. Using a threshold about two-fold the normal breast expression levels pooled, genomic amplification and over-expression correlation were estimated using the Fisher's exact test. All statistical tests were two-sided at the 5% level of significance.

Real-time quantitative PCR (QPCR) RT-PCR (RQ-PCR). DNA and RNA form BCLs were isolated utilizing RNeasy Mini Kit (QIAGEN) and used for QPCR and RQPCR in a LightCycler FastStart DNA Master^{plus} SYBR Green I (Roche, Mannheim, Germany). DNA copy number was measured for 27 genes spanning the 8p12-11 region. Primers for the SYBR Green detection were designed with the software Primer 3 (http://frodo.wi.mit.edu/primer3/) and listed below. The relative DNA copy

number of each gene analyzed was computed with respect to internal standard *G6PD*. The expression mRNA level of *ZNF703* was computed with respect to the internal standard *GUSB* gene to normalize for variations in the quality of RNA and the amount of input cDNA.

Primers used for Q-PCR				
Genes	Primers sequence 5' to 3'		Hybridation temperature (°C)	Q-PCR product size (bp)
TMEM66	**F**	GATCATGTTAGCTCCACCTTCC	63	288
	R	TGTTAGGGTGACATTTGACTGG		
ZNF703	**F**	ACCCTCTTCCAGACACCTAAGC	60	110
	R	TTGAGGAAAGGCATTAAACTCG		
ERLIN2	**F**	TTGATAAGCACACTCCCTTGG	63	336
	R	GGGAAAGTGAGGGTAGTGAGG		
PROFC	**F**	CATTCACTCATATGGCTCTTTCC	63	246
	R	TTGATGCTTTTTCTAGCAGTTCC		
BRF2	**F**	GCGACTAACTTGTTCGTTTTCC	63	311
	R	GCTAGGAAAATTGCCTGTATGG		
RCP/RAB11FIP1	**F**	GCTTCTGGTTCTGTGATCTGC	63	220
	R	GGAAGAAGGATGTGGCTAAGG		
EIF4EBP1	**F**	AAGGGGACAGGAGAGTAAGACC	61	344
	R	GGGCTATGATGAGTAGCATTCC		
LFM1	**F**	TTCCATCTCGAAGCAGAACC	63	251
	R	CTGTCCTTTTCCCACTCTGC		
BAG4	**F**	TCAGATGCCTGTCACTGAGG	65	369
	R	CCACACTCCTGAGCTTCTCC		
DDHD2	**F**	TCAACTTCAGAGAAGCCTAGTGG	58	225
	R	CCATTGATAGCTAGGTCATTTGC		
PPAPDC1B	**F**	ACCTGGAGGACTGAGGTATGG	63	261
	R	CAACTAATGGCTTTGCTTTGC		
WHFC1L1	**F**	TGGACAACCAGTAGGCATAGG	63	216
	R	CTGACCACAGACCTTCAAAGC		
LETM2	**F**	TTTGCTCCTAGAGAAGGAGTGC	63	276
	R	GGCAAATAGCTTGGATACTTGG		
FGFR1	**F**	AGATATCCCAGGGAACACACC	63	351
	R	GGGAGGGTCTGTGTTAAAAGC		
TM2D2	**F**	TCTGTTTGTGCCATTACACTCC	61	272
	R	AGCCCAGACTCCTACTCCTACC		
C8orf4	**F**	AGGAGAACAAAATGCTTTCAGC	60	217
	R	CTCCCTTCCCCTATCTATTTGG		
ZMAT4	**F**	TGTTCAATGTCCCAGATACAGC	60	133
	R	CTTCTCGTCATCCAAACCTACC		
FFRP1	**F**	GACTTCTCCAATCCTGTGAACC	60	145
	R	TGCCTGAGTTCTCTTTCCTACC		
GOLGA7	**F**	CTGCTATTGGAGCTGATTTTCC	63	311
	R	CATTCTTCTACCCAGACCAAGG		
GINF4	**F**	GCACTTCCTAGTCACGTTCTCC	60	152
	R	CTAAATGAACTTGGGCTTCAGG		
AGPAT6	**F**	GCATTCTGCCTCTGAATTTACC	60	225
	R	ATCAATGGATGAAAGGATGACC		
NKX6-3	**F**	TTGCTTGGGATAGTTGAAGAGG	63	272
	R	TTTCTACCGCCTTAGAGTCTGG		
ANK1	**F**	GAAACAAAATGTGGCTTCTTCC	60	246
	R	CTTCCAGCACTTGGTTAAATCC		
	R	ATCCCACGTCATTCCATATAGC		
G6PD	**F**	AACCACAGCCTGCAAATACC	63	348
	R	CAGCTTAGCTCCCTGTCACC		

F: Forward primer R: Reverse primer

Primers used for RQ-PCR				
Genes	Primers sequence 5' to 3'		Hybridation temperature (°C)	QPCR product size (bp)
ZNF703	**F**	ATCAGGGTCCTGAAGATGCT	60	108
	R	CTTGGCGTCCAGCTCAAT		
GUSB	**F**	GAAAATATGTGGTTGGAGAGCTCATT	60	100
	R	CCGAGTGAAGATCCCCTTTTTA		

Multi Experiment Matrix (MEM) Analysis. Co-expression search for *ZNF703* (probeset 222760_at) was done using MEM web tool [5]. We selected only studies performed on Affymetrix GeneChip® Human Genome U133 Plus 2.0 [HG-U133_Plus_2] platform. Dataset selection was done using standard deviation σ=0.29 and by eliminating dataset related to transcriptional profiling of artificially modified cell lines. From the 764 starting datasets we computed 432 datasets. Correlation between *ZNF703* and coexpressed genes was determined using Pearson correlation test.

