SUPPLEMENTAL METHODS

Quantitative RT-PCR

Total RNA was isolated with the Triazol method (Invitrogen). Two micrograms of RNA were used, with 100 ng of random examers, in a reverse transcription reaction (SUPERSCRIPT II, Invitrogen). One-tenth ng of cDNA was amplified, in triplicate, in a reaction volume of 25 μL with 10 pMol of each gene specific primer and the SYBR-green PCR MasterMix (Applied Biosystems). Real-time PCR was carried out on the ABI/Prism 7700 Sequence Detector System (Perkin-Elmer/Applied Biosystems), using a pre-PCR step of 10 min at 95°C, followed by 40 cycles of 15 s at 95°C and 60 s at 60°C. Specificity of the amplified products was confirmed by melting curve analysis (DISSOCIATION CURVE™ Perkin-Elmer/Applied Biosystems) and by 6% PAGE. Preparations with RNA template without reverse transcriptase were used as negative controls. Samples were amplified with primers for each genes (for details see Q-PCR primer list below) and GAPDH as a housekeeping gene (other housekeeping genes, including rRNA 18S and beta-actin were also tested with comparable results). The Ct values were normalized to the GAPDH curve and the relative expression of each gene was expressed as the ratio relative to mock (dl312) infected myotubes.

Primers and oligonucleotides

The following primers and oligonucleotides were used:

Mouse Q-RT-PCR oligos

Gene	Oligo	Sequence
TRCP4	Qrip1f	GGAATGACGACAAACACACGC

TRCP4	Qrip1r	GCTGAGAAGAGCAGCTTGGTTG
TAF3	Qtaf-1f	CCCTTTCGAGTTTTCTTCCG
TAF3	Qtaf-1r	GTTATTGGAGGAAGTGGAGGCT
SMU-1	Q36-1F	CAAAGAGAAAAGGAGAGCAGCG
SMU-1	Q36-1r	GCTGCCACTTCAATGCCTG
SKIN	mSK1019	GTCCCGTGCAAGAAGCCTC
SKIN	mSK1137	TGCAATTCGCTTTGTGTACCC
SF3B1	Qsf3b1f	GCCTCTGTTTCTGGCGGC
SF3B1	Qsf3b1r	TCTTCGTGAGTCTTGGCGATC
RRM2	Qrrm1f	CTGACAAGGAGAACACGCCC
RRM2	Qrrm1r	GCGCTTTACTTTCCAGCTCG
PTB2	Qptb-1f	AGATGTGCAGCGCGTGAAG
PTB2	Qptb-1r	GCCCATTAAGATGATTCATGGC
NSAP	QsyncrF	AGGTGCCCAACAACAAAGAGG
NSAP	QsyncrR	TGCAAGCTTCTGGTGTAATCCC
NP95	MmNP775	ACGGCAACATCAGGCTCTTG
NP95	MmNP868	GGCTAGCGATCAAAGGCCTC
NASP	Qnasp3f	AGAGGTGAATGGAGGCAGTGG
NASP	Qnasp3r	CTCTGCTGTTTGGCTCTCAGC
mXTP1	QmXTP1F	CATTGCTTCACTGCATCGGAG
mXTP1	QmXTP1R	GCAGCTGGACTGTTTGTTTGC
MGC22679	Q10E-1	CCTGCCCATGCTCAATGC
MGC22679	Q10E-2	GAACTCGGCCTCTTCGGC
MCM7/cdc21	Qcd21f1	CATCACCATCAGTGGCATGG
MCM7/cdc21	Qcd21r1	TGCCTCGATCTATCTCCACCC
MCM6/MIS5	mMIS511	CCACAATCTCTGCACGAGCC
MCM6/MIS5	mMIS512	CCCCACGAAGAGAGGTTCC
MCM4/cdc47	mC47-12	GTGTGGCCAAATCTCAGCTCC
MCM4/cdc47	mC47-13	ACACAGCTGCTGTGAGCCC
LBR	LBR1254	GTTCTTTTGTGAGTTGCGCCC
LBR	LBR1364	ATCATTGCCAAGGACGGAGC
K1594	Q99-1f	CTCTGCATCATGGCATGAGG
K1594	Q99-1r	ACGGCCTTTCTAAGTCAACAGG
K0648	mK648-1	CCCAGTCTCCCGATGAATCC
K0648	mK648-2	TCCTTTGGTGAGTCTGCATTGC
HAT1	QHAT-1f	GCTTTTTGCGACCGTAGGC
HAT1	QHAT-1r	GGCCCTGACCTTGAAATGG
G3BP2	QG3bp1f	AGAAGTTCAGTCTCAGCCACCG
G3BP2	QG3BP1r	TGTTCCATATCTCCTCTGCCTG
FLJ37562	Q143c1F	TCCTGAAGAAATCTGAGCAGCG
FLJ37562	Q143c1r	CGCCGCCTCTTTCGACTC
E2F1	QmE2F1f	CAGAAACGGCGCATCTATGAC
E2F1	QmE2F1r	CATGGTGTGGCTGCCTAGC
CML66	Q98-1f	ACCCCGAGAGGTGTTCCG
CML66	Q98-1r	CCATCTGACAAGGTAACCCAGG
CHTOG	Qtog-1f	AAGAGAAACCTCAGCGCACAC
CHTOG	Qtog-1r	GGATGCCCACTCAGGCTTC
CCNE2	QcE2-2f	TGTGCATTCTAGCCATCGACTC

CCNE2	QcE2-2r	CCATTCCAAACCTGAAGCTTTC
CCNE1	mCE881	TGTTACAGATGGCGCTTGCTC
CCNE1	mCE988	TTCAGCCAGGACACAATGGTC
C3orf4	Q115-1f	ACTCACTGGTATGCGCCACC
C3orf4	Q115-1r	CGCTATTGTGGTTGCCGG
BAT1	Qbat-1F	GGTGTCTGTGCTGGTGATGTG
BAT1	Qbat-1R	CACTGCCACCTTGACATTCG
GDH	QmGDHf	CCCACTCTTCCACCTTCGATG
GDH	QmGDHr	GTCCACCACCCTGTTGCTGTAG

Human Q-RT-PCR oligos

Gene	Oligo	Sequence
Actin beta	QActB1f	AGGTCATCACCATTGGCAATG
Actin beta	QActB1r	GCGGATGTCCACGTCACAC
18S	Q18S1F	CGCCGCTAGAGGTGAAATTC
18S	Q18S1R	CTTTCGCTCTGGTCCGTCTT
Ch-TOG	QhsTOG1	TGTATGTTGGTCCCTCTTTGCG
Ch-TOG	QhsTOG2	GGTGGACTTTGTCCCTGCATC
CycA2	QhCyA2f	AGCTGGCCTGAATCATTAATACG
CycA2	QhCyA2r	GGTGAAGGTCCATGAGACAAGG
CycB1	QhCyB1f	TCCATTATTGATCGGTTCATGC
CycB1	QhCyB1r	TCAGTCACAAAAGCAAAGTCACC
CycE1	QhCyE1f	TGCAGAGCTGTTGGATCTCTGTG
CycE1	QhCyE1r	GGCCGAAGCAGCAAGTATACC
Gapdh	QhGDHf	ATCAGCAATGCCTCCTGCAC
Gapdh	QhGDHr	TGGCATGGACTGTGGTCATG
MGC22679	Qh10EF1	ATGGACAGAAGGTTTGCCAATG
MGC22679	Qh10ER1	AACTCAGCCTCTTCTGCTGGAG
SF3B1	QhSF3B1	AGATCGCCAAGACTCACGAAGA
SF3B1	QhSF3B2	CCTGTAGAATCGAGGCCCACTC
SKIN	QhSKN2f	AAATCCGGCCTGTCAGCAG
SKIN	QhSKN2r	GATGGTGGTCCTTGTCCAGTG
SMU-1	QhsSMU1	CTTTGTTTGCTGTGCCCTCTCT
SMU-1	QhsSMU2	CATCCTTCTCGTGCACTGTCAA
TRCP4	QhTRCP1	TCTGCCGGATTCTGGCTG
TRCP4	QhTRCP2	TGCAGAAGGCCCCAAACTC
XTP1	XTP425F	AATGGAATGATCTCCCACCAGG
XTP1	XTP541R	AAGCTGGCACCTCTCCAATTG
STAT1	QhSTA1F	CCCAGAATGCCCTGATTAATG
STAT1	QhSTA1R	CCCAGAATGCCCTGATTAATG
IFITM1	QIFIM1F	TGATCAACATCCACAGCGAGAC
IFITM1	QIFIM1R	TCCTGTCCCTAGACTTCACGGA

RNAi oligos

Gene	Oligo	Position	Sequence
SKIN	Oligo1	434	GGAGAAGCACAGAGAUAUUdTdT
SCRAMBLED	Scrambled control		AGACGAACAAGUCACCGACdTdT
SKIN	Oligo 2	245	UCACCUUGCACUCGAUUCAdTdT

Tissue Microarray design and analysis

TMAs were designed and prepared as previously described (1). Briefly, small tissue cylinders (diameter 0.6 mm) were removed from formalin-fixed and paraffin-embedded tumors (the donor blocks) and placed into a single empty "recipient" block, using a custom-built precision instrument (Tissue Arrayer - Beecher Instruments). Two representative areas (cores) of the donor block, previously identified on hematoxylin-eosin-stained sections, were deposited on the recipient block. In general, normal samples from the same patient (from surrounding normal tissue) were also deposited in duplicate on the recipient block. In some cases (~ 25 % cumulatively for breast, colon, kidney larynx, lung, prostate and stomach tumors, see Supplemental Table 9) this was not possible, due to unavailability of tissue. In the case of uterus, melanoma and brain, no normal tissue was available. In these cases, overexpression was evaluated by internal comparison among tumor samples, and the definition "overexpression" should be considered tentative. Three μm sections of the resulting recipient block were cut, mounted on glass slides, and processed for (ISH) or immunohistochemistry (IHC) analysis.

For the purpose of this study, we prepared five multi-tumor TMAs with samples derived from the most common types of cancer: colon, lung, breast, stomach, uterus, larynx, prostate, melanoma, glioblastomas, and sarcomas. The exact composition of each TMA is reported in Supplementary Table 9, together with specifics about which TMA was used for the individual genes

We also created a colon progression TMA (utilized in Figure 4 of the main text) including normal epithelia, hyperplastic polyps, adenomas and adenocarcinomas (number of samples are in the legend to Figure 4A of the main text). In the colon progression TMA, we deposited, whenever possible, matched samples from the same patient (normal-adenoma-tumor from 8 patients; normal-tumor from 18 patients; adenoma-tumor from 5 patients).

In situ hybridization (ISH) was performed as previously described (2). Briefly, we employed S-UTP-labeled sense and antisense riboprobes, generated from the most specific region (300bp) of each gene, identified by BLAST searches. The identified probes were produced by PCR using oligos flanked by T3 and T7 RNA polymerase promoters, followed by in vitro transcription. Sequences of probes are available upon request. TMA sections were dewaxed, digested with Proteinase K (20 mg/ml), post-fixed, acetylated and dried. After overnight hybridization at 50^0 C, sections were washed in 50% formamide, 2X SSC, 20 mM 2-mercaptoethanol at 60^0 C, coated with Kodak NTB-2 photographic emulsion, and exposed for three weeks. The slides were lightly H&E counterstained for morphologic evaluation.

Immunohistochemical staining for the proliferation marker Ki67 was performed on the same tissue microarrays using the EnVision+ system (Dako). Formalin-fixed, paraffin-embedded tissue sections were deparaffinized, rehydrated, and exposed to the primary antibody (Mib1 clone, DBA) at a dilution of 1:200 for 2 hours at room temperature. Microwave-pretreatment in EDTA pH 8.0 was used to retrieve antigenicity. Positive and negative controls were included in each experiment. The correlation with Ki67, and the other histological parameters, was determined by the

Contingency Table analysis with Likelihood-Ratio Chi-square (G^2) (JMP™ IN 5.1).

Microarray data normalization and analysis

Affymetrix microarray intensities, from our screening of 46 lymph node negative (N0), ER+, breast cancer patients (M. Vecchi. F. Bianchi and P.P. Di Fiore, unpublished, see however Supplemental Table 13 for relevant data, henceforth referred to as the "IFOM dataset"), were scaled to a TGT factor of 500 by MAS 5.1 software (Affymetrix®). Normalized data were then imported in Genespring® 6.2 (Silicon Genetics, Redwood City, California), where global per-chip median normalization and log (natural base) data transformation was applied. Class-D genes where analyzed and only the genes with at least 1.5 fold regulation and a p-value less than 0.01 (Welch t-test) in at least one of the two N0 class (relapse or not), compared to the level of expression in normal samples (12 samples analyzed in the same screening), were selected. Three Class-D genes passed this selection, SKIN, ch-TOG and TRPC4AP.

cDNA microarray data from the study of van't Veer and co-workers (3) (http://www.rii.com/publications/2002/vantveer.html, henceforth referred to as the "VV dataset") were downloaded (all N0 ER+ patients, 60 patients) and used without additional normalization steps. We performed a Welch t-test on Class-D genes, between the N0 and N0+ classes, with the following p values: SF3B1, 0.64; SMU-1, 0.23; TRCPA4, 0.06; ch-TOG, 0.02; SKIN, 0.01. Thus, SKIN and ch-TOG passed the statistical test, while TRCPA4 was borderline.

We, therefore, built a 2-gene model (SKIN and ch-TOG). The exclusion of TRCPA4 was due to two reasons. First, in the VV dataset the significance of

TRCPA4 was borderline. Second in a series of Q-RT-PCR experiments, we found differences in the direction of expression (over or under expressed) of TRCPA4, when comparing data obtained with Q-RT-PCR and Affymetrix (not shown). The reason for this discrepancy is not clear. It however compelled the exclusion of TRCPA4 from further analyses.

The following procedures were then implemented:

Kaplan-Meier survival curves

For Kaplan-Meier survival curves the following steps were implemented:

Step 1:

To categorize the patients into "favorable" and "unfavorable" prognostic classes, using the 2-gene model, we built a binary matrix based on the level of expression of each gene in each patient using the 50^{th} percentile cutoff point. The direction of regulation of a gene was assigned to belong to the "favorable" or "unfavorable" signature based on the difference of the median values between the relapse and the disease-free N0 class. In brief:

compute,
a = median (N0)
compute,
b = median (N0+)
compute,
z = median (N0, N0+)
then,
IF i (**g**) > z_i where $a_i < b_i$
g is defined as high risk
IF i (**g**) > z_i where $a_i > b_i$
g is defined as low risk
or,
IF i (**g**) < z_i where $a_i < b_i$
g is defined as low risk
IF i (**g**) < z_i where $a_i > b_i$
g is defined as high risk
or
IF i (**g**) = z_i
g is not defined

where, *i* is the normalized (median scaled intensity of the entire array) gene value in each *g* patient.

Each patient was then classified weighing up the votes in the binary matrix of the classifier.

Step 2:

Using the JMP-IN 5.1 software package (SAS Institute Inc.), we tested the bona-fide level of the classifier non-parametrically, using Kaplan-Meier survival curves; the relative null hypothesis (there is no difference between the two populations in the probability of an event) was tested by the log-rank test.

Diagonal Linear Discriminant Analysis

For Diagonal Linear Discriminant Analysis, we used BRB ArrayTools developed by Drs. Richard Simon and Amy Peng Lam. Diagonal Linear Discriminant Analysis is a version of linear discriminant analysis that ignores correlations among the genes in order to avoid over-fitting the data (4, 5), and was employed to further verify the performance of the classifier. The misclassification rate of the classifier is reported in percentage after the leave-one-out cross-validation procedure, and the relative p-value means the proportion of the random permutations of class labels (10,000 permutations) that gave as small a cross-validated misclassification rate as was obtained with the real class labels. The 2-gene model (SKIN and ch-TOG) predicted with an accuracy of 67% and a p-value of 0.029 in the IFOM dataset, and with an accuracy of 70% and a p value of 0.003 in the VV dataset.

Random co-occurrence test

For the Random co-occurrence test, the following steps were implemented:

Step 1:

In our unpublished screening (IFOM dataset), we analyzed 46 lymph node negative (N0), ER+, breast cancer patients, and 12 normal non-neoplastic breast samples. We used Genespring® 6.2 (Silicon Genetics, Redwood City, California) to select genes which are differentially expressed between the normal samples and all the N0 breast tumors. The list is composed of 966 genes with a p-value of less than 0.001 (Welch t-test and multiple testing correction: Benjamini and Hochberg False Discovery Rate). We refer to this list of genes as List A.

An alternative list was created by selecting all the probesets flagged "present or marginal" by the MAS 5.1 software (Affymetrix®) in the 46 N0 patients of the IFOM dataset (the 966 probesets of list A were excluded from this list). This resulted in a list of 6417 probesets, which we refer to as List B.

We then performed two parallel analyses (described underneath) using List A and B.

Step 2:

List A (966 probesets). Using the R 2.1.0 environment (written by Robert Gentleman and Ross Ihaka), we generated 10,000 random 2-gene signatures, and we performed a Diagonal Linear Discriminant Analysis with leave-one-out cross validation, to assess the performance of each individual 2-gene signature on the IFOM and VV datasets. Affymetrix probesets and cDNA array probes were matched through Unigene ID (release 184). Of 10,000 random 2-gene signatures, 1168 performed as our 2-gene model (≥67%) on the IFOM dataset. Of these 1168 signatures, 849 were present in the VV dataset. Only 8 out of 849 2-gene signatures performed on the VV dataset as our 2-gene model (≥70%).

List B (6417 probesets): Of 10,000 random 2-gene signatures, generated as above, 1223 performed as our 2-gene model (≥67%) on the IFOM dataset. Of these 1223 signatures, 933 were present also on the VV dataset. 16 out of these 933 signatures performed on the VV dataset as our 2-gene model (≥70%).

Therefore, 8 of 10,000 (list A approach), and 16 of 10,000 (list B approach) random 2-gene signatures performed as efficiently as our 2-gene model on two independent datasets.

Human SKIN cDNA sequence

ATGAACATAGACGTGGAGTTCCACATCCGGCACAACTACCCCTGGAACAA
GTTGCCGGCCAACGTGAGACAGAGTCTTGGAAATTCACAGAGAGAATATG
AAAAGCAGGTTGTCCTGTACAGTATCCGCAATCAGTTACGATATAGAAAT
AACTTAGTTAAACATGTCAAGAAAGATGAACGCAGATACTATGAGGAACT
GCTAAAGTACAGCCGAGATCATCTCATGCTGTACCCTTACCATCTATCGG
ATATTATGGTGAAAGGCTTGAGGATAACACCATTTTCATATTATACTGGG
ATTATGGAGGATATTATGAACAGTGAGAAAAGTTATGATTCATTGCCCAA
TTTTACTGCTGCTGACTGTCTAAGGCTTCTTGGCATAGGAAGAAACCAGT
ATATTGATCTTATGAATCAGTGTAGATCATCAAAAAAATTCTTCAGAAGG
AAAACAGCCCGTGATCTTCTACCAATAAAGCCAGTGGAAATTGCCATAGA
GGCGTGGTGGGTGGTGCAGGCTGGCTATATCACAGAAGATGACATCAAGA
TATGCACTTTGCCTGAGAAATGCGCTGTTGATAAGATCATCGATTCAGGC
CCTCAACTCTCTGGATCACTAGATTACAATGTAGTACATAGTTTGTATAA
CAAAGGATTTATTTATCTGGATGTACCAATATCTGATGACAGTTGTATAG
CAGTTCCACCTCTTGAAGGTTTTGTAATGAATCGAGTGCAAGGTGATTAT
TTTGAAACTCTACTCTATAAGATATTTGTTTCAATAGATGAGCACACAAA
TGTTGCAGAGCTTGCAAATGTCCTTGAGATTGACTTATCCCTGGTTAAGA
ATGCTGTTTCAATGTATTGCCGATTGGGCTTTGCCCATAAGAAGGGACAA
GTAATTAATTTGGATCAACTTCATTCATCATGGAAGAATGTTCCATCCGT
AAACAGATTAAAGAGTACCTTAGATCCACAGAAGATGCTCTTGTCATGGG
ATGGAGGGGAAAGTAGGAGTCCTGTACAAGAAGCTTCATCGGCAACTGAC
ACTGATACAAATAGTCAAGAAGATCCAGCTGACACAGCCAGTGTAAGCAG
CCTGAGTCTGTCTACAGGACACACGAAGCGCATCGCATTCCTGTTTGACT
CCACTCTTACTGCCTTCTTAATGATGGGAAATCTTTCACCAAACTTGAAA
AGTCATGCAGTCACAATGTTTGAAGTAGGCAAACTCTCAGATGAGTCTCT
GGACAGCTTTCTTATAGAACTAGAAAAGGTTCAGAGCACTGGTGAAGGAG
AAGCACAGAGATATTTTGATCATGCACTTACTCTGAGAAACACAATACTG
TTTCTGCGTCATAACAAAGATCTAGTTGCGCAAACTGCACAGCCAGACCA
ACCCAATTATGGTTTTCCTCTGGATCTCTTACGCTGTGAAAGCCTTCTTG
GTTTGGACCCTGCAACTTGCAGCAGAGTTCTAAACAAAAATTACACGCTG
CTTGTTTCCATGGCTCCCCTCACCAATGAAATCCGGCCTGTCAGCAGCTG
CACCCCTCAGCATATTGGACCAGCTATCCCAGAAGTCAGCTCTGTCTGGT
TTAAACTGTACATTTATCATGTCACTGGACAAGGACCACCATCCCTTTTA
TTGTCCAAAGGTACAAGACTTCGAAAACTGCCAGATATATTTCAGAGTTA
TGATCGATTGCTAATAACATCTTGGGGTCATGATCCTGGAGTAGTTCCTA
CCTCAAATGTGCTCACGATGTTGAATGATGCTTTAACACATTCTGCAGTT
TTAATTCAGGGGCATGGTCTGCATGGGATAGGAGAAACTGTCCATGTCCC
ATTTCCATTTGATGAAACAGAACTACAAGGAGAGTTCACTCGTGTCAATA
TGGGTGTTCATAAAGCATTGCAGATACTAAGGAACAGAGTGGACTTACAG
CATCTCTGTGGATATGTCACCATGTTGAATGCTTCCAGCCAACTTGCAGA
TAGAAAACTCAGTGATGCTTCTGATGAGAGAGGAGAACCTGATTTGGCTT

CTGGCTCAGATGTAAATGGGAGTACAGAGTCATTTGAAATGGTCATTGAG
GAAGCAACTATAGATTCAGCAACAAAGCAAACCTCTGGTGCCACAACAGA
AGCAGATTGGGTTCCTCTCGAGCTGTGCTTTGGAATTCCACTGTTCAGTT
CCGAATTAAACCGGAAAGTTTGTAGGAAAATTGCTGCACATGGCCTTTGC
AGAAAAGAGAGCCTTCAAAACCTCTTACATTCCAGTAGAAAACTCTCTCT
GCAAGTCCTTAACTTTGTTCACTCATTCCAGGAAGGTGCTTCAATATTGG
ATATTCACACAGAGCCCAGTTTTTCAAGTTTGCTTTCACAGTCATCGTAT
GCTGACATGGGTGTTCCACTTCCTGCAAAAAACTTAATATTTAAAGATGG
TGTCTTATCAGAATGGAGTGGACGGTCACCTTCCTCACTTCTTATTGCTA
ATCTCCATTTGCAATAA

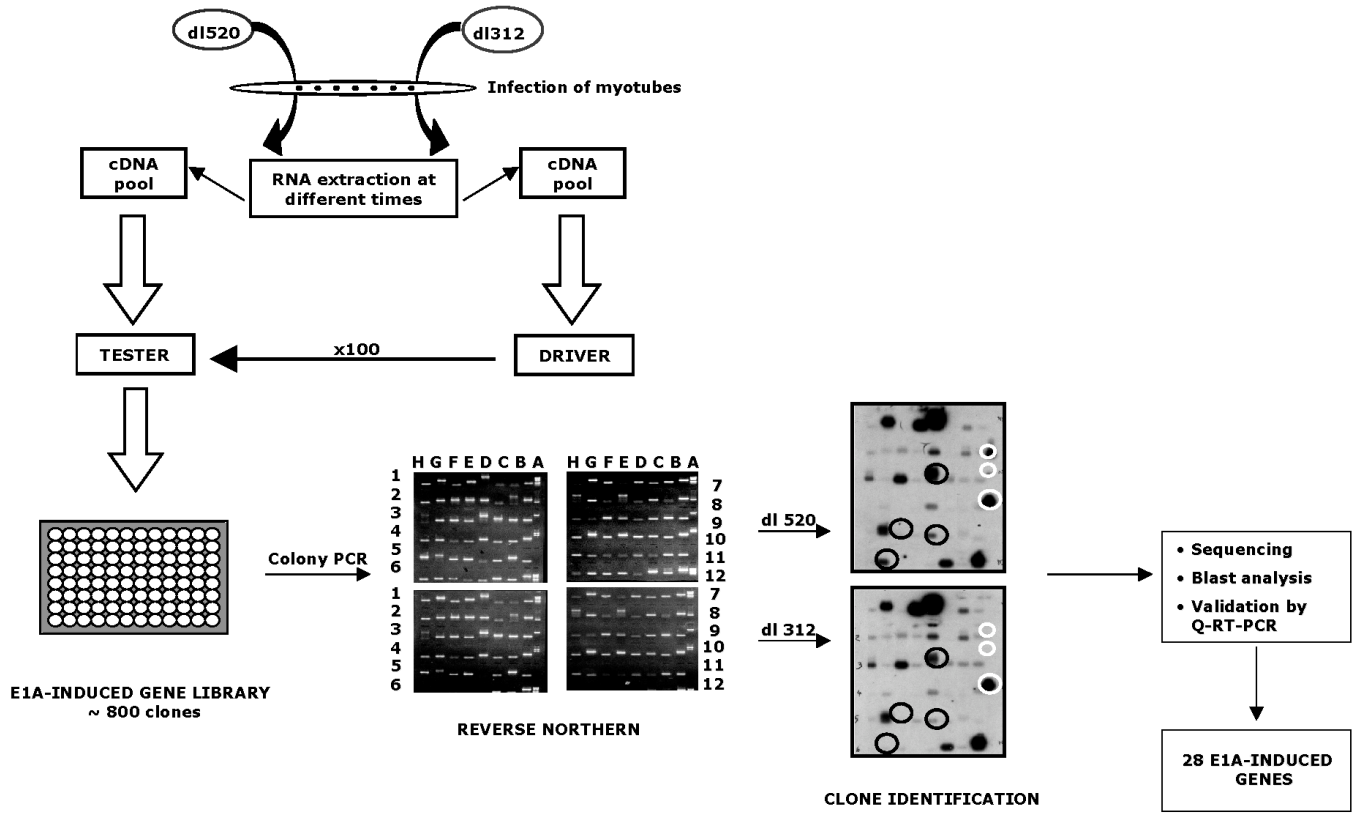
Mouse SKIN cDNA sequence

ATGAACATCGACGTTGAGTTCCACATCCGGCACAATTATCCTTGGAGCAA
GTTGCCAACGAACGTGAAGCAGAGTCTTGGAAACTCACAGAGAGAATATG
AAAAGCAGGTTGTTTTGTACAGTATTCGCAATCAGTTGCGCTACAGAAAT
AACCTAGTTAAACATGTCAAGAAAGATGAACGCAAGTACTATGAGGAGCT
GCTCAAGTACAGTCGTGACCATCTCATGCTGTACCCTTACCACTTATCCG
ACATCATGGTCAAGGGCCTGAGGATAACCCCATTCTCATACTATGCTGGG
ATCATGGAGGACATCATGAACAGTGAGAAAAGTTACGATTCATTGCCTAA
TTTTACTGCTGCTGACTGTCTGCGACTTCTTGGCATAGGAAGGAACCAGT
ACATTGACCTGATGAATCAGTGCAGATCTTCAAAAAAATTCTTCAGAAGG
AAAACTGCCCGTGACCTTCTACCAATGAAGCCAGTGGAAATTGCCATCGA
GGCGTGGTGGGTGGTGCAGGCTGGCTATATCACAGAAGATGACATCAAGA
TATGCACTTTTCCTGAGAAAGGCGCCATTGATAAGATCATCGACTCAGGT
CCTCAGCTCTCTGGGTCACTAGATTACAATGTCGTACACAGTTTGTATAA
CAAAGGCTTTATTTATCTGGATGTGCCCATATCAGATGACAGTTGTATAG
CTGTCCCCCCTCTGGAAGGCTTTGTAATGAATCGGGTACAAGGTGACTAT
TTTGAAACCCTACTCTATAAGATATTTGTGTCTATAGATGAACACACTAA
TGTTGCAGAGCTTGCAAATGTCCTTGAGATAGACTTATCGCTGGTGAAGA
ATGCTGTTTCAATGTATTGCCGATTGGGCTTTGCCCATAAGAAGGGACAA
GTGATAAATTTGGATCAACTTCATTCGTCATGGAAGAATGTTCCATCTGT
GAATAGACTAAAGAGCACATTAGACCCACAAAAGATGCTCTTATCGTGGG
ATGGCGGGGAGAGTAGGAGTCCCGTGCAAGAAGCCTCTTCTGCTACTGAC
ACAGACACAAATAGTCAAGAGGATCCAGCTGACACGGCCAGTGTGAGCAG
TTTGAGTCTGTCCACTGGGTACACAAAGCGAATTGCATTCCTGTTTGACT
CCACTCTGACTGCATTCTTAATGATGGGAAACCTTTCACCGAACTTGAAA
AGCCACGCAGTCACCATGTTTGAAGTAGGCAAACTCTCAGATGAGTCTCT
GGACAGCTTCCTTATAGAATTAGAGAAGGTTCAGAGCACTGGTGAAGGAG
AAGCACAGAGATACTTTGACCACGCACTTACTCTGAGGAACACTATCCTC
TTCCTGCGTCATAATAAGGATCTGGTGGCTCAGACCTCTCAGCCAGACCA
GCCCAACTATGGTTTCCCCCTGGATCTCTTACGCTGCGAAAGCCTTCTTG
GTTTGGACCCAGCAACTTGCAGCAGAGTTCTAAACAAAAACTACACACTG
CTTGTTTCCATGGCGCCTCTTACCAATGAGATCCGGCCAGTCAGCAGCTG
CACCCCGCAGCATATCGGACCAGCTATCCCAGAAGTCAGTTCTGTCTGGT
TTAAATTGTATATTTACCATGTCACTGGGCAAGGACCACCATCTCTCTTA
CTGTCCAAAGGTACAAGACTTCGGAAACTGCCAGATATATTCCAGGGTTA
TGATCGGTTACTTATAACCTCTTGGGGCCATGACCCTGGGGTCGTTCCTG
CTTCAAACGTTCTCACGATGCTGAATGATGCTTTAACCCACTCAGCAGTG
CTCATTCAGGGACATGGTTTGCATGGGGTAGGAGAGACTGTCCATATACC
CTTTCCATTTGATGAAGCAGAGCTGCAAGGAGAGTTCACCCGTGCCAGTA
TGGGTGTTCATAAAGCACTGCAGATTCTGCGGAGCAGAGTGGACTTGCAG
CATTTCTGCGGCTATGTCACCATGTTGAATGCTTCTAGCCAGCTTGCAAG
TAGGAAGCTCAGTGAGGCTTCTGATGAGAGAGGCGAGCCTGATTTