SUPPORTING INFORMATION

METHODS

Massively parallel sequencing and shRNA screen data analysis pipeline

Genomic DNA extraction and purification from surviving MCF7 cells in the genome wide screen was carried out using a Gentra Puregene kit (Qiagen). shRNA sequences integrated into the genomic DNA of the screening cells were recovered by PCR amplification using the following primers:

p5+mir30: AATGATACGGCGACCACCGACTAAAGTAGCCCCTTGAATTC
p7+Loop: CAAGCAGAAGACGGCATACGATAGTGAAGCCACAGATGTA

PCR was performed using Amplitaq Gold polymerase. PCR products from multiple identical parallel reactions (sufficient to amplify DNA from 1000 cells per shRNA construct in each viral pool) were subsequently pooled, concentrated and then sequenced on an Illumina GAIIx platform according to the manufacturer's instructions.

Sequence image analysis and base calling were performed using the Genome Analyser pipeline v1.5 (Illumina). Short reads were aligned to shRNA library using a bespoke software package, shALIGN (Sims *et al*., manuscript in preparation). In alignment, we allowed up to two mismatches to the reference sequence. Statistical analysis of screen results was performed in R using a bespoke software package, shRNA-seq (Sims *et al*., manuscript in preparation). Briefly, read counts per shRNA in each sample were log2 transformed and then the ratio of reads in 4OHT-treated *vs.* vehicle treated samples was calculated. Each log ratio was then normalised to the average shRNA abundance in each pool using non-linear regression, to account for biases in shRNA abundance. Normalised log ratios were then re-scaled by the pool Median Absolute Deviation (MAD) to ensure comparable distributions between different pools. The resultant Drug Effect (DE) log ratios were then quantile normalised to allow comparison between biological replica screens. A similar approach was used to calculate the effect of each shRNA upon cell viability in the absence of 4OHT. Here we compared shRNA frequency data from vehicle-treated samples with shRNA frequency in the original plasmid pool used to generate virus.

Hit detection was performed using three parallel methods. In the first method, replicate DE scores for each shRNA were summarised using a regularised t-test (to penalise hairpins with high variance across replicates) and a Z-score threshold of >2 or <-2 was used to call hits. All hairpins with no predicted target, or with greater than one predicted target were removed from this analysis. As alternative methods, RSA and RIGER were also used as previously described (see Refs. 1, 2).

96 well plate method (see Ref. 3).
MCF7 cells were transfected with siRNA duplexes in replica plates and 48 hours later exposed to a range of 4OHT concentrations or drug vehicle. Five days later cell viability was determined using an assay measuring cellular ATP levels (Cell Titer Glo, Promega).

GFP competition assay (see Ref. 4)
Fluorescence Activated Cell Sorting (FACS) was used to track the survival advantage conferred by specific shRNAs in partially transduced MCF7 cells. Shifts in GFP percentage were monitored two days post-infection (t=0), and 5 days after treatment with 500 nM 4OHT or vehicle control using a BD Biosciences LSRII flow cytometer.

Immunoblotting
Western blotting was performed as described previously (see Ref. 5). Proteins were detected by using the following antibodies: anti-phospho p44/p42 MAPK (ERK1/2) (Thr202/Tyr204) (9101, Cell Signaling), anti-p44/p42 MAPK (ERK1/2) (9102, Cell Signaling), anti-Ezrin (3145 Cell Signaling), and anti-Actin (C-2, Santa Cruz Biotech). Activated Ras was assessed by immuno-precipitation using a Ras Activation Assay Kit (Millipore). In brief, activated GTP-Ras was isolated from cell lysates by the use of agarose beads conjugated to the Ras-binding domain of c-RAF. The amount of GTP-Ras was quantified by the western blotting of purified samples with a mouse monoclonal antibody recognizing all three isoforms of Ras.

Supplementary MATERIALS

Antibodies targeting the following epitopes were used: PTEN C-terminus (138G6, Cell Signalling), NF1 (NB100-418, Novus Biologicals), c-RAF (9422, Cell Signalling), RRAS2 (ab56859 Abcam), KRAS (F234, Santa Cruz Biotech), NAE1 (14863-1-AP, ProteinTech Group Inc.), UBA3 (E-22, Santa Cruz Biotech), NIPBL (NB100-93320, Novus Biologicals), RAD21 (L611, Cell Signalling), SMC3 (A300-060A, Bethyl Labs) and EDF1 (ab33588, abcam).

Supplementary Figure and Table LEGENDS

Fig. S1. Tamoxifen genome-wide RNA interference (RNAi) screen **(a)** High-throughput Screen (HTS) workflow. Schematic of the screen procedure employed to identify modifiers of tamoxifen response, orderly describing main steps and conditions involved. **(b)** Scatter Plot of normalised Drug Effect (DE) Z-scores per shRNA, from the tamoxifen genome-wide screen performed in triplicate. Hit selection thresholds ($Z>2$ and $Z<-2$) are indicated in red.

Fig. S2. Representative FACS profiles from a genome-wide screen replicate. Infectivity results shown for MCF7s three days after transduction with 6 shRNA pools (≈10K molecules each).

Fig. S3. Validation of the PTEN effect. **(a)** GFP FACS profiles from MCF7 cells infected with PTEN shRNAs. Increase in the GFP+ fraction after 4OHT treatment indicated a resistance effect. **(b)** Dose-response curves to anti-estrogen agents, 4OHT and ICI 182,780, for MCF7 cells transfected with either control or PTEN siRNAs. **(c)** RNAi reagents against PTEN effectively suppressed expression of their intended target when tested by western blotting.

Fig. S4. Western blots showing gene silencing for: NF1, NAE1, UBA3, NIPBL (NIPBL-specific band is annotated by an arrow), RAD21, SMC3, KRAS, RRAS2, RAF1 and EDF1.

Fig. S5. Low expression levels of EDF1 correlate with a favourable outcome in tamoxifen-treated breast cancer patients. The combined EDF1 effect among five studies is significant at $p<0.05$. Kaplan-Meier survival curve for the most significant study is also shown.

Fig. S6. Kaplan-Meier survival curve from GUYS77 dataset showing that tamoxifen treated patients with low NF1 expression (defined as lowest quartile expression) were at significantly higher risk of distant relapse, $p = 0.05$ (log-rank test).

Table S1. Run statistics for massively parallel sequencing. % PF represents the percentage of short DNA read clusters that passed the quality filter used on the Illumina GAII pipeline. % matched reads describes the number of short DNA reads that mapped to the shRNA library sequence.

Table S2. Tamoxifen sensitization and resistance-causing effects identified from the genome-wide functional screen using the intersection of Z score threshold, RIGER and RSA methods.

