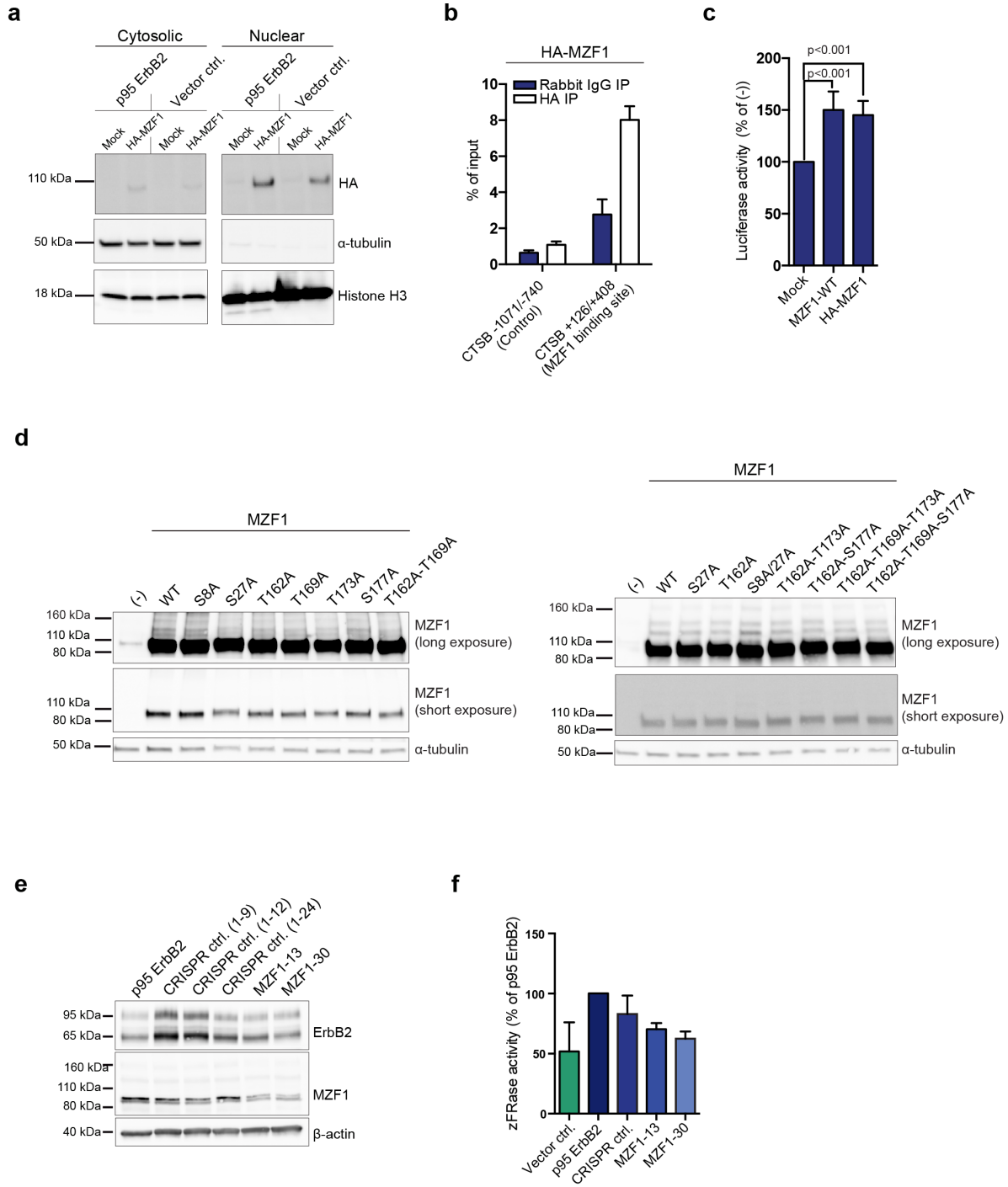
Supplementary Figures and supplementary Materials and Methods

Release of transcriptional repression via ErbB2-induced, SUMO-directed phosphorylation of Myeloid Zinc Finger-1 Serine 27 activates lysosome redistribution and invasion

Ditte Marie Brix, Siri Amanda Tvingsholm, Malene Bredahl Hansen, Knut Bundgaard Clemmensen, Tiina Ohman, Valentina Siino, Matteo Lambrughi, Klaus Hansen, Pietri Puustinen, Irina Gromova, Peter James, Elena Papaleo, Markku Varjosalo, José Moreira, Marja Jäättelä and Tuula Kallunki

Supplementary Figures:



Supplementary Figure S1. MZF1 is phosphorylated at S27 in response to ErbB2 signaling

a The HA-tagged MZF1 is predominantly expressed in the nucleus. The p95-ErbB2-MCF7 cells were transiently transfected with HA-MZF1, and subjected to nuclear and cytosolic fractionation. The fractions were blotted for the detection of HA, α-tubulin, and histone H3. The blot is a representative of three independent experiments.