Western blotting. Cells were lysed in extraction buffer (1% (v/v) Triton X-100, 50 mM Hepes, pH 7, 1 mM EDTA, 1 mM EGTA, 150 mM NaCl, 100 mM sodium fluoride, 1 mM Na_3VO_4, and one tablet of CompleteTM inhibitor mix (Roche Applied Science) per 25 ml of buffer and loaded onto SDS-polyacrylamide gels. Blots were incubated with the respective primary antibodies diluted in TBST (containing 0.05% Tween20 and 5% non fat dry milk) either overnight at 4°C or for 2 hours at room temperature. Then, blots were washed and incubated with appropriate secondary antibodies (xxxx) and detected using SuperSignal West Pico Chemiluminescent Substrate (Pierce). For cellular fractionation Nuclear Extract Kit was used (Active Motif Europe, Belgium). Antibodies used for Western blotting were Anti-ZNF703 (Abcam ab40861, 1:200), anti-α-tubulin (Sigma-Aldrich T5168, 1:4000), anti-GFP (Roche 11814460001,

1:1000), anti HSP60 (Abcam ab46798, 1:1000), anti-DCAF7 (Sigma-Aldrich, HPA022962, 1:200), anti-ERα (Neomarkers Fremont CA, RB-9016-P1, 1:200), anti-GATA3 (Santa-Cruz Sc-268 1:200), anti-FOXA1 (Abnova Taiwan, H00003169M01, 1:200) anti-Rb (Santa-Cruz Sc-102 1:200) and anti-p27Kip1 (Dako, M7203, 1:200).

1. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5: 557-572.
2. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, et al. (2007) High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol 8: R215.
3. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, et al. (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 40: 1166-1174.
4. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, et al. (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949-951.
5. Adler P, Kolde R, Kull M, Tkachenko A, Peterson H, et al. (2009) Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. Genome Biol 10: R139.

Supplementary figure legends

Supplementary figure 1. Comparison of mRNA and protein expression of ZNF703, in breast cancer cell lines, measured by RQ-PCR, cDNA microarray, and immunoblot. A. ZNF703 gene expression was measured by reverse transcriptase-quantitative-polymerase chain reaction experiments (RTQ-PCR) and oligonucleotide microarrays in 11 breast cell lines. Correlation coefficients (r) calculated between RT-Q-PCR and oligonucleotide microarrays were 0.96. **B**. ZNF703 protein expression was evaluated in HCC1500, MDA-MB-134, 184-B5, MCF-10A and HME1 cell lines by immunoblotting using 100 µg of total soluble proteins separated in a 7.5% acrylamide gel. We used α-tubulin as a loading control.

Supplementary figure 2. Copy number alterations frequency and mRNA expression for 8p12 genes in breast cancer cell lines. For each breast cancer cell line, measures of gene copy number and their transcript expression levels are respectively reported in diagram **(A)** and **(B)**, for 8p12 genes, ordered from telomere (left) to centromere (right) on X-axis. In **(A)**, Y-axis represents gene copy number determined by QPCR (related to normal breast cell lines; HME1, 184-B5 and MCF-10A) while in **(B)** it represents mRNA expression level determined by cDNA microarrays. Each analysed cell lines are represented by different colored lines and dashed lines as indicated in respective diagrams **(A)** and **(B)** (left panels). Amplification was stringently defined with a gene copy number > 4 (larger yellow line) **(A).** Baselines for up and down regulated expression are respectively mentioned (larger yellow and violet lines, respectively) **(B)**.

Supplementary figure 3. *ZNF703* expression is regulated by estrogen stimulation. A. ZNF703 mRNA and protein expression in MDA-MB-134 breast cancer cell lines after 17-β estradiol (E2) stimulation. ZNF703 expression was determined for different time points after E2 addition in culture media. Experiments were made in duplicate. **B**. ZNF703 protein expression was evaluated in MDA-MB-134 and HCC1500 cell lines by western blotting. Different treatments were used: standard culture medium (FBS), without E2 (FBS-Charcoal), and with two E2 concentrations (1nM and 10nM). When medium is depleted in E2 (FBS-Charcoal) ZNF703 expression decreases. ZNF703 expression is restored when E2 is supplied.

Supplementary figure 4. GFP-ZNF703 localized both to the cytoplasm and the nucleus in MCF7 cell line. Nuclear and cytoplasmic fraction of MCF7 GFP and GFP-ZNF703 were analysed by immunoblotting after SDS-PAGE resolution. Anti-Sp1 and anti-α-tubulin antibodies were used as control of compartment contamination. Anti-GFP was used to detect GFP and GFP-ZNF703 proteins.

Supplementary figure 5. Immunofluorescence analysis of ZNF703 co-localisation with HSP60 and PML in MCF7 GFP and GFP-ZNF703 cell lines. MCF7 overexpressing GFP-ZNF703 and GFP cells were fixed and immunostained with anti-ZNF703 (**A**), anti-PML (**B**) and anti-HSP60 (**C**) antibodies. GFP-ZNF703 (green) nuclear dot-like structures do not colocalizes with HSP60 (red) and PML (red). Immunostaining with anti-ZNF703 (red) shows complete colocalization with

GFP-ZNF703 (green). Moreover, anti-ZNF703 staining reveals nuclear dot-like structures in MCF7 GFP breast cancer cell line.

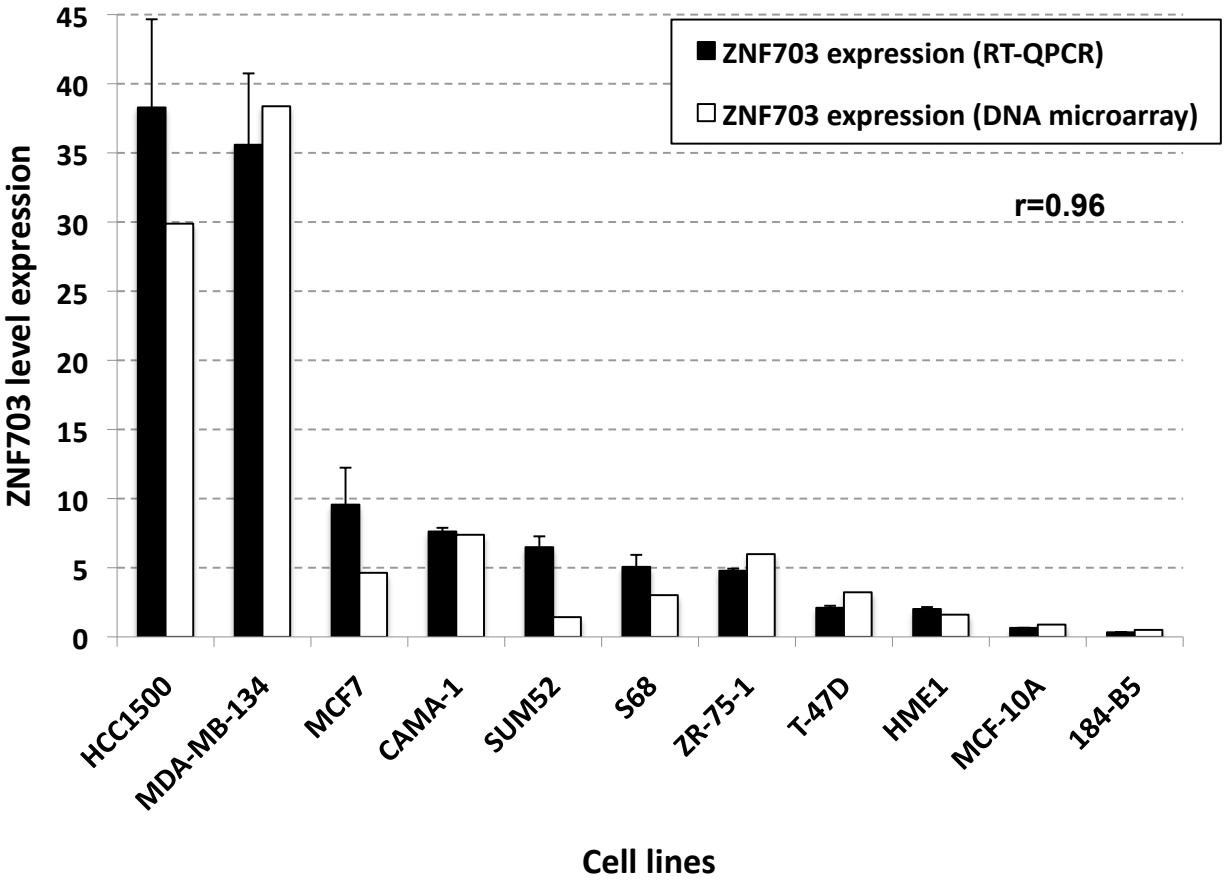
Supplementary Figure 6. GFP-ZNF703 interact with HSP60 and DCAF7 in MCF7 GFP-ZNF703 breast cancer cell lines. The interaction between GFP-ZNF703 and HSP60 or DCAF7 was confirmed by GFP-immunoprecipitation, followed by Western blot analysis using anti-HSP60 or anti-DCAF7 antibody. Input represents whole cell extract from MCF7 GFP and MCF7 GFP-ZNF703.

Supplementary Figure 7. Validation of ZNF703 knock-down efficiency using ZNF703-siRNA in MDA-MB-134 cell line. MDA-MB-134 cells lipofected with ZNF703-siRNA (using three independent siRNA constructs) or with control-siRNA were collected after 30 hours of lipofection. ZNF703 protein expression was measured by immunobloting. All three ZNF703-siRNA presented a significant efficiency in knocking down ZNF703 protein expression compared to control-siRNA.

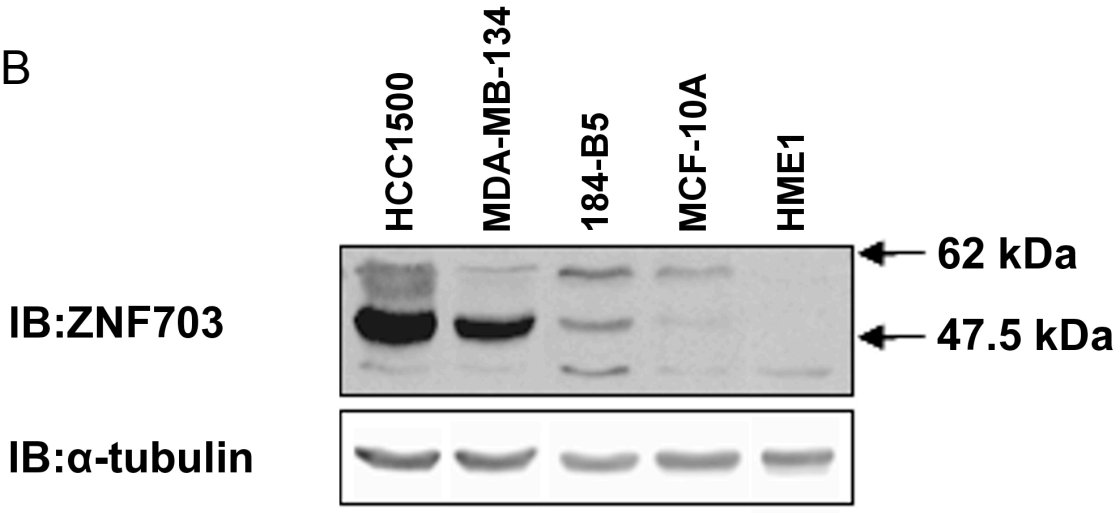
Supplementary Figure 8. Role of ZNF703 in the regulation of cell cyle progress. To determine the effect of ZNF703 on the cell cycle we analyzed cell cycle status of synchronized cells from MCF7 GFP (**A** and **B**) and MCF7 GFP-ZNF703 (**C** and **D**) cell line in the presence or absence of E2. This analysis was done every 2 hours over 24 hours after FBS stimulation. Results are summarized in the left panel with graphs presenting the percentage of cells in G0/G1, S, and G2/M. Examples of cell cycle flow charts are presented on the right panel for four time points (0h, 6h, 10h, 22h). MCF7 GFP needs E2 stimulation to enter the cell cycle whereas ZNF703 overexpression is sufficient to induce cell cycle entry independently of E2 stimulation.

Interestingly, in the presence of E2, G0/G1 phase is significantly shorter (4h) in MC7 GFP-ZNF703 cells compared to cells cultured in the absence of E2 (6h).

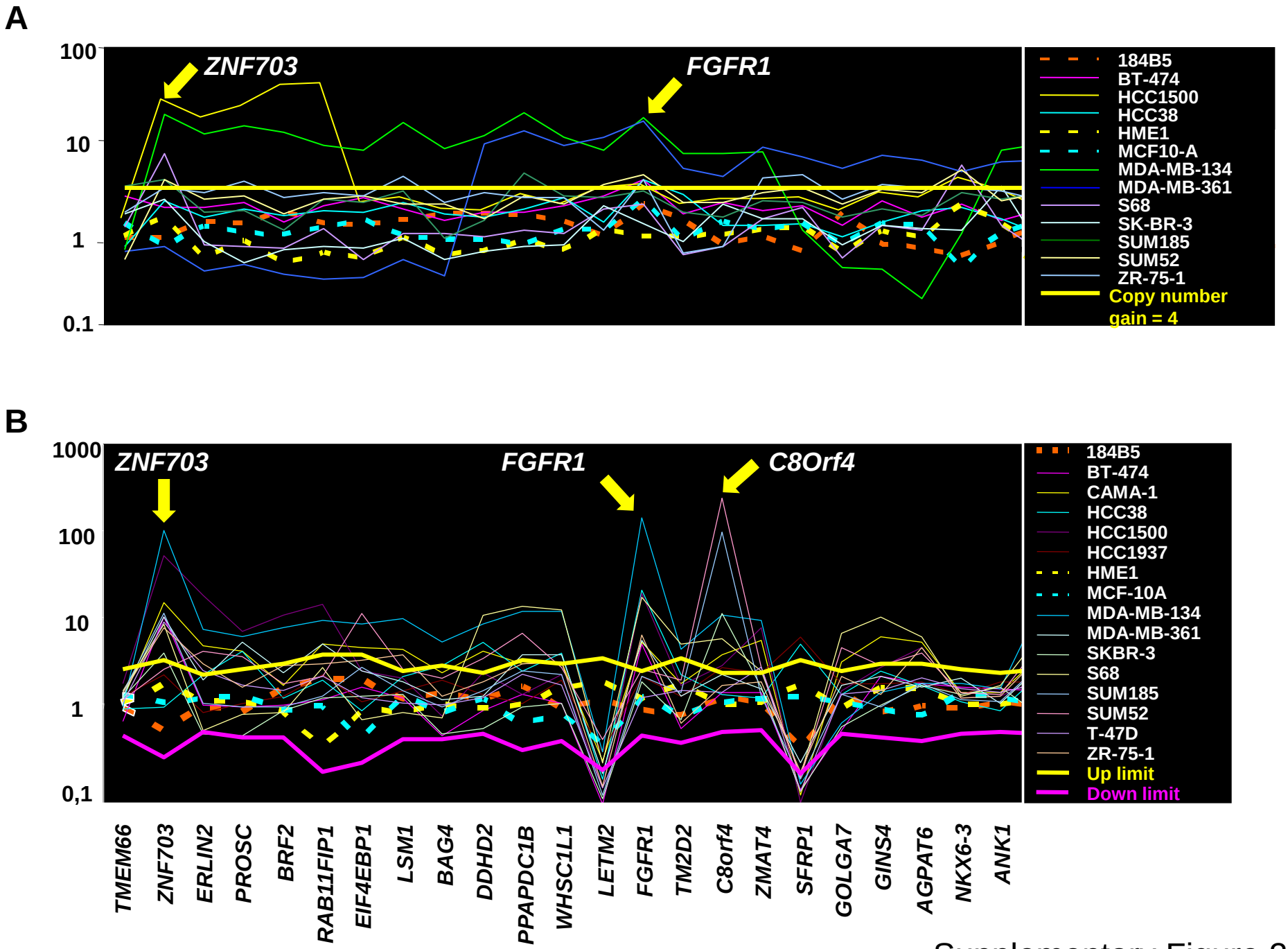
A



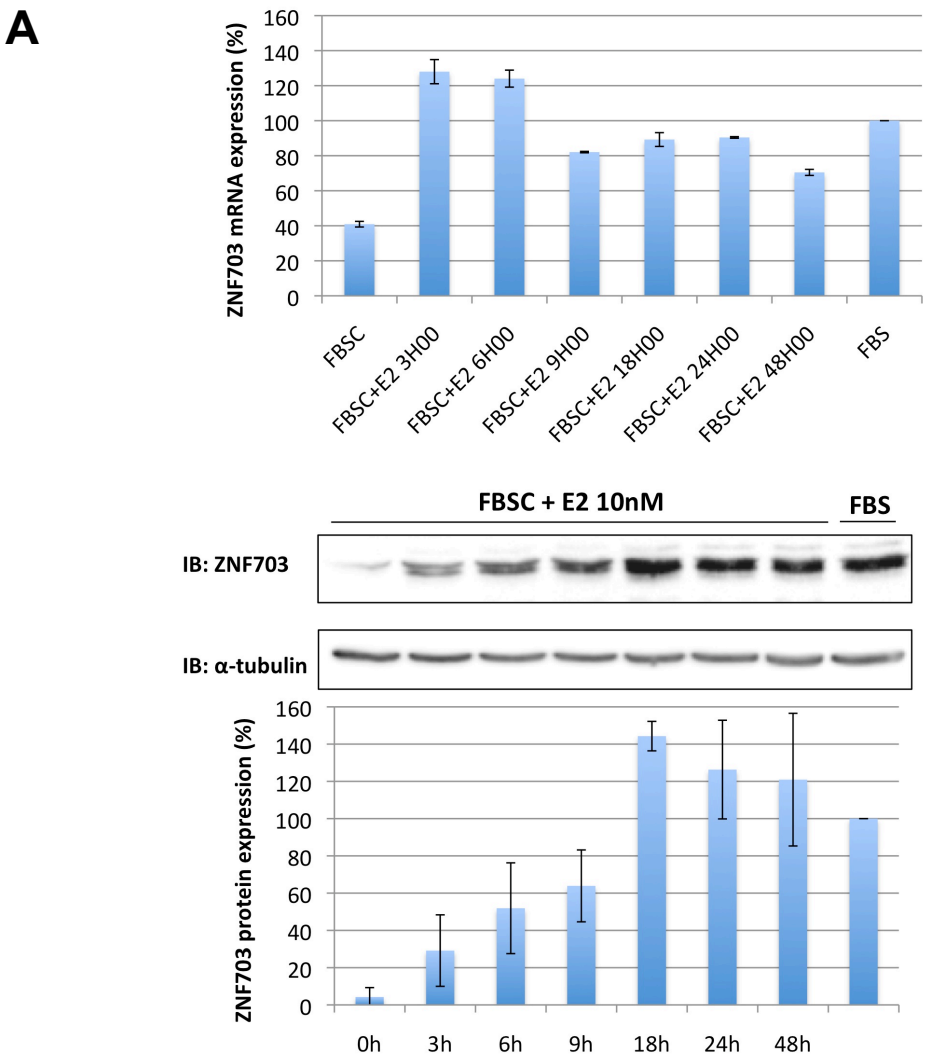
B

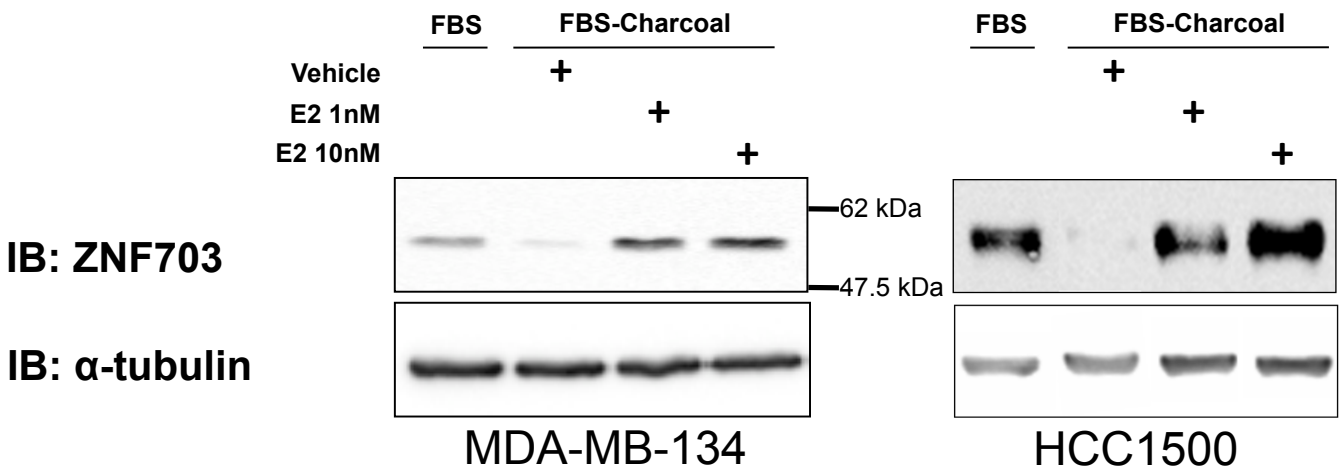


Supplementary Figure 1

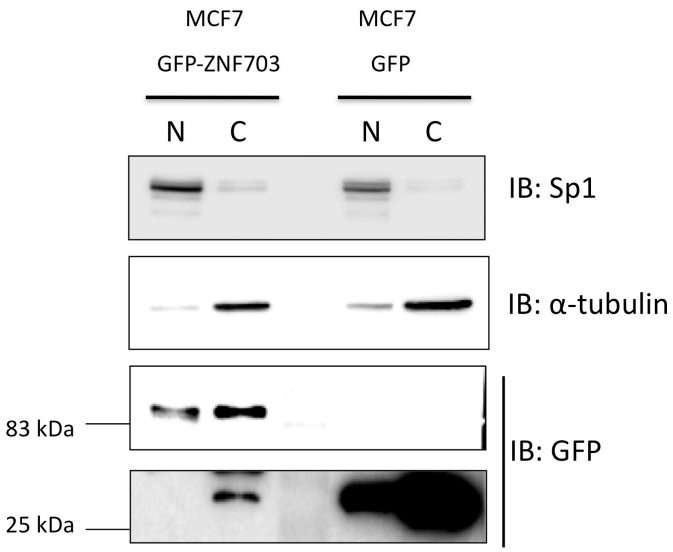


Supplementary Figure 2

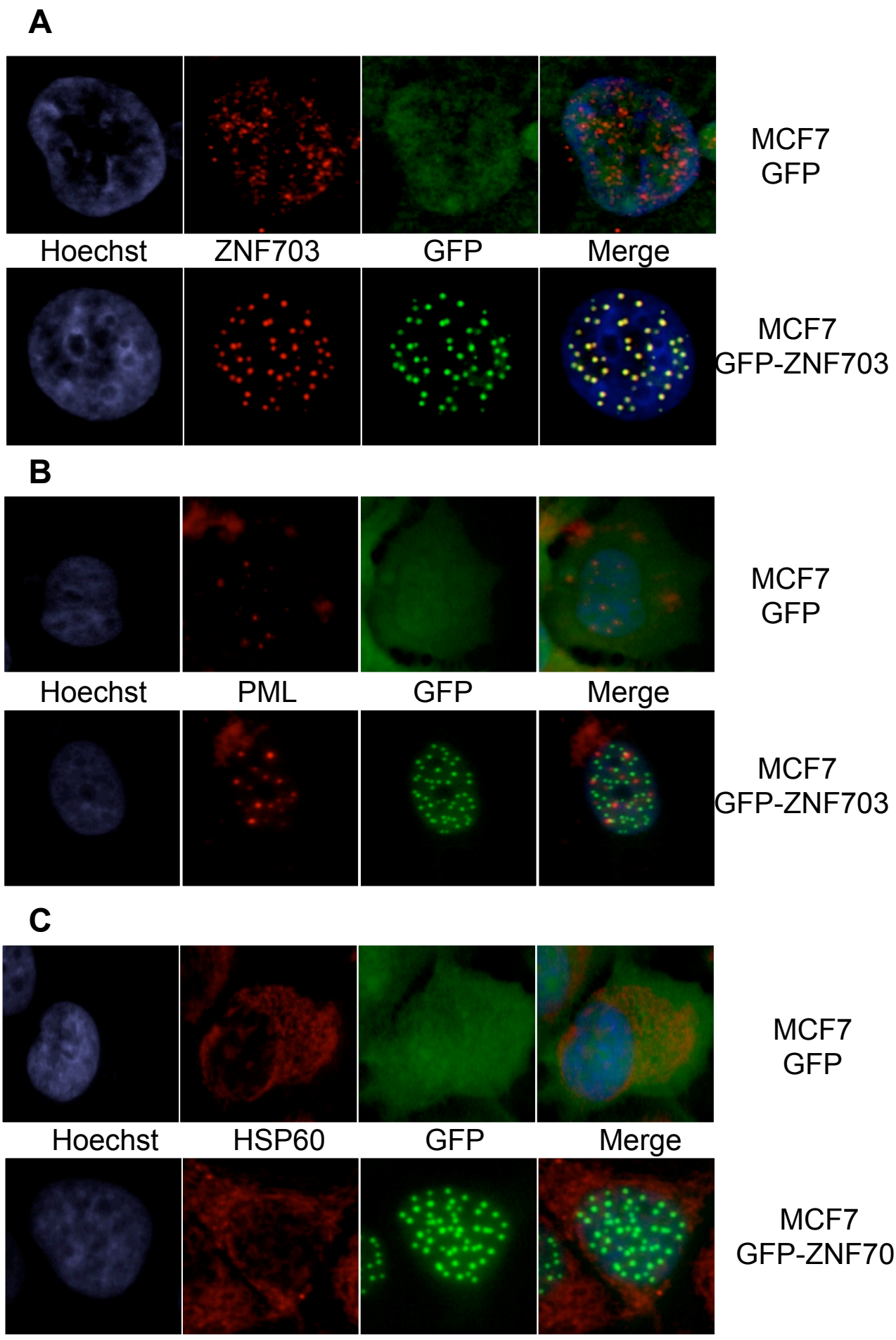




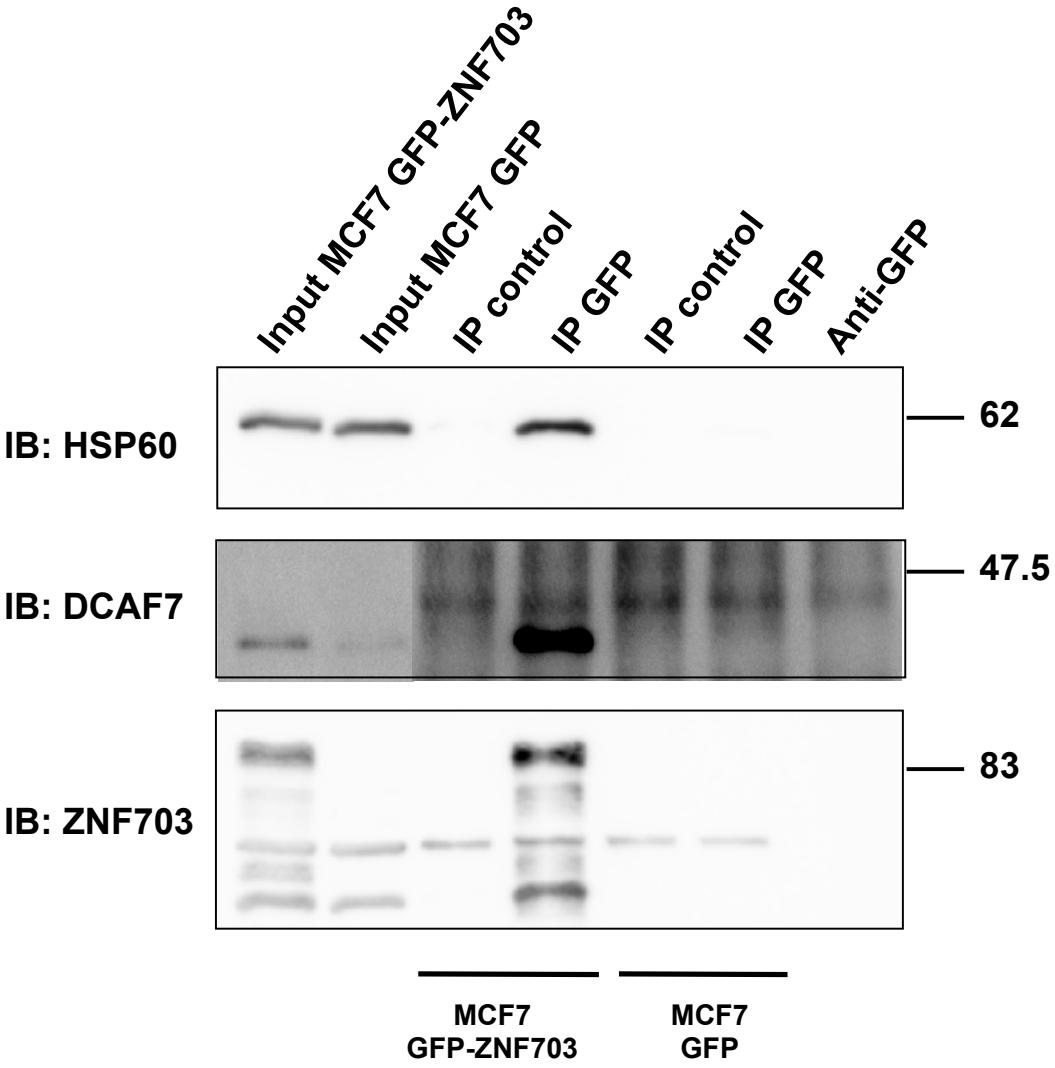
Supplementary Figure 3

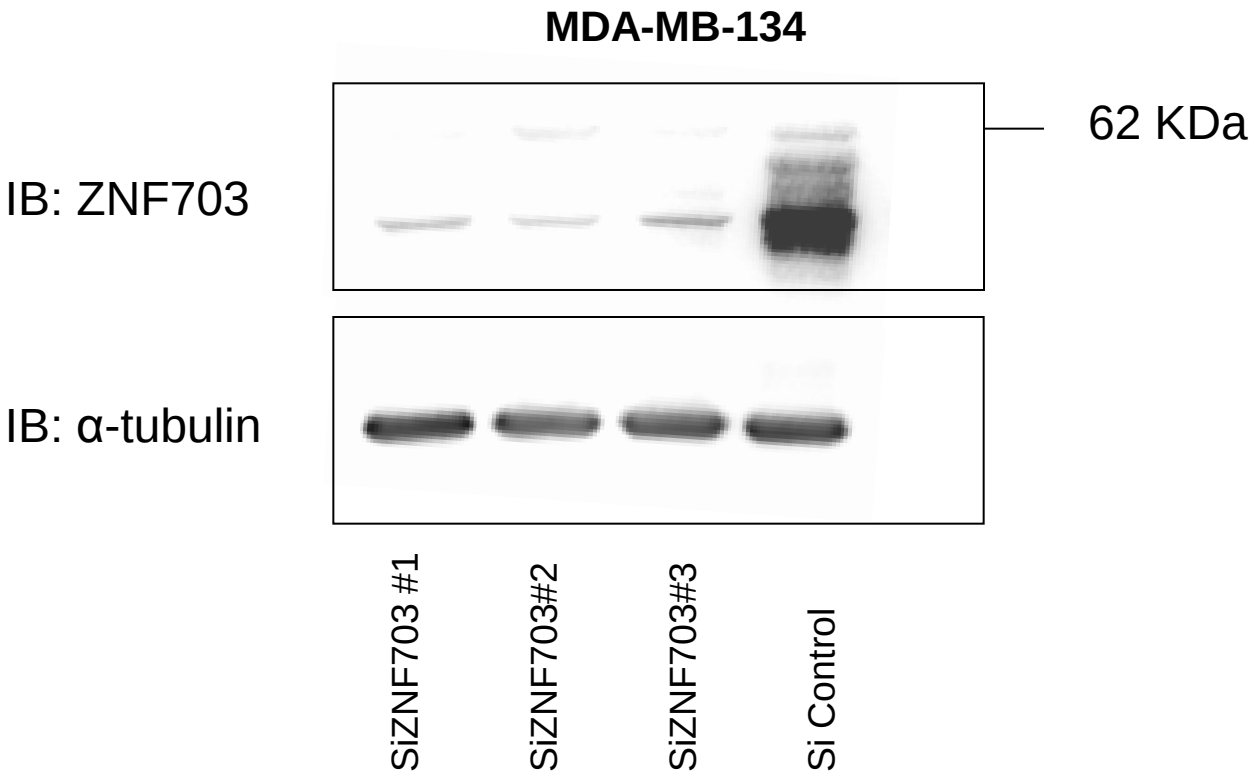


Supplementary Figure 4

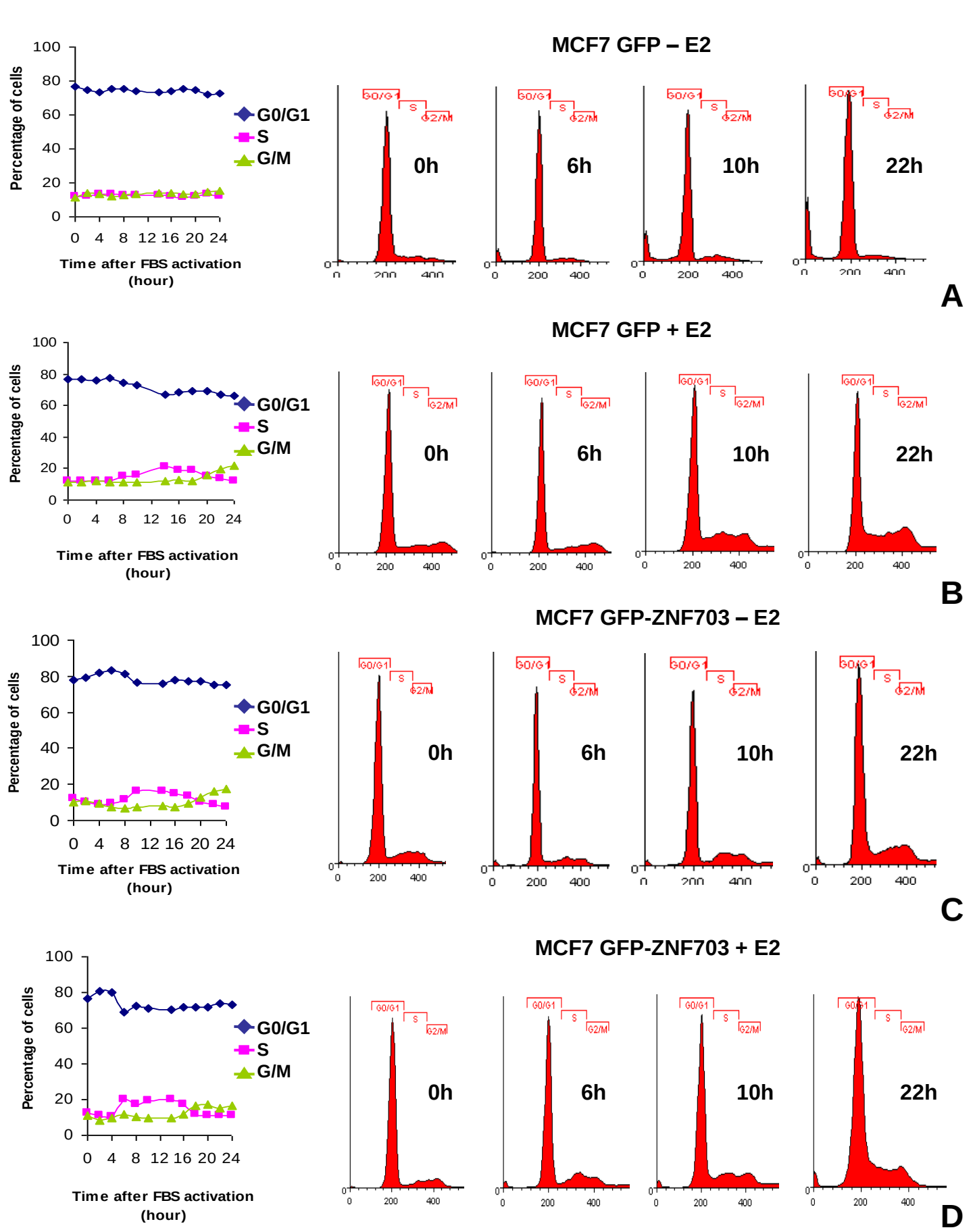


Supplementary Figure 5


Input MCF7 GFP-ZNF703
Input MCF7 GFP
IP control
IP GFP
IP control
IP GFP
Anti-GFP
IB: HSP60
62
IB: DCAF7
47.5
IB: ZNF703
83
MCF7
GFP-ZNF703
MCF7
GFP



Supplementary Figure 7



Supplementary Figure 8