Table S3. RNAi molecules targeting validated hits: sequences for scoring shRNAs and siRNAs used for validation. Non-targeting shRNA sequences also listed.

Table S4. 4OHT SF_{60} values from Fig. 2.

Table S5. Gene annotation of validated hits (obtained from STRING). Resistance and sensitivity-causing hits, in blue or red respectively. EC numbers correspond to enzyme nomenclature from NC-IUBMB.

Table S6. Tumor characteristics, according to NF1 expression, where "low" is defined as lowest quartile expression in five independent clinical datasets.

Supplementary References

1. Subramanian A*, et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102(43):15545-15550.
2. Luo B*, et al.* (2008) Highly parallel identification of essential genes in cancer cells. *Proc Natl Acad Sci U S A* 105(51):20380-20385.
3. Iorns E*, et al.* (2008) Identification of CDK10 as an important determinant of resistance to endocrine therapy for breast cancer. *Cancer Cell* 13(2):91-104.
4. Burgess DJ*, et al.* (2008) Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc Natl Acad Sci U S A* 105(26):9053-9058.
5. Mendes-Pereira AM*, et al.* (2009) Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO Mol Med* 1(6-7):315-322.

Fig. S1

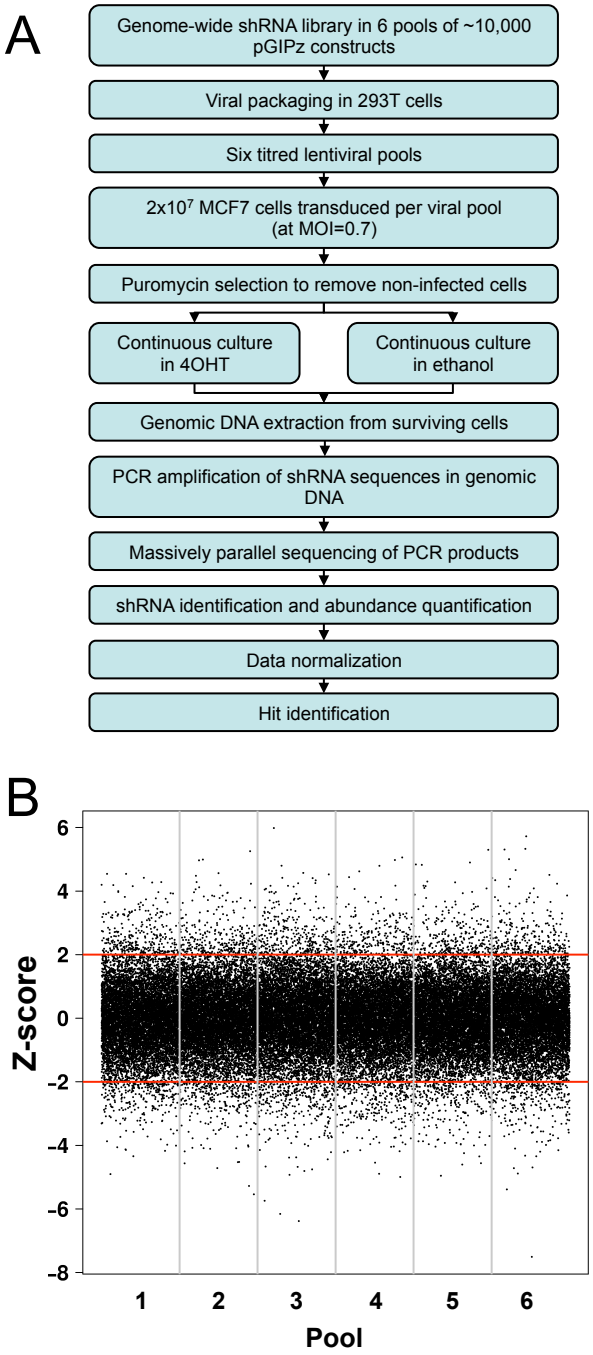
A
Genome-wide shRNA library in 6 pools of ~10,000 pGIPz constructs
Viral packaging in 293T cells
Six titred lentiviral pools
2x10^7 MCF7 cells transduced per viral pool (at MOI=0.7)
Puromycin selection to remove non-infected cells
Continuous culture in 4OHT
Continuous culture in ethanol
Genomic DNA extraction from surviving cells
PCR amplification of shRNA sequences in genomic DNA
Massively parallel sequencing of PCR products
shRNA identification and abundance quantification
Data normalization
Hit identification
B
Z-score
6
4
2
0
-2
-4
-6
-8
1
2
3
4
5
6
Pool

Fig. S2

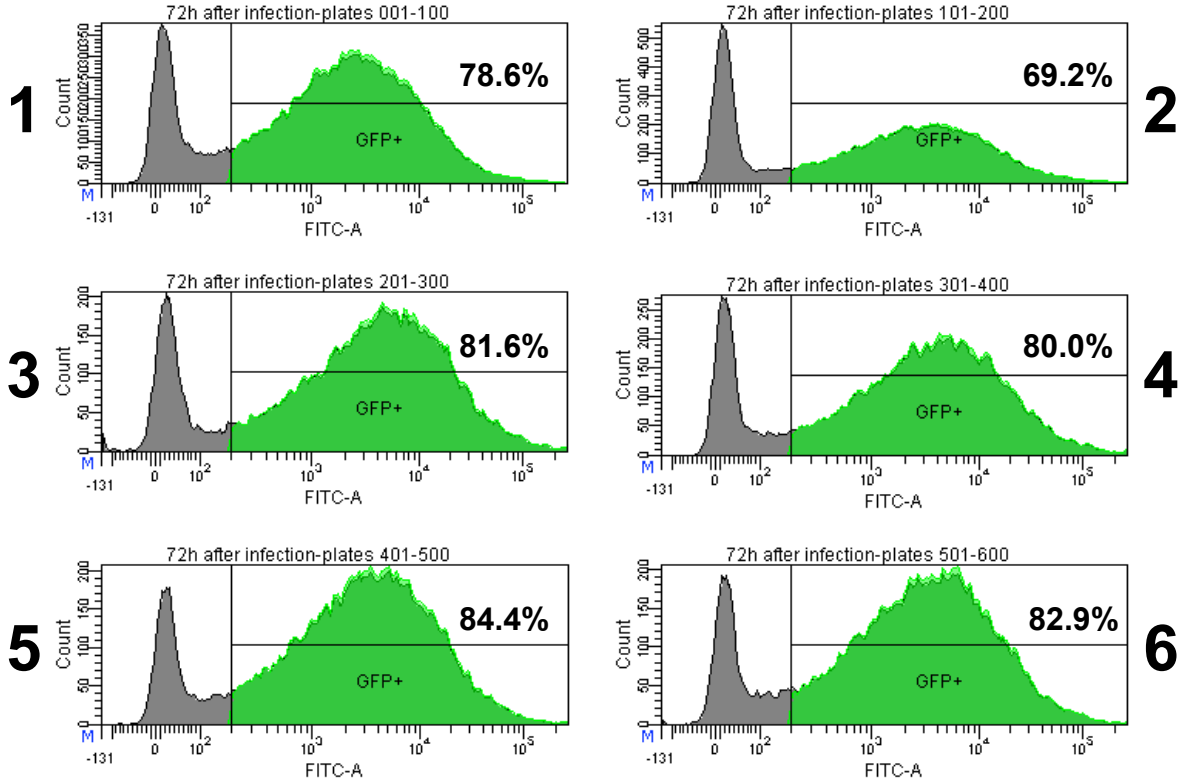
1
72h after infection-plates 001-100
78.6%
GFP+
Count
FITC-A
2
72h after infection-plates 101-200
69.2%
GFP+
Count
FITC-A
3
72h after infection-plates 201-300
81.6%
GFP+
Count
FITC-A
4
72h after infection-plates 301-400
80.0%
GFP+
Count
FITC-A
5
72h after infection-plates 401-500
84.4%
GFP+
Count
FITC-A
6
72h after infection-plates 501-600
82.9%
GFP+
Count
FITC-A