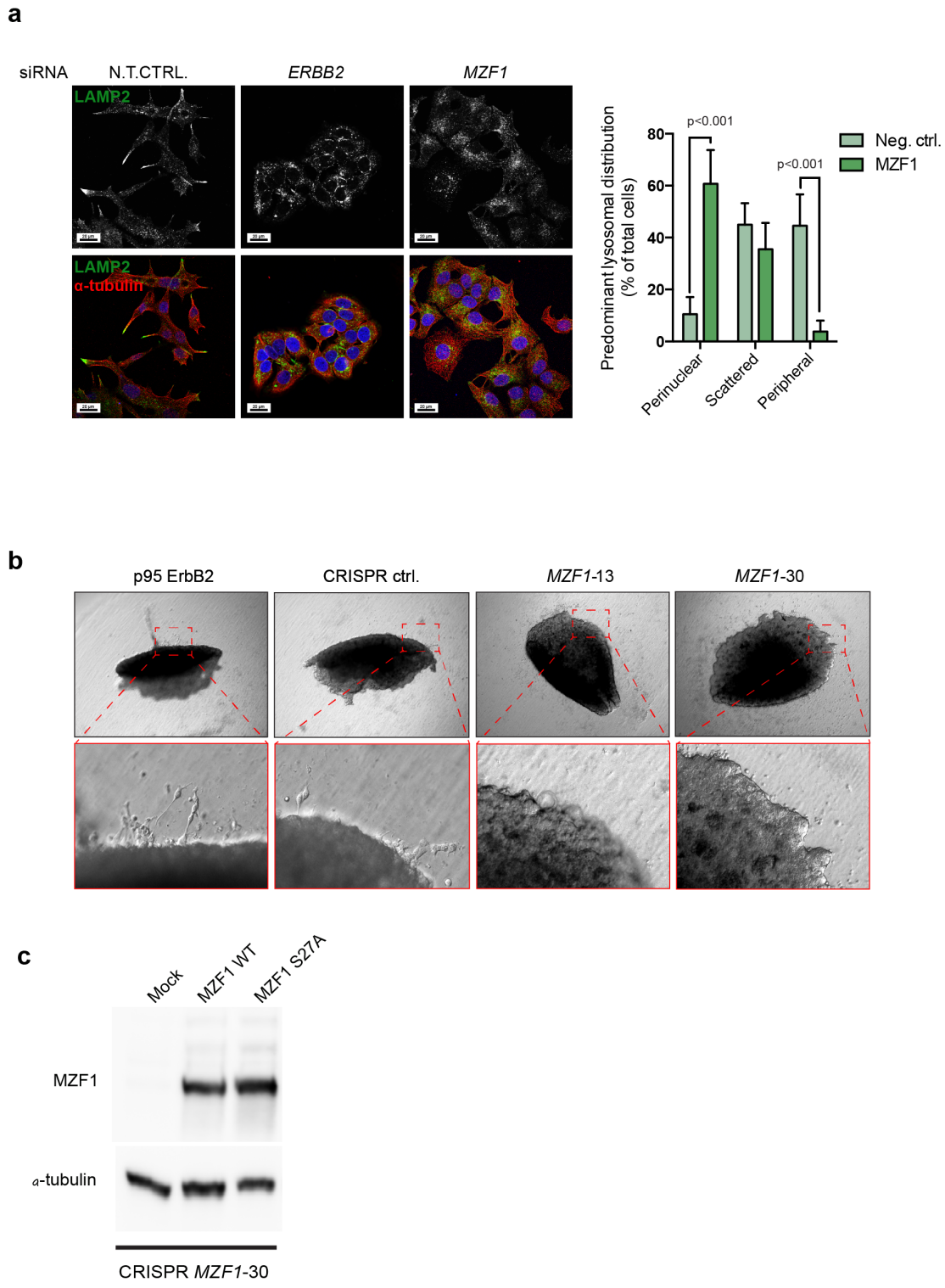
b HA-tagged MZF1 binds to the ErbB2-induced enhancer in the *CTSB* first intron. Chromatin immunoprecipitation assay for detecting the HA-tagged MZF1 (HA-MZF1) binding to the *CTSB* enhancer region *in vivo*. The HA-MZF1 was immunoprecipitated with anti-HA antibody, and the *CTSB* enhancer fragment (+126/+408) was amplified by quantitative RT PCR. The results are presented as percentage of input. The data are representative of two experiments. Values present mean ± standard deviation of technical replicates from one experiment.

c HA-MZF1 can activate *CTSB* luciferase expression in response to ErbB2 in a similar manner than the non-tagged WT MZF1. *CTSB* reporter assay. The p95-ErbB2-MCF7 cells were transfected with the *CTSB* luciferase and *Renilla* luciferase and either with the empty (mock), wild type MZF1-WT or HA-MZF1 pcDNA3.1 plasmids. The reporter activity was calculated as the firefly luciferase activity divided with *Renilla* luciferase activity and presented as the percentage of the mock control. Data are presented as mean ± standard deviation of four independent experiments. Statistical significance was calculated with one-way ANOVA with Dunnett's correction.

d WT and mutant MZF1 constructs are expressed at the same level. The immunoblot analysis to detect the MZF1 in p95-ErbB2-MCF7 cells that were transiently transfected with either empty (mock), MZF1-WT or indicated mutant MZF1 pcDNA3.1 plasmids. The membranes were blotted with the MZF1 and ErbB2 antibodies, and α-tubulin was used as loading control. The immunoblot is a representative of at least three independent experiments.

e MZF1 expression is decreased in MZF1 CRISPR knockdown cell lines *MZF1*-13 and *MZF1*-30. Immunoblot analysis for the detection of ErbB2 and MZF1 in lysates of p95-ErbB2-MCF7 MZF1 CRISPR cell lines. β-actin was used as loading control.

f Cysteine cathepsin (zFRase; cathepsin B and L) activity is decreased in MZF1-depleted cells. Cysteine cathepsin activity measurements of p95-ErbB2-MCF7 MZF1 CRISPR knockdown cell lines. Data are presented as mean ± standard deviation of two independent experiments.



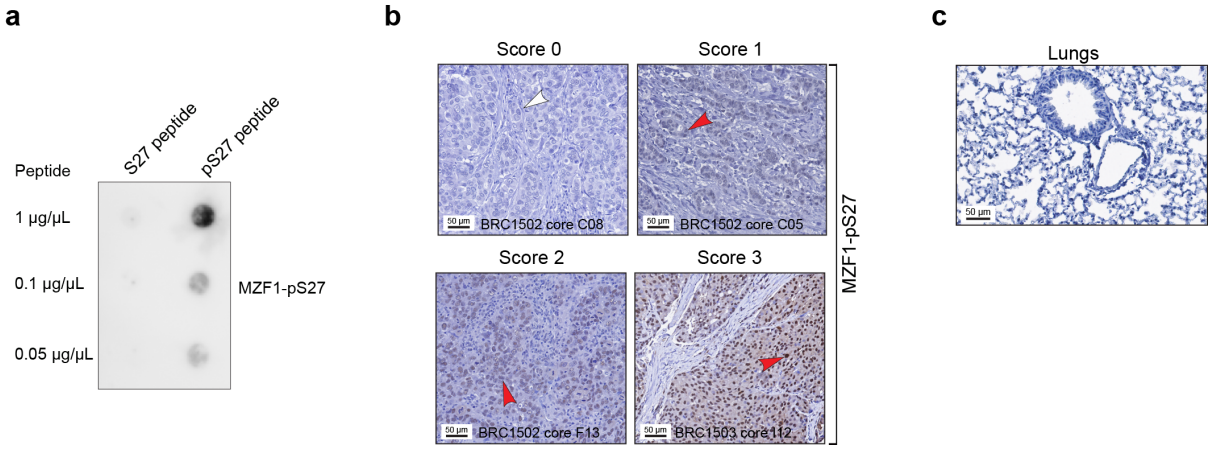
Supplementary Figure S2. MZF1-WT, but not MZF1-S27A, can mediate invasive ErbB2 signaling in breast cancer cells

a *ERBB2* and *MZF1* depletion results in the distribution of lysosomes at the perinuclear region. Left: Representative confocal images of p95-ErbB2-MCF7 cells transfected with 20

nM concentrations of control, *ERBB2* or *MZF1* siRNA for 72h. The cells were fixed and stained stained for LAMP2 (green), α-tubulin (red) and Hoechst (blue). Right: Quantification of predominant lysosomal distribution. Data are presented as mean ± standard deviation of three independent experiments. Statistical significance was calculated with Student's *t*-test (unpaired with Welch's correction).

b MZF1 depleted *MZF1*-13 and 30 cells are less invasive in 3D Matrigel invasion assay than corresponding MZF1 expressing control cells. Images of the representative invasion assays quantified in Figure 2c.

c Control immunoblot analysis for the invasion assay in Figure 2d. Lysates from *MZF1*-30 cells that were transiently transfected with indicated plasmids, and used for the invasion assays in Figure 2d. Transfected cells were set up for invasion assays 24h after transfection and the rest of the cells were plated, cultured for additional 48h, and harvested at 72h after transfection for the detection of the transiently expressed MZF1.

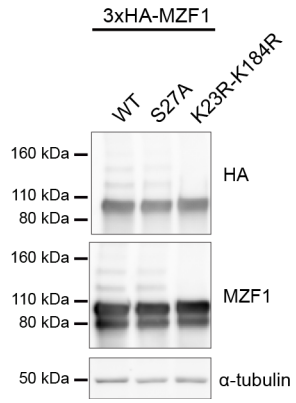


Supplementary Figure S3. Phosphorylation of S27 correlates significantly with high HER2 and EGFR status

a The anti-MZF1-pS27 antibody preferentially recognizes S27-phosphorylated MZF1 peptide. Dot blot with S27- and pS27-containing peptides blotted with pS27 antibody.

b Example of the scoring of the MZF1-pS27 expression in breast cancer TMA including 321 primary samples of normal and breast cancer tissue: scores 0-3.

c IHC negative control: IHC staining of negative control sample, mouse lung, with the anti-MZF1-pS27 antibody. Negative control was stained together with the breast cancer TMA (Figure 3d and e, Supplementary Figure S3b).



Supplementary Figure S4. SUMOylation of K23 cooperates with the phosphorylation of S27 in MZF1 activation

WT and mutant Δt-3xHA MZF1 constructs are expressed at the same level. Immunoblot analysis for the detection of HA-tagged MZF1 constructs in p95-ErbB2-MCF7 CRISPR *MZF1*-30 cells transiently transfected with either MZF1-WT or indicated mutant MZF1 Δt-3xHA plasmids. Membranes were blotted with HA and MZF1 antibodies and α-tubulin was used as loading control. The immunoblot is a representative of at least three independent experiments.

Supplementary Materials and Methods:

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-LAMP2	Developmental studies Hybridoma Bank	Cat#H4B4
Anti-HA clone H-7	Sigma-Aldrich	Cat#H3663
Anti-MZF1	Abcam	Cat#ab64866
Anti-phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (20G11)	Cell Signaling Technology	Cat#4376S
Anti-PAK4	Cell Signaling Technology	Cat#3242
Anti-SUMO1	Abcam	Cat#ab11672

Anti-SUMO2/3	Abcam	Cat#ab3742
Anti-histone H3 (FL-136)	Santa Cruz	Cat#sc-10809
Anti-α-tubulin	Abcam	Cat#ab15246
Anti-HER-2/c-erbB-2/neu Ab-17	Thermo Scientific	Cat#MS-730-P0-A
Anti-β-actin	Sigma-Aldrich	Cat#A2228
Anti-phospho-S27-MZF1	This paper	N/A
Alexa Fluor 555 Phalloidin	Life Technologies	Cat#A34055
Alexa Fluor 488 Donkey Anti-Mouse IgG (H+L) antibody	Life Technologies	Cat#A-21202
Alexa Fluor 594 Donkey Anti-Rabbit IgG (H+L) antibody	Life Technologies	Cat#A-21207
Alexa Fluor 488 Donkey Anti-Rabbit IgG (H+L) antibody	Life Technologies	Cat#A-21206
Negative control mouse IgG	DAKO	Cat#X0931
Rabbit IgG Isotype Control	Invitrogen	Cat#31235
anti-mouse-HRP (WB)	DAKO	Cat#P0260
anti-rabbit-HRP (WB)	Vector Laboratories	Cat#PI-1000
anti-rabbit-HRP (IHC)	Envision+Systems, DAKO	Cat#K4002

Biological Samples		
Breast cancer tissue micro arrays	Pantomics, Inc.	BRC1501; BRC1502; BRC1503

Chemicals, Peptides, and Recombinant Proteins		
Lapatinib Ditosylate	Santa Cruz Biotechnology	Cat#SC-202205
PF-37583909 (PAK4 inhibitor)	Selleck Chemicals	Cat#S7094
Doxycycline Hyclate	Sigma-Aldrich	Cat#D9891
N-ethylmaleimide	Sigma-Aldrich	Cat#E3876
PhosSTOP (Phosphatase Inhibitor Cocktail)	Sigma-Aldrich	Cat#04906837001
cOmplete Mini (Protease Inhibitor Cocktail)	Sigma-Aldrich	Cat#04693124001
Benzonase Nuclease	Sigma-Aldrich	Cat#E1014
Lambda protein phosphatase (λPPase)	New England Biolabs	Cat#P0753
DMSO	VWR	Cat#AMREN182
S27 peptide sequence: VKLEDSEEEGEA-Cys	JPT Peptide Technologies GmbH	N/A
Phospho-S27 peptide sequence: VKLEDS(p)EEEGEA-Cys	JPT Peptide Technologies GmbH	N/A

Commercial Assays/kits		
Dual-luciferase™ Reporter Assay System	Promega	Cat#E1960
Pierce™ BCA Protein Assay Kit	Life Technologies	Cat# 23228
Nucleobond® Extra Maxi EF	Macherey-Nagel	Cat#740424

Imject® Maleimide Activated mcKLH	Thermo Fisher Scientific	Cat#77605
SulfoLink™ Immobilisation Kit for Peptides	Thermo Fisher Scientific	Cat#44999
Brilliant III UltraFast SYBR Green QPCR Master Mix	Agilent Technologies	Cat# 600882

Oligonucleotides (5´-> 3´) and siRNAs		
CTSB fwd: CCGACAGGGGATGGAAAGAG CTSB rev: CAGCGCTGGGTGGATCTAGG	TAG Copenhagen	N/A
MYC-forward: GGGAGGCTATTCTGCCCATT MYC-reverse: TTCTCCTCCTCGTCGCAGTA´	TAG Copenhagen	N/A
PRKCA fwd: GATAAGGGACCCGACACTGA PRKCA rev: CATGGAGCTTTTCATCAGCA	TAG Copenhagen	N/A
PPIB fwd: GGGAGATGGCACAGGAGG PPIB rev: TGGGAGCCGTTGGTGTCT	TAG Copenhagen	N/A
S27A fwd: GGTGAAGCTAGAGGACGCTGAGGAGGAGGGTG S27A rev: CACCCTCCTCCTCAGCGTCCTCTAGCTTCAC	TAG Copenhagen	N/A
T162A fwd: GAAACTGAGCCTCCAGCTCCAGAGCCTGGG T162A rev: CCCAGGCTCTGGAGCTGGAGGCTCAGTTTC	TAG Copenhagen	N/A
T169A fwd: CCTGGGCCCAAGGCACCTCCTAGGAC T169A rev: GTCCTAGGAGGTGCCTTGGGCCCAGG	TAG Copenhagen	N/A
S177A fwd: GGACTATGCAGGAAGCACCACTGGGCCTGC S177A rev: GCAGGCCCAGTGGTGCTTCCTGCATAGTCC	TAG Copenhagen	N/A
K23R fwd: CCTGTCATGGTGAGGCTAGAGGACTC K23R rev: GAGTCCTCTAGCCTCACCATGACAGG	TAG Copenhagen	N/A
K184R fwd: GCCTGCAGGTGAGAGAGGAGTCAGAG- K184R rev: CTCTGACTCCTCTCTCACCTGCAGGC-3	TAG Copenhagen	N/A
Negative control siRNA	Qiagen	Cat#1027281
MZF1 siRNA	Santa Cruz	Cat#SC-45714A
ERBB2 siRNA (pool of three)	Sigma-Aldrich	Cat#SASI_Hs01_00010288; Cat#SASI_Hs01_00010289; Cat#SASI_Hs01_00010291
ChIP: +126/+408 CTSB promoter fragment forward: 5'-CAGCGCTGGGTGAGTGCTGG-3' ChIP: +126/+408 CTSB promoter fragment reverse: 5'-CACAGCCCGGGGAGGGGAG-3'	TAG Copenhagen	N/A
ChIP: +1071/-740 Negative control fragment forward: 5´-GTG CGA TCT CTG GTC ACC GCA-3´ ChIP: +1071/-740 Negative control fragment reverse: 5'-CCT CCT CCA GAC ACT CCC CAA ACA-3'	TAG Copenhagen	N/A
S27E forward: 5´-GGT GAA GCT AGA GGA CGA GGA GGA GGA GGG TG-3´ S27E reverse: 5´-CAC CCT CCT CCT CCT CGT CCT CTA GCT TCA CC-3´	TAG Copenhagen	N/A

Ser27Glu-mut-forw: 5´- GTG ATG GTG AAA CTG GAA GAT GAG GAG GAA GAG GGC GAG G-3´ Ser27Glu-mut-rev: 5´- CCT CGC CCT CTT CCT CCT CAT CTT CCA GTT TCA CCA TCA C-3´	TAG Copenhagen	N/A
MZF1-Met-Ala-mut-forw: 5´ -CAA AAA AGC AGG CTT CGC GCG GCC TGC CGT GCT G-3´ MZF1-Met-Ala-mut-rev: 5´- CAG CAC GGC AGG CCG CGC GAA GCC TGC TTT TTT G-3´	TAG Copenhagen	N/A

Recombinant DNA/plasmids		
pLVX-Tet-On	Clontech	632162
pCR2.1-TOPO	Thermo Fisher Scientific	K4500-40
pcDNA3.1 (+)	Invitrogen	V79020
Δt-3xHA/DEST	Drs. Stephan Geley and Marin Barisic	N/A
pTAL-CTSB+126/+408-Luc	Rafn et al.,2014, Mol. Cell	N/A
pGL3 basic null Renilla luciferase vector	Promega	E2231
SRα-HA-PAK4 KM	Prof. Audrey Minden Qu et al., 2001, MCB	N/A
pTO_HA_StrepIII_c_GW_FRT	Prof. Markku Varjosalo Varjosalo et al., 2013, Cell Rep	N/A
pOG44 Flp recombinase expression vector	Thermo Fisher Scientific	V600520
Synthetic MZF1 clones in pENTR221 for MZF1-WT, MZF1-K23+K184, MZF1-S27A	Thermo Fisher Scientific	N/A

Software and Algorithms		
ZEISS ZEN Microscope Software	ZEISS	https://www.zeiss.com/microscopy/int/products/microscope-software/zen.html
Nano Zoomer Digital Pathology NDP.view2	Hamamatsu	https://www.hamamatsu.com/jp/en/U12388-01.html
Mascot	Matrix Science	http://www.matrixscience.com
MaxQuant	MaxQuant	http://www.coxdocs.org/doku.php?id=maxquant:start
Perseus	Perseus	http://www.coxdocs.org/doku.php?id=perseus:start
Proteome Discoverer software	Thermo Fisher Scientific	https://www.thermofisher.com/order/catalog/product/OPTON-30795

Cell lines

MCF7 M8-tTAS-pTRE-ΔNErbB2 cells (p95-ErbB2-MCF7 cells) and the corresponding vector control MCF7 M8-tTAS-pTRE cells were established and cultured in the presence of tetracycline, as described previously (1, 2). Prior to setting up the experiments, the expression of p95 ErbB2 was induced by washing off tetracycline (1, 2). Stable MZF1 CRISPR-Cas9 indel cell lines were generated from p95 ErbB2-MCF7 cells using the CRISPR-Cas9 construct (HS0000221698) designed by Sigma-Aldrich. The CRISPR cell line *MZF1*-13 and *MZF1*-30 indels contained five-nucleotide and one-nucleotide deletions, respectively. All cell lines of MCF7 origin were cultured in RPMI 1640 Glutamax™ (GIBCO™, Life Technologies) that was supplemented with 6% FBS and 0.25% penicillin and streptomycin. Lysates of various breast cancer cell lines including normal breast-like (MCF-10A), luminal A (ZR-751, MCF7, T47D, HCC1428), luminal B (BT474), and HER2/ErbB2 (HCC1954, HCC202, HCC1419, SK-BR-3) cells, were provided by I. Gromova. Cells were cultured in DMEM (GIBCO™, Life Technologies) that was supplemented with 10% FBS and 0.25% penicillin and streptomycin. To establish MZF1-expressing Flp-In 293 T-REx cell lines, tetracycline-inducible expression vectors for each MZF1 mutant and GFP (with nuclear localization signal, negative control) with N-terminal Strep-HA (SH) tag were generated by LR recombination into the corresponding destination vector (3). The Flp-In 293 T-REx cells (Invitrogen) were cultured according to the manufacturer´s instructions. The constructs were transfected with FuGENE HD Transfection Reagent (Promega) together with the pOG44 Flp recombinase expression vector (Thermo Fisher Scientific). The cells were grown under selection with hygromycin B (Thermo Fisher Scientific) to create stable cell lines. Cells were routinely tested to confirm the absence of mycoplasma contamination.

Clinical samples

Breast cancer tissue microarrays (BRC1501, 1502 and 1503), each containing 75 duplicate cores of normal, premalignant, and malignant breast tissues were obtained from Pantomics, Inc.

Drug treatment

Lapatinib ditosylate (Santa Cruz Biotechnology) was used at 1 μM or 10 μM for 16 h or 24 h. Stocks were dissolved in DMSO and diluted in full growth medium prior to treatment.

Plasmids

The plasmid for HA-tagged wild-type (WT) MZF1, HA-MZF1 was constructed by inserting a 2xHA tag with an *EcoRI* restriction site at the 5'-end of the N-terminus of WT MZF1 (734 aa) in a pCR2.1-TOPO vector using PCR. The HA-tagged insert was then subcloned into the pcDNA3.1(+) expression vector. Point mutations were introduced into the WT MZF1 using a pCR2.1-TOPO vector and site-directed mutagenesis using the PfuUltra II Fusion HS DNA polymerase (Agilent) according to the manufacturer's protocol. All the primer pairs are listed in the Key Resources Table. For those constructs, in which several sites were mutated, the mutations were introduced sequentially. Following sequence verification, the mutated inserts were then inserted into the pcDNA3.1(+) expression vector. HA-tagged mutant plasmids (Δt-3xHA/DEST) were constructed using the Gateway cloning system (Gateway) from the entry constructs derived from synthetic MZF1 by Thermo Fisher Scientific Geneart GmbH. Site directed mutagenesis was used to convert the first Metionine of the synthetic MZF1 pENTR constructs into Alanine prior to cloning to the N-terminal Strep-HA (SH) tag vector for establishing the MZF1-expressing Flp-In 293 T-REx cell lines. The pCMV3-Myc-CTCF and its empty control vector were from Sino Biological Inc. The *CTSB* firefly luciferase reporter pTAL-CTSB+126/+408-Luc, which contains the +126/+408 region of *CTSB* subcloned into a pTAL-Luc firefly luciferase vector, has been described previously (2). The pGL3 basic null *Renilla* luciferase vector (Promega) was used to assay basal luciferase expression. The plasmids encoding WT and dominant negative PAK4 (SRα-HA-PAK4 KM) were obtained from Audrey Minden (4).

Transfections

Plasmid transfections were performed with either the Lipofectamine 2000 Transfection Reagent (Life Technologies) or the FuGENE HD Transfection Reagent (Promega) in 6-well or 9-cm culture dishes. Transfections with siRNA (25nM) were performed in 6-well culture plates with the Lipofectamine RNAiMAX Transfection Reagent (Invitrogen).

Nuclear/cytosolic fractionation

For subcellular fractionation, the cells were lysed in lysis buffer (10mM HEPES (pH 7.9), 10mM KCl, 0.1mM EDTA, 0.1mM EGTA, 1mM DTT, 10% NP40) that was supplemented with 1X protease (Sigma-Aldrich) and phosphatase (Sigma-Aldrich) inhibitors. After 10 min at 4°C the cells were centrifuged. The cytosolic fraction was present in the supernatant and

the nuclear fraction was present in the pellet. The nuclear fraction was lysed for 1 h at 4°C in nuclear lysis buffer (20mM HEPES (pH 7.9), 0.4M NaCl, 1mM EDTA, 1mM EGTA, 1mM DTT, 10% glycerol) that was supplemented with 1X protease (Sigma-Aldrich) and phosphatase (Sigma-Aldrich) inhibitors for 1h.

Immunoblotting

Cells were harvested and lysed using RIPA lysis buffer (50 mM Tris (pH 8), 150 mM NaCl, 1% sodium deoxycholate, 0.1% SDS) that was supplemented with 1X protease and phosphatase inhibitors (Sigma-Aldrich) and N-ethylmaleimide (Sigma-Aldrich) to inhibit de-SUMOylation. The lysates were sonicated in a Bioruptor™ (Diagenode), the protein concentrations were measured using the Pierce™ BCA Protein Assay Kit (Life Technologies), and the lysate volume was adjusted to ensure equal loading. After transfer, the membranes were incubated overnight at 4°C on a rocking device with primary antibodies in 5% BSA. The primary antibodies (and dilutions) used for immunoblotting were: anti-MZF1 (1:1000); anti-phospho-S27-MZF1 (1:1000); anti-HA clone H-7 (1:1000); anti-SUMO1 (1:1000); anti-SUMO2/3 (1:1000); anti-phospho-Erk1/2 (1:1000); anti-HER-2 (1:1000); anti-PAK4 (1:1000); anti-α-tubulin (1:500); anti-histone H3 (FL-136) (1:1000); and anti-β-actin (1:1000). The membrane was then washed and incubated at room temperature (RT) for 1 h with HRP-conjugated secondary antibodies (1:2000; DAKO and Vector Laboratories) diluted in 5% milk. The immunoblots were developed using Clarity™ Western ECL (Bio-Rad Laboratories), and the luminescent image analyzer LAS-4000 mini (Fujifilm, GE Healthcare). When required, the membranes were stripped using the Restore™ PLUS Western Blot Stripping Buffer (Thermo Scientific). The immunoblots were quantified using the ImageJ software.

Immunoprecipitation

For immunoprecipitation, cells were harvested in lysis buffer (150 mM NaCl, 50 mM Tris (pH 7.4), 1 mM EDTA, 1% Np40 (IGEPAL), 0.25% Na-deoxycholate, 2mM $MgCl_2$, 0.1% SDS). For co-immunoprecipitation, the cells were incubated in EBC lysis buffer (150 mM NaCl, 50 mM Tris (pH7.4), 1 mM EDTA, 0.5% NP-40 (IGEPAL)). The lysis buffers were supplemented with 1x protease (Sigma-Aldrich) and phosphatase (Sigma-Aldrich) inhibitors and 25 U/ml Benzonase Nuclease (Sigma-Aldrich). For precipitation, 500-2500 µg of whole cell lysate were used with 2-4 µL of anti-HA clone H-7 antibody coupled to magnetic

Dynabeads A/G (Life Technologies) for 2 h at 4°C according to the manufacturer's instructions.

Immunofluorescence microscopy

Cells were seeded on glass coverslips 24 h prior to treatment. After fixation, the cells were covered with blocking buffer (PBS with 1%BSA, 0.3% Triton X-100) that was supplemented with 5% universal goat serum (DAKO) for a minimum of 20 min at RT. The cells were incubated for 1-2 h at RT with primary antibody diluted in blocking buffer, followed by washing in wash buffer (PBS with 0.25% BSA, 0.1% Triton X-100). For F-actin staining, Alexa Fluor 555 Phalloidin, diluted 1:100, in wash buffer that was supplemented with 5% universal goat serum, was added and incubated for 1 h at RT. The cells were then incubated for 1 h at RT with secondary antibody diluted in wash buffer that was supplemented with 5% universal goat serum (DAKO). The primary antibodies (and dilutions) used for immunofluorescence were: anti-LAMP2 (1:400), anti-MZF1 (1:500), anti-phospho-S27-MZF1 (1:500) and anti-α-tubulin (1:600). Secondary Alexa Flour 488 and 594 antibodies were obtained from Life Technologies and used at 1:1000 dilution. After washing, the coverslips were mounted with preheated ProLong Antifade Gold Mounting Medium (Invitrogen), which was left to solidify overnight at RT. Images were taken with the LSM700 Meta microscope (Carl Zeiss Inc).

Immunohistochemistry

Cells and tissue were subjected to antibody staining following hematoxylin and eosin (HE) staining. For antibody staining, sections were deparaffinized in TissueTeck, covered with 99% ethanol, covered with blocking peroxidase solution (12% H_2O_2 in 99% ethanol) for 15 min, and then rinsed in 96% and 70% ethanol. The sections were then subjected to protein retrieval by boiling them in 0.01 M sodium citrate buffer (MilliQ with 0.01 M Na citrate) for 10 min. After cooling, the sections were rinsed in TBS (50 mM Tris.HCl (pH 7.4), 150 mM NaCl) with 1% FBS. Primary antibody, diluted 1:600 in TBS with 10% FBS, was added and the samples were incubated for 1 h at RT in a humidified chamber. Following a wash in TBS, the sections were incubated with secondary anti-mouse EnVision (DAKO), for 30 min and then washed with TBS. Color was developed by incubating with enhanced DAB (DAKO). For HE staining, the sections were hydrated in a graded ethanol series, stained with HE, and dehydrated in a graded ethanol series. After drying at RT, the slides were coverslipped with a Tissue-Tek Film Coverslipper (Sakura, Finetek Europe B.V., Leiden, Netherlands). The

slides were scanned in the NanoZoomer digital slide scanner (Hammamatsu Photonics K.K., Japan).

Luciferase reporter assay

Cells were transfected with Lipofectamine 2000 Transfection Reagent (Life Technologies) or FuGENE HD Transfection Reagent (Promega) for 72 h with *CTSB* luciferase reporter pTAL-CTSB+126/+408-Luc, pGL3 *Renilla* null and pcDNA3.1 or Δt-3xHA plasmids that express the indicated proteins. Luciferase activities were measured using the Dual-Luciferase™ Reporter Assay System (Promega) and the Enspire Multilabel Plate Reader (PerkinElmer). All the firefly luciferase reporter activities were normalized to the activity of the internal control *Renilla* luciferase.

Kinase assay

The p95 ErbB2-MCF7 cells were transiently transfected with PAK4-HA and empty-HA plasmids. The cells were lysed in RIPA buffer and the lysates were cleared by centrifugation. HA-beads (HA clone H-7; Sigma-Aldrich) were used to purify HA-tagged proteins. The immunopurifed, HA-tagged proteins were washed extensively in RIPA buffer and further washed three times in PAK4 kinase buffer (20 mM HEPES (pH 7.6), 2 mM DTT, 10 mM $MgC1_2$, 1 mM EDTA, 1 mM EGTA, 1 mM PNPP, 10 mM β-glycerophosphate and 100 μM Na-vanadate). HA-tagged proteins were released from the beads using a HA-peptide, and the supernatant was immediately assayed for kinase activity for 20 min at 30°C using the S27-containing peptide (VKLEDSEEEGEA-Cys). More specifically, 2 μl peptide (1μg) and 10 μl of PAK4 were added to 10 μl of 2x PAK4 kinase buffer that was supplemented with 1 μM ATP and 10 μCi of [γ-^{32}P] ATP. The PAK4 inhibitor PF3758309 (500 nM) was used as a negative control. The reactions were warmed to 30°C briefly and spotted onto a nitrocellulose membrane. The samples were dried, and the incorporation of labeled phosphate into the peptide was quantified using a Typhoon 9410 imaging reader (Amersham Biosciences) after a 4 h exposure to a Phosphoimager plate (Fujifilm). Phosphorylation was quantitated using the FujiFilm MultiGauge ver. 3.2 software.

Cysteine cathepsin activity measurement

Cysteine cathepsin zFRase activity was measured as described previously (2). Briefly, the cells were washed carefully with PBS and lysed on ice for 15 min in a lysis buffer (25 mM

HEPES, 5 mM $MgCl_2$, 1 mM EGTA, 0.025% Triton X-100, 0.1 mM DTT and 0.5 mM Pefabloc) and centifuged at 3000x g for 5 min at 4°C. Then, 50 µl of the lysate were mixed with 50 µl of cathepsin reaction buffer (50 mM NaOAc (pH 6), 8 mM EDTA, 8 mM DTT, 0.5 mM Pefabloc, 20 mM z-FR-AFC). The zFRase activity readouts i.e., the V_{max} of the liberation of the AFC fluorochrome were obtained as the average slope from 17 reads of fluorescence at 400 nm excitation, 489 nm emission and 475 nm cutoff over a period of 15 min at 37°C, using the SpectraMax Gemini Fluorometer (Molecular Devices). The zFRase activity readouts were normalized to the respective protein concentrations.

Phosphor mass spectrometry

HA-MZF1 immunoprecipitates were resolved by SDS-PAGE and stained for 10 min with Coomassie Brilliant Blue R250 X in 10% glacial acetic acid, 45% methanol, 45% ddH2O, followed by overnight destaining in 10% glacial acetic acid, 10% methanol, 80% ddH2O. The bands corresponding to the molecular mass of MZF1 were excised and subjected to in-gel digestion with modified porcine trypsin (Promega) in 50 mM ammonium bicarbonate (AMBIC) overnight at 37°C. Peptides were eluted twice using 75% acetonitrile/formic acid (ACN/FA). Tryptic peptides were dissolved in 0.1% (v/v) FA before phosphor-enrichment, which was achieved using TiO_2 magnetic beads (GE Healthcare Life Sciences). Phosphopeptides were eluted twice using 28% ammonium hydroxide and non-phosphopeptides were kept and purified using C18 StageTips according to the manufacture's protocol (Proxeon). All the peptide samples were finally dissolved in 0.1% (v/v) FA and analyzed with the LTQ-Orbitrap instrument (ThermoFisher) coupled to the Eksigent nano-LC 2D system (Eksigent Technologies). For the phosphopeptide and the non-phosphopeptide analyses, 6 µL and 1 µL of peptide sample, respectively, were loaded and washed for 15 minutes onto a pre-column (Thermo Fisher Scientific) at a constant flow of 5 µl/min solvent B (0.1% FA in ACN). Phosphopeptides were loaded onto an RP analytical column that was packed in-house with C18 material ReproSil-Pur and separated using a 78-min gradient (3% to 90%) of solvent B at a flow rate of 300 nl/min. The gradient was followed by a 20-min column washing with 90% ACN, and 0.1% FA, and 15 min of re-equilibration with 3% solvent B. The HPLC system was coupled to an LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific), which was operated in data-dependent acquisition (DDA) mode. The ten most intense ions from the survey scan performed by the Orbitrap were fragmented by collision-induced dissociation (CID) in the LTQ (normalized collision energy 35, parent

mass selection window 0.5 Da, 30 ms activation time and minimum signal threshold for MS/MS scans set to 100). Unassigned charge states and singly charged ions were excluded from the fragmentation process.

Database analysis for phosphor mass spectrometry

The Mascot and MaxQuant programs were used to analyze the MS/MS spectra. Mascot was mainly used for peptide and phosphosite identification, whereas MaxQuant was used for both peptide and phosphosite identifications and relative quantification of the data. When using both software packages, the enzyme specificity was set to trypsin as the protease and two miss/cleavages were allowed. The instrument was set to ion-trap, and the peptide concentration and fragment mass tolerance were set to 10 ppm and 0.8 Da, respectively. Data were searched against the human database of Uniprot (84,236 sequences, 48,506,780 residues in Mascot; and 55539852 sequences, 17,689,561,164 residues in MaxQuant). Carbamidomethyl-cysteine was set as a fixed modification. Oxidation of methionine and phosphorylation of tyrosine, serine and threonine were set as variable modifications. For Mascot, the protein and pepetide identification FDR was set to <0.05. Identification of modified peptides in MaxQuant was accepted based on a minimum Andromeda score of 40 with a false discovery rate and site decoy fraction of 0.01. Label-free quantification of data based on spectral intensities was done using MaxQuant, applying a match between runs. The MaxQuant output was further analyzed with the Perseus software. Protein and peptide identifications were accepted based on a confidence score of at least 0.01, with a false discovery rate of 0.01. Phosphosite location probability was accepted based on a score of at least 0.75.

Mass spectrometry analysis for associating proteins

Affinity Purification

For each pull-down experiment, approximately 5×10^7 cells (5 x 15 cm dishes) in two biological replicates were induced with 2 μg/ml tetracycline (Sigma) for 24 h. Cells were washed with 0.1 mM $MgCl_2$, 0.1 mM $CaCl_2$ in PBS and harvested in 1 mM EDTA in PBS. Cells were pelleted and lysed on ice in the lysis buffer HENN (50 mM HEPES (pH 8.0), 5 mM EDTA, 150 mM NaCl, 50 mM NaF), supplemented with 0.5% NP-40, 1 mM DTT, 1.5 mM Na_3VO_4, 1 mM PMSF, and 1x protease inhibitors cocktail (Sigma). Clear cell lysates were loaded on spin columns (Bio-Rad, Laboratories) that contained 200 ml of Strep-Tactin

beads (IBA GmbH), washed four times with 1 ml of HENN buffer. Bound proteins were eluted twice with 300 µl of freshly prepared 0.5 mM D-biotin (Thermo Fisher Scientific) in HENN buffer. The samples were reduced with 5 mM (Tris(2-carboxyethyl) phosphine) (TCEP) and alkylated for cysteine bonds with 10 mM iodoacetamide. The sample proteins were digested with 2 g of trypsin (Promega) overnight at 37 °C, and tryptic peptides were purified using C-18 microspin columns (The Nest Group), according to manufacturer's instructions. The vacuum-dried samples were finally reconstituted in 30 µl of HPLC buffer A (0.1% formic acid, 1% acetonitrile).

Affinity mass spectrometry

The LC-MS/MS analysis was performed in a Q Exactive ESI-quadrupole-orbitrap mass spectrometer that was coupled to an EASY-nLC 1000 nanoflow LC (Thermo Fisher Scientific)), using the Xcalibur ver. 3.1.66.10 (Thermo Scientific) software. For each sample two technical replicates were loaded onto a C18-packed precolumn (Acclaim PepMap™100, 75 µm x 2 cm, 3-µm particle size, 100-Å pore size, Thermo Scientific) in buffer A. Peptides were transferred onwards to a C18-packed analytical column (Acclaim PepMap™RSLC 75 µm x 15 cm, 2 µm, 100 Å, Thermo Scientific) and separated using a 60-min linear gradient of 5% to 35% of buffer B (98% ACN and 0.1% FA in HPLC-grade water) at a flow rate of 300 nl/min. This was followed by a 5-min gradient of 35% to 80% buffer B, a 1-min gradient of 80% to 100% B, and a 9-min column wash with 100% B at a constant flow rate of 300 nl/min. The MS analysis was performed in data-dependent acquisition and in positive ion mode. MS spectra were acquired from m/z 200 to m/z 2000 with a resolution of 70,000 with a Full AGC target value of 1,000,000 ions, and a maximal injection time of 100 ms, in profile mode. The 10 most abundant ions for which the charge states were 2+ to 7+ were selected for subsequent fragmentation (higher energy collisional dissociation; HCD) and MS/MS spectra were acquired with a resolution of 17,500 with an AGC target value of 5000, a maximal injection time of 100 ms, and the lowest mass fixed at m/z 120, in centroid mode. The dynamic exclusion duration was 30 s.

Database analysis for affinity mass spectrometry

The SEQUEST search algorithm in the Proteome Discoverer software (Thermo Fisher Scientific) was used for peak extraction and protein identification with the human reference proteome in the UniProtKB database (www.uniprot.org, Uniprot_human 10_2016 with

20118 sequences). The allowed error tolerances were 15 ppm and 0.05 Da for the precursor and fragment ions, respectively. Database searches were limited to fully tryptic peptides allowing one missed cleavage (in peptide mapping of the semi-tryptic peptides, one missed cleavage was allowed), and carbamidomethyl +57 021 Da (C) of cysteine residue was set as fixed, and oxidation of methionine +15 995 Da (M) was accepted as a dynamic modification. For peptide identification, the FDR was set to <0.05. The high confidence protein-protein interactions were identified using stringent filtering against GFP control samples. The bait-normalized relative protein abundances (% to the MZF1) were calculated from the spectral counts. Statistically significant differences were assessed using the Student's *t*-test.

Molecular modeling and molecular dynamics simulations

The first conformer of the NMR ensemble (5) was used to model the disordered N-terminal tail in an extended conformation using Modeller version 9.15 (6). This produced the 1-128 MZF1 construct (MZF_{1-128}). Since MZF1 is known to be a dimer in the native state, the protein was modeled in its dimeric assembly. Disorder prediction was carried out using the consensus method implemented in MobiDB (7). The MZF_{1-128} construct was used as the starting structure for both unphosphorylated and phosphorylated (pS8 and pS111) variants of $MZF1_{1-128}$ for all-atom explicit solvent molecular dynamics (MD) simulations at 298 K for 400 ns with Gromacs 4.6 (8). MD simulations was carried out with Amber03ws force field (9). The protein was solvated in a dodecahedral box with a minimum distance between protein and box edges of 1.5 nm applying periodic boundary conditions for a total of 550 000 atoms for each system. His51 was simulated as the Nε2-H tautomer, according to NMR CD2-HD2 chemical shift (BMRB accession number: 6957) which is < 122 ppm (10). The same simulation protocol that was previously applied to the MZF1 SCAN domain was employed (11). Productive MD simulations were carried out in the canonical ensemble at 298 K for 400 ns. It was verified that the MZF1 (residues 35-128) folded SCAN domain remained stable during the simulation and no loss of secondary structure was observed. To model the complex of MZF_{1-128} with SUMO1-SUMO2 chains, the model of MZF_{1-128} described above was employed. Using both the Pymol and Modeller version 9.15 programs, the complexes of MZF_{1-128} with SUMO2 and SUMO1 were modeled, using as templates the first conformer of the NMR ensemble of SUMO2 (PDB entry 2N1W) and the X-ray structure of SUMO1 (PDB entry 2VRR) (12). In detail, the first conformer of the NMR ensemble of SUMO2 (PDB entry 2N1W) was used to model the isopeptide bond between K23 of MZF_{1-128} and G93 of

SUMO2. The X-ray structure of SUMO1 (12) was used to model the isopeptide bond between K11 of SUMO2 and G97 SUMO1.

Supplementary References:

1. Egeblad M, Mortensen OH, Jaattela M. Truncated ErbB2 receptor enhances ErbB1 signaling and induces reversible, ERK-independent loss of epithelial morphology. Int J Cancer. 2001;94(2):185-91.
2. Rafn B, Nielsen CF, Andersen SH, Szyniarowski P, Corcelle-Termeau E, Valo E, et al. ErbB2-driven breast cancer cell invasion depends on a complex signaling network activating myeloid zinc finger-1-dependent cathepsin B expression. Mol Cell. 2012;45(6):764-76.
3. Varjosalo M, Sacco R, Stukalov A, van Drogen A, Planyavsky M, Hauri S, et al. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. Nat Methods. 2013;10(4):307-14.
4. Qu J, Cammarano MS, Shi Q, Ha KC, de Lanerolle P, Minden A. Activated PAK4 regulates cell adhesion and anchorage-independent growth. Mol Cell Biol. 2001;21(10):3523-33.
5. Peterson FC, Hayes PL, Waltner JK, Heisner AK, Jensen DR, Sander TL, et al. Structure of the SCAN domain from the tumor suppressor protein MZF1. J Mol Biol. 2006;363(1):137-47.
6. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinformatics. 2016; 54:5 6 1-5 6 37.
7. Potenza E, Di Domenico T, Walsh I, Tosatto SC. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. Nucleic Acids Res. 2015;43(Database issue): D315-20.
8. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput. 2008;4(3):435-47.
9. Best RB, Zheng W, Mittal J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput. 2014;10(11):5113-24.
10. Sudmeier JL, Bradshaw EM, Haddad KE, Day RM, Thalhauser CJ, Bullock PA, et al. Identification of histidine tautomers in proteins by 2D 1H/13C(delta2) one-bond correlated NMR. J Am Chem Soc. 2003;125(28):8430-1.
11. Nygaard M, Terkelsen T, Vidas Olsen A, Sora V, Salamanca Viloria J, Rizza F, et al. The Mutational Landscape of the Oncogenic MZF1 SCAN Domain in Cancer. Front Mol Biosci. 2016; 3:78.
12. Knipscheer P, Flotho A, Klug H, Olsen JV, van Dijk WJ, Fish A, et al. Ubc9 sumoylation regulates SUMO target discrimination. Mol Cell. 2008;31(3):371-82.