Fig. S3

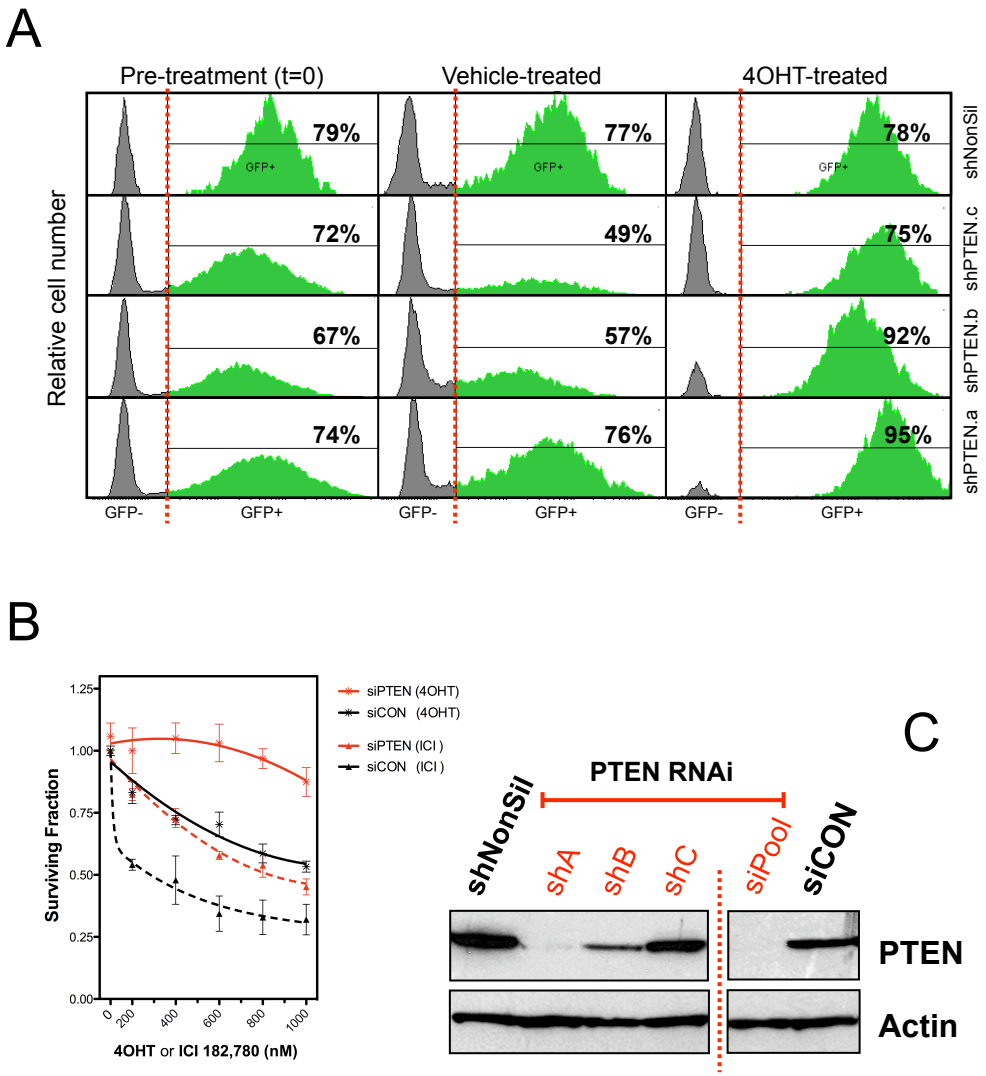
A
Pre-treatment (t=0)
Vehicle-treated
4OHT-treated
Relative cell number
79%
77%
78%
72%
49%
75%
67%
57%
92%
74%
76%
95%
GFP+
GFP-
shNonSil
shPTEN.c
shPTEN.b
shPTEN.a
B
Surviving Fraction
4OHT or ICI 182,780 (nM)
siPTEN (4OHT)
siCON (4OHT)
siPTEN (ICI)
siCON (ICI)
C
PTEN RNAi
shNonSil
shA
shB
shC
siPool
siCON
PTEN
Actin

Fig. S4

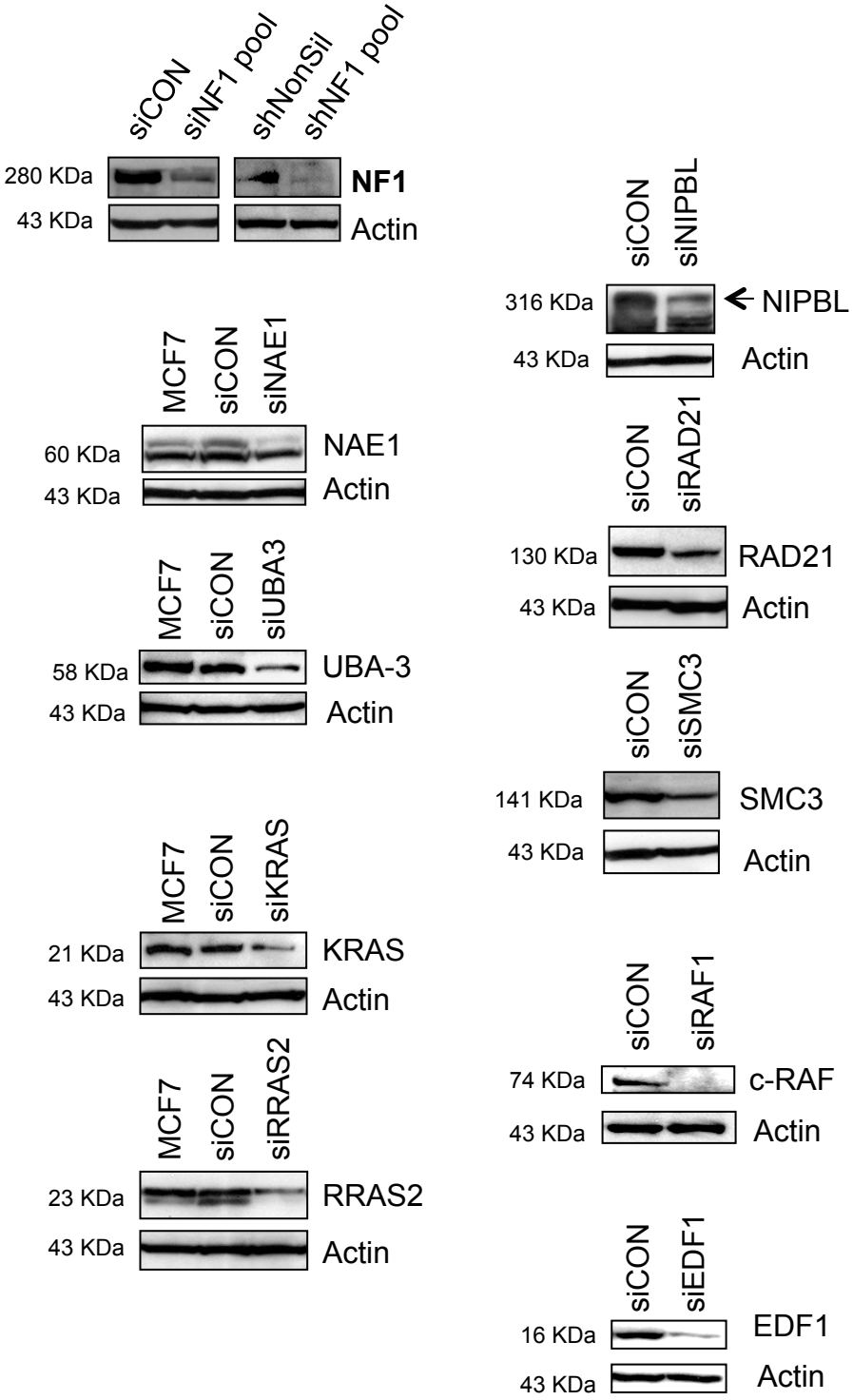
siCON
siNF1 pool
shNonSil
shNF1 pool
280 KDa
NF1
43 KDa
Actin
MCF7
siCON
siNAE1
60 KDa
NAE1
43 KDa
Actin
MCF7
siCON
siUBA3
58 KDa
UBA-3
43 KDa
Actin
MCF7
siCON
siKRAS
21 KDa
KRAS
43 KDa
Actin
MCF7
siCON
siRRAS2
23 KDa
RRAS2
43 KDa
Actin
siCON
siNIPBL
316 KDa
NIPBL
43 KDa
Actin
siCON
siRAD21
130 KDa
RAD21
43 KDa
Actin
siCON
siSMC3
141 KDa
SMC3
43 KDa
Actin
siCON
siRAF1
74 KDa
c-RAF
43 KDa
Actin
siCON
siEDF1
16 KDa
EDF1
43 KDa
Actin

Fig. S5

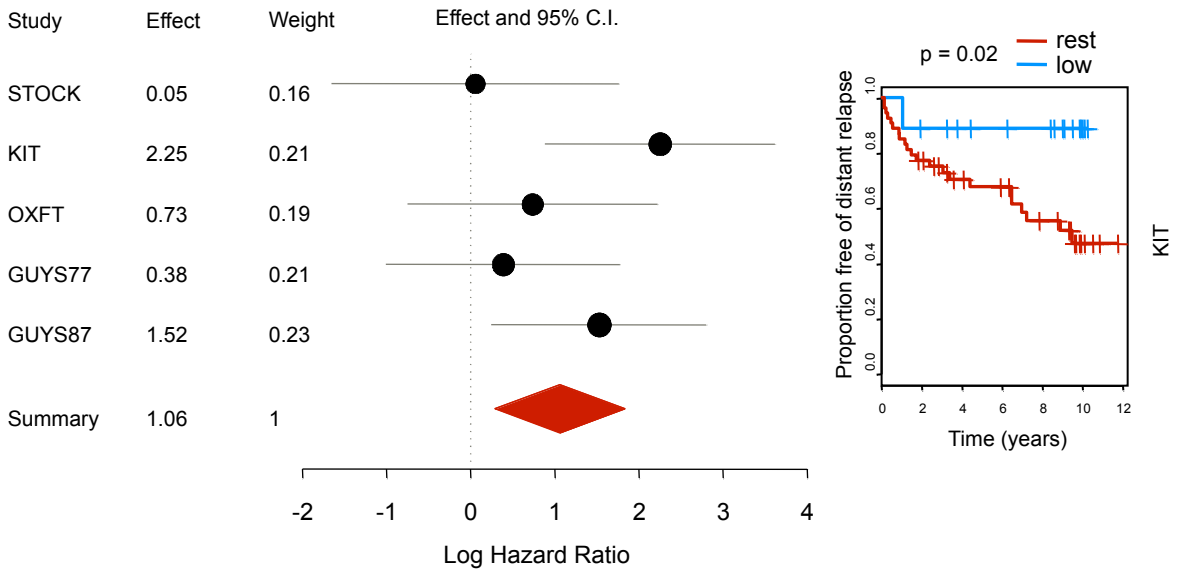
Study
Effect
Weight
Effect and 95% C.I.
STOCK 0.05 0.16
KIT 2.25 0.21
OXFT 0.73 0.19
GUYS77 0.38 0.21
GUYS87 1.52 0.23
Summary 1.06 1
-2 -1 0 1 2 3 4
Log Hazard Ratio
p = 0.02
rest
low
Proportion free of distant relapse
0.0 0.2 0.4 0.6 0.8 1.0
0 2 4 6 8 10 12
Time (years)
KIT

Fig. S6

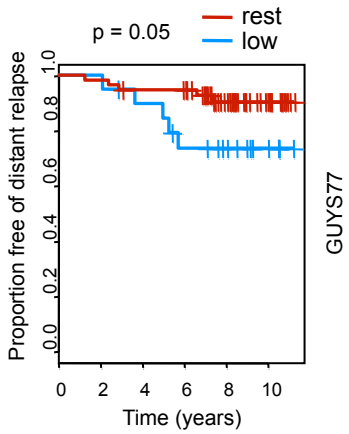
p = 0.05
rest
low
Proportion free of distant relapse
0.0 0.2 0.4 0.6 0.8 1.0
0 2 4 6 8 10
Time (years)
GUYS77

Table S1

Run	Lane	Pool	Treatment	Replicate	% PF Clusters	% Mapped
Run27	2	1	Vehicle	1	80%	97%
Run27	3	2	Vehicle	1	80%	98%
Run27	4	3	Vehicle	1	81%	99%
Run27	5	4	Vehicle	1	89%	99%
Run27	6	5	Vehicle	1	88%	99%
Run27	7	6	Vehicle	1	88%	99%
Run28	2	1	Tamoxifen	1	90%	96%
Run28	3	2	Tamoxifen	1	90%	97%
Run28	4	3	Tamoxifen	1	90%	98%
Run28	5	4	Tamoxifen	1	91%	97%
Run28	6	5	Tamoxifen	1	90%	98%
Run28	7	6	Tamoxifen	1	91%	98%
Run38	1	1	Vehicle	2	85%	91%
Run38	2	2	Vehicle	2	87%	94%
Run38	3	3	Vehicle	2	87%	95%
Run38	4	4	Vehicle	2	87%	94%
Run38	5	5	Vehicle	2	86%	94%
Run38	6	6	Vehicle	2	86%	93%
Run39	1	1	Tamoxifen	2	87%	96%
Run39	2	2	Tamoxifen	2	89%	97%
Run39	3	3	Tamoxifen	2	88%	98%
Run39	4	4	Tamoxifen	2	88%	97%
Run39	5	5	Tamoxifen	2	88%	97%
Run39	6	6	Tamoxifen	2	89%	98%
Run67	1	1	Vehicle	3	93%	97%
Run67	2	2	Vehicle	3	93%	93%
Run67	3	3	Vehicle	3	91%	98%
Run67	5	4	Vehicle	3	87%	96%
Run67	6	5	Vehicle	3	93%	99%
Run67	7	6	Vehicle	3	93%	99%
Run68	1	1	Tamoxifen	3	92%	97%
Run68	2	2	Tamoxifen	3	90%	98%
Run68	3	3	Tamoxifen	3	93%	99%
Run68	5	4	Tamoxifen	3	92%	99%
Run68	6	5	Tamoxifen	3	93%	99%
Run68	7	6	Tamoxifen	3	92%	99%
				AVERAGE	89%	97%

Table S2

Sensitisation effects

Gene Symbols

ABCA13	CBR3	HES2	MAX	RP11-292E2.1	TMPRSS2
AC079953.2	CCDC11	HIPK4	MDM2	RP11-3B7.1	TMTC3
AC087521.2	CD163	HMGA1	MED13L	RP4-604K5.1	TPM4
ARHGAP18	CGA	HOOK1	MPST	RPS27L	TRAK2
ARHGAP28	CLCN5	HOXC13	MYST3	SAP130	TTC29
ARPC2	COG2	HPSE2	NCKAP1	SCUBE1	UBAP2L
ATBF1	CYTSB	HSD11B2	NDFIP1	SEZ6L	UBXN10
ATXN2L	DACT3	IFNAR2	NMNAT2	SFPQ	UCK1
BCL9L	DDX50	IGF1R	NRP2	SFRS11	VPS13C
C11orf1	DGKI	IL13	NSFL1C	SH3YL1	WNT8B
C12orf70	DNAJC8	ILF3	NUDT4	SLC11A2	XG
C14orf23	EFCAB2	ING5	OR4D6	SLC26A6	XPA
C15orf24	ERH	JAK2	OR4K13	SLC6A5	ZFP62
C15orf55	ESR1	KCNH6	OVCH2	SPINLW1	ZMIZ1
C17orf75	FAM26E	KCNJ9	P2RY13	ST6GALNAC1	ZNF391
C19orf63	GLA	KDM5C	PBRM1	STAM2	ZSWIM4
C1orf116	GLS	KHDRBS1	PCBD2	STXBP5L	
C3orf67	GOLM1	KRAS	PLXNA4	SYNJ2	
C4orf29	GPR88	LAMA3	PPP1R15B	TAS2R1	
CACNA1C	GULP1	LUC7L	PRUNE2	TEKT3	
CAPZA2	HECTD1	MAP2	PTS	THOC2	

Resistance-causing effects

Gene Symbols

AC069234.1	CCDC42	GPR15	NDUFB2	RBMS3	TAF15
AC092143.3	CCNC	GRM8	NF1	RELT	TCEAL3
ACAD8	CDKN2B	HERC1	NFATC2IP	REV1	TFB2M
AFTPH	CHD4	HOOK2	NFE2L3	RGS16	TIGD7
AK1	CLDN11	HPS3	NPAS2	RMND5A	TMEM133
AMBP	CLPP	IL26	NRN1	RNASE8	TMEM155
ANKRD12	CPNE4	ITGAV	NSD1	RP11-40M23.1	TMEM48
AP002448.1	CRADD	KANK1	OR4B1	S100A3	TNFRSF11A
AP003774.1	CREB5	KEAP1	ORC5L	SFRS8	TOMM40L
APOD	CUL3	LARP4	OTOL1	SLC38A10	TRAIP
ARMC4	CYP7B1	MAP2K1	PABPC5	SLC4A7	TRIM56
BAP1	CYTL1	MAP4	PENK	SLFN12	TRIP12
BCOR	DNASE1L3	MAPK11	PLD2	SMC3	TSPAN13
BECN1	EHHADH	MAPKAPK5	PLEKHC1	SMCR7L	TYRP1
C12orf61	ERG	MARCKSL1	PLEKHG7	SOX12	UBE2W
C1orf27	FAM118A	MBD6	PPIL6	SPAM1	UBQLN1
C1orf61	FAM63B	MED1	PTEN	SPDYE2	UBR1
C6orf140	FBXO7	MPI	PTPRF	SRFBP1	VPS41
C6orf64	FUBP1	MRPS14	PTPRJ	STAG2	ZBTB25
CAND1	GBP2	MTM1	PYCARD	STARD3	ZNF256
CASP8AP2	GFRA1	NBEAL1	RAD21	STEAP3	ZNF627
CBLN2	GPN3	NCOA6	RASA1	SUCLG2	

Table S3

Gene Symbol	shRNA Z-score	shRNA 19mer Target Sequence	shRNA Clone ID	siRNA Pool Sequences
BAP1	3.28 2.63	CAACTCTGCCCTTAGGTAT GAGTTCATCTGCACCTTTA	V2LHS_246612 V2LHS_41473	GAGUUCAUCUGCACCUUUA CCACAAGUCUCAAGAGUCA GAUGAUACGUCCGUGAUUG GAGCAAAGGAUAUGCGAUU
CLPP	2.46 2.33	CTCAAGAAGCAGCTCTATA GAAGCAGCTCTATAACATC	V2LHS_71906 V2LHS_71903	GAAGGAGCCUGUAGAAGCA GAGAGGGACCGCUACAUGA AGAAGCAGCUCUAUAACAU GGGCCAAGCCACAGACAUU
GPRC5D	2.05	ACAATTCTGTGCATTGCTA	V2LHS_30673	GGUAUGAUGUUUGUGAAUA GCUCAUGCCUCCAAUCUAG GGAAGGCUCAUCUUUAUCA GGAUGUAGCAUUAACUUCA
NAE1	3.97 2.05	GAAACAATTTCTTCCTTCA GTATTGGTTCGTTTACAAT	V2LHS_47901 V2LHS_47903	GAUGAUCGCUGCAUAAAUA GUAAUCAUGUUGCCAAAUU UGAAACAAAUGGACGAAUA GGGUUGUGCUUUAGUCUGU
NF1	4.14 2.80 2.30 2.07	GACTTCAAATACGGACCAA CTCAATATCGCATTACTTA CAGATACACCTGTCAGCAA CACCGAGTCTTACATTTAA	V2LHS_190255 V2LHS_260806 V2LHS_189526 V2LHS_76029	GGAAUAAGAUGGUAGAAUA GAUAGAAGCUACAGUAAUA CAACAAAGCUAAUCCUUAA CGCAGUGAAGUUGAAGAUG
NIPBL	2.03	GAGCCTTTCTTATTTCTTT	V2LHS_258577	CUGAUAAACUAGAACGAAA GGGAAUAUGAAGAGCGUGA UGGCUGAAAUUGAGCCAAU GAAAAUGAGUCAAGCGACA
NSD1	5.33 4.42 3.27 2.89	GGTTAAATGTTTGTGATAA GGATTCAAGTGACATAGAA GTGAATTTGTGAATGAGTA CGAAGTGATTCCATTAGTA	V2LHS_81480 V2LHS_81476 V2LHS_238790 V2LHS_239055	GGACGAGAAUUCUUUGAUU GAACUGCAGUGGCUUCUUG GCAAGUGGCUUUUGGAAUA GAAUGGAGAUACCCGUGUA
PTEN	4.90 3.73 3.73 3.44 2.91	CTATATTTATCCAAACATT GGCGCTATGTGTATTATTA GTGAAGATATATTCCTCCA CAATATGATGTGTACAGGA GAGACAGACTGATGTGTAT	V2LHS_92314 V2LHS_92317 V2LHS_231477 V2LHS_119551 V2LHS_92319	GUGAAGAUCUUGACCAAUG CGCAGAUAAUGACAAGGAG GUAUAGAGCGUGCAGAUAA GAUCAGCAUACACAAAUUA
RAD21	3.37 3.44 2.91	CTTATAATGCCATTACTTT CATTGGAGCCTATTGATAT GACACAGAACCCTTTGAGA	V2LHS_57223 V2LHS_57226 V2LHS_57224	GGAAGAAGCAUUUGCAUUG GAACAGAGCACCAGCAAUC GAGCCCAACUUAGUGAUUA GGGAGUAGUUCGAAUCUAU
RARG	3.16	CCCTACATGTTCCCAAGGA	V2LHS_239268	GAAAUGACCGGAACAAGAA UAGAAGAGCUCAUCACCAA CAAGGAAGCUGUGCGAAAU UCAGUGAGCUGGCUACCAA
SMC3	3.70 2.35	CAGTGCAACACAGAATTAA CAGAAATATTGAAAGGATT	V2LHS_249997 V2LHS_69325	CCCACACAUGGUUAAUUGG CGGGCAGAAAUGGAUCUGG ACCUCAAACAGUUGUUUCG AGCAGAAAUAUUGAAAGGA
UBA3	3.61	ATGCTGATATCTCTTCTAA	V2LHS_46794	CCCACAGACUGUACUAUUC CAACGACACUUUCUAUCGA ACAGAACACUGUAUUGAGU CGCAUACAUUCCCUUGAAU

Gene Symbol	shRNA Z-score	shRNA 19mer Target Sequence	shRNA Clone ID	siRNA Pool Sequences
C10orf72	-2.87	CTTGCGATGATATATTTAA	V2LHS_35944	CAGGAGGCCUUGAUGGUGA AGGAACAAGUGGACGGCCU GAUCAAGGGCAUUACGUCU GCUCCUCAGCCACGGAAAU
C15orf55	-2.34 -2.34	GACCGGATATGAGCATGAA CACTGAATGTTCATTCTTA	V2LHS_55320 V2LHS_55319	GGAGCUGCCUGUACAAAUA GCAGAAAGGUUCAUGGAGU GGUGACCGCUCCAAAAUUU CAAGGACGUUUAUGAGAAC
EDF1	-3.25	ATGTGGAGACTTCCAAGAA	V2LHS_23260	CCACGAAAAUCAAUGAGAA GGCCAGAACAAACAACAUU AAUCCAAGCAGGCUAUCUU GCAAGGGGCUUACGCAGAA
ESR1	-3.13 -2.20 -2.12 -2.02	CCTCTATTATGGCACTTCA CTATCAATGTAGGTTGCAA GTGACCTTATTGTCTGTAA CTATCAATGTAGGTTGCAA	V2LHS_239572 V2LHS_239449 V2LHS_83798 V2LHS_239269	CGCUAGUUAUGUGAAAGGC AGGACUUACUGAUAAUUUA ACCUAUCAAUGUAGGUUGC AGGUGACCUUAUUGUCUGU
ING5	-4.44 -2.77	CAAAGCCTGTTCGCACAGA GATGTGAATTTGTTGTGAA	V2LHS_138565 V2LHS_138566	GGAAUACAGUGACGACAAA CCUACGAGAUGGUGGAUAA CCACGAAACCCAAAGGAAA AAGCAGAUCUGAAGGACAA
KRAS	-2.20	CCTATGGTCCTAGTAGGAA	V2LHS_169384	CGAAUAUGAUCCAACAAUA UAAGGACUCUGAAGAUGUA GACAAAGUGUGUAAUUAUG GCUCAGGACUUAGCAAGAA
NOC3L	-3.64	CTCTGGATTTCACGAAATA	V2LHS_256682	GGGCAGAAAUUACGAAGUU CAACAAAGCACUCUCAAGA CAAUGAAGGUGUUGAGAUU GGAACAAGAUCCUGAUGUG
PPP1R15B	-2.78	CTCTAAACATTCAACGCAT	V2LHS_177770	GAGAUAACCCAACACAGUU CCGAAUAAGUGUAGUUGAU UACCUGGACUGCUUUCCUA CCGACCAGAUGGCUAGAUU
RRAS2	-2.61	GTCATTTCCAGACAAATTT	V2LHS_19741	CAGUUAGCACGGCAGCUUA CCGUGAUGAGUUCCCAAUG CCCAGACAAAUUUAAAGGA CUGUCAGCCUUUGUUAACC
TMPRSS2	-2.83	CCGGCAATGTCGATATCTA	V2LHS_5610	GCAAUGUCGAUAUCUAUAA GCUAUUGGACCUUACUAUG ACGGGAAUGUGAUGGUAUU CGGACUGGAUUUAUCGACA
TPM4	-3.20	CCAACTTCATTTCCATACT	V2LHS_171493	CCAAGCACAUUGCGGAAGA GCGGAGGUGUCUGAACUAA GAACGUGGGCUUACAUCAG GGAUCAGACACUAAACGAA

Gene Symbol	shRNA Z-score	shRNA 19mer Target Sequence	shRNA Clone ID
non-targeting	0.49	CAAACGAGACCTATGGAAT	V2LHS_162428
	0.47	AGGACAAGGGTTGTACTTA	V2LHS_211724
	0.40	CCGACATAGTTAATTCATT	V2LHS_148920
	0.37	CGATTACAATGCAAGGTAT	V2LHS_125866
	-0.03	CCAGGGCAAAGATAAACTA	V2LHS_235246
	-0.07	CTGACATAATCTCCAAACA	V2LHS_243633
	-0.39	GAGAACATGAGCATGTGAA	V2LHS_189022
	-0.43	GCAGCAGAATTTGGAATTA	V2LHS_67354

Table S4

siRNA Target	4OHT SF_{60} (nM)
ESR1	120
C10orf72	131
KRAS	197
PPP1R15B	279
ING5	280
EDF1	304
TMPRSS2	319
NOC3L	387
TPM4	389
C15orf55	426
RRAS2	472
siCON	805
NAE1	990
NIPBL	1009
SMC3	1019
UBA3	1068
GPRC5D	1134
BAP1	1139
CLPP	1158
RAD21	1222
NF1	1292
PTEN	1341
NSD1	*>1500*
RARG	*>1500*

Table S5

GENE Symbol	DESCRIPTION
BAP1	Ubiquitin carboxyl-terminal hydrolase BAP1 (EC 3.4.19.12) (BRCA1- associated protein 1) (Cerebral protein 6); Deubiquitinating enzyme which may be involved in BRCA1 signal transduction pathway (729 aa)
C10orf72	Uncharacterized protein C10orf72 precursor (320 aa)
C15orf55	Chromosome 15 open reading frame 55. Alias: NUT, nuclear protein in testis (1132 aa)
CLPP	Putative ATP-dependent Clp protease proteolytic subunit, mitochondrial precursor (EC 3.4.21.92) (Endopeptidase Clp); Clp cleaves peptides in various proteins in a process that requires ATP hydrolysis. Clp may be responsible for a fairly general and central housekeeping function rather than for the degradation of specific substrates (277 aa)
EDF1	Endothelial differentiation-related factor 1 (EDF-1) (Multiprotein- bridging factor 1) (MBF1); Transcriptional coactivator stimulating NR5A1 and ligand-dependent NR1H3/LXRA and PPARG transcriptional activities. Enhances the DNA-binding activity of ATF1, ATF2, CREB1 and NR5A1. Regulates nitric oxid synthase activity probably by sequestering calmodulin in the cytoplasm. May function in endothelial cells differentiation, hormone-induced cardiomyocytes hypertrophy and lipid metabolism (148 aa)
GPRC5D	G-protein coupled receptor family C group 5 member D (345 aa)
ING5	Inhibitor of growth protein 5 (p28ING5); May play a role in the regulation of transcription through epigenetic modification of chromatin (By similarity) (241 aa)
KRAS	GTPase KRas precursor (K-Ras 2) (Ki-Ras) (c-K-ras) (c-Ki-ras); Ras proteins bind GDP/GTP and possess intrinsic GTPase activity (189 aa)
NAE1 (APPBP1)	NEDD8-activating enzyme E1 regulatory subunit (Amyloid protein-binding protein 1) (Amyloid beta precursor protein-binding protein 1, 59 kDa) (APP-BP1) (Protooncogene protein 1) (HPP1); Regulatory subunit of the dimeric UBA3-NAE1 E1 enzyme. E1 activates NEDD8 by first adenylating its C-terminal glycine residue with ATP, thereafter linking this residue to the side chain of the catalytic cysteine, yielding a NEDD8-UBA3 thioester and free AMP. E1 finally transfers NEDD8 to the catalytic cysteine of UBE2M. Necessary for cell cycle progression through the S-M checkpoint. (534 aa)
NF1	Neurofibromin (Neurofibromatosis-related protein NF-1). Stimulates the GTPase activity of Ras. NF1 shows greater affinity for Ras GAP, but lower specific activity. May be a regulator of Ras activity (2839 aa)
NIPBL	Nipped-B-like protein (Delangin) (SCC2 homolog); Probably plays a structural role in chromatin. Involved in sister chromatid cohesion, possibly by interacting with the cohesin complex (By similarity) (2804 aa)
NOC3L	Nucleolar complex protein 3 homolog (NOC3 protein homolog) (NOC3-like protein) (Nucleolar complex-associated protein 3-like protein) (Factor for adipocyte differentiation 24); May be required for adipogenesis (By similarity) (800 aa). Formerly “C10orf117”
NSD1	Histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific (EC 2.1.1.43) (H3-K36-HMTase) (H4-K20-HMTase) (Nuclear receptor-binding SET domain-containing protein 1) (NR-binding SET domain-containing protein) (Androgen receptor-associated co; Histone methyltransferase. Preferentially methylates 'Lys-36' of histone H3 and 'Lys-20' of histone H4 (in vitro) (By similarity). Transcriptional intermediary factor capable of both negatively or positively influencing transcription, depending on the cellular context (2696 aa)
PPP1R15B	Protein phosphatase 1, regulatory subunit 15B; Maintains low levels of EIF2S1 phosphorylation in unstressed cells by promoting its dephosphorylation by PP1 (By similarity) (713 aa)
PTEN	Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase and dual- specificity protein phosphatase PTEN (EC 3.1.3.67) (EC 3.1.3.16) (EC 3.1.3.48) (Phosphatase and tensin homolog) (Mutated in multiple advanced cancers 1); Tumor suppressor. Acts as a dual-specificity protein phosphatase, dephosphorylating tyrosine-, serine- and threonine- phosphorylated proteins. Also acts as a lipid phosphatase, removing the phosphate in the D3 position of the inositol ring from phosphatidylinositol 3,4,5-trisphosphate, phosphatidylinositol 3,4-diphosphate (403 aa)
RAD21	Double-strand-break repair protein rad21 homolog (hHR21) (Nuclear matrix protein 1) (NXP-1) (SCC1 homolog); Cleavable component of the cohesin complex, involved in chromosome cohesion during cell cycle, in DNA repair, and in apoptosis. The cohesin complex is required for the cohesion of sister chromatids after DNA replication. The cohesin complex apparently forms a large proteinaceous ring within which sister chromatids can be trapped. At metaphase-anaphase transition, this protein is cleaved by separase/ESPL1 and dissociates from chromatin, allowing sister chromatids to segregate. (631 aa)
RARG	Retinoic acid receptor gamma-1 (RAR-gamma-1); This is a receptor for retinoic acid. This metabolite has profound effects on vertebrate development. Retinoic acid is a morphogen and is a powerful teratogen. This receptor controls cell function by directly regulating gene expression (454 aa)
RRAS2	Ras-related protein R-Ras2 precursor (Ras-like protein TC21) (Teratocarcinoma oncogene); It is a plasma membrane-associated GTP-binding protein with GTPase activity. Might transduce growth inhibitory signals across the cell membrane, exerting its effect through an effector shared with the Ras proteins (204 aa)
SMC3	Structural maintenance of chromosomes protein 3 (Chondroitin sulfate proteoglycan 6) (Chromosome-associated polypeptide) (hCAP) (Bamacan) (Basement membrane-associated chondroitin proteoglycan); Involved in chromosome cohesion during cell cycle and in DNA repair. Central component of cohesin complex. The cohesin complex is required for the cohesion of sister chromatids after DNA replication. The cohesin complex apparently forms a large proteinaceous ring within which sister chromatids can be trapped. At anaphase, the complex is cleaved and dissociates from chromatin. (1217 aa)
TMPRSS2	Transmembrane protease, serine 2 precursor (EC 3.4.21.-) (Serine protease 10) [Contains- Transmembrane protease, serine 2 non-catalytic chain; Transmembrane protease, serine 2 catalytic chain] (492 aa)
TPM4	Tropomyosin alpha-4 chain (Tropomyosin-4) (TM30p1); Binds to actin filaments in muscle and non-muscle cells. Plays a central role, in association with the troponin complex, in the calcium dependent regulation of vertebrate striated muscle contraction. Smooth muscle contraction is regulated by interaction with caldesmon. In non-muscle cells is implicated in stabilizing cytoskeleton actin filaments (284 aa)
UBA3 (UBE1C)	NEDD8-activating enzyme E1 catalytic subunit (EC 6.3.2.-) (Ubiquitin- activating enzyme 3) (NEDD8-activating enzyme E1C) (Ubiquitin- activating enzyme E1C); Catalytic subunit of the dimeric UBA3-NAE1 E1 enzyme. E1 activates NEDD8 by first adenylating its C-terminal glycine residue with ATP, thereafter linking this residue to the side chain of the catalytic cysteine, yielding a NEDD8-UBA3 thioester and free AMP. E1 finally transfers NEDD8 to the catalytic cysteine of UBE2M. Down-regulates steroid receptor activity. Necessary for cell cycle progression (463 aa)

Table S6

Study	test	low	Rest	p-value	Summary
STOCK	N	22	65		
	Mean NF1 expression	-0.659	0.199		
	Median age	68	62	0.374	NS
	Median Tumour size (cm)	2	1.95	0.051	NS
	Grade 1 (%)	5.0	33.3	0.028	*
	Grade 2 (%)	40.0	43.3	0.794	NS
	Grade 3 (%)	55.0	23.3	0.018	*
	Node positive (%)	65.0	69.8	0.897	NS
	PR positive (%)	81.8	83.1	0.893	NS
KIT	N	18	54		
	Mean NF1 expression	-0.643	0.157		
	Median age	68.5	65	0.261	NS
	Median Tumour size (cm)	2.51	2.4	0.131	NS
	Grade 1 (%)	5.6	21.6	0.238	NS
	Grade 2 (%)	44.4	68.6	0.124	NS
	Grade 3 (%)	50.0	9.8	0.001	**
	Node positive (%)	82.4	65.4	0.309	NS
	PR positive (%)	83.3	96.2	0.197	NS
OXFT	N	28	81		
	Mean NF1 expression	-0.548	0.178		
	Median age	66.5	64	0.331	NS
	Median Tumour size (cm)	2.25	2.3	0.260	NS
	Grade 1 (%)	4.4	29.9	0.027	*
	Grade 2 (%)	60.9	56.7	0.918	NS
	Grade 3 (%)	34.8	13.4	0.051	NS
	Node positive (%)	35.7	37.3	0.880	NS
	PR positive (%)	NA	NA		
GUYS77	N	20	57		
	Mean NF1 expression	-0.547	0.135		
	Median age	64	66	0.407	NS
	Median Tumour size (cm)	1.98	2.28	0.158	NS
	Grade 1 (%)	20.0	25.0	0.737	NS
	Grade 2 (%)	50.0	31.3	0.442	NS
	Grade 3 (%)	30.0	43.8	0.653	NS
	Node positive (%)	30.0	52.6	0.138	NS
	PR positive (%)	70.0	78.9	0.613	NS
GUYS87	N	22	65		
	Mean NF1 expression	-0.82	0.155		
	Median age	61.5	62	0.242	NS
	Median Tumour size (cm)	2.25	2.1	0.735	NS
	Grade 1 (%)	16.7	26.9	0.578	NS
	Grade 2 (%)	66.7	48.1	0.277	NS
	Grade 3 (%)	16.7	25.0	0.689	NS
	Node positive (%)	63.6	67.7	0.931	NS
	PR positive (%)	75.0	75.4	0.972	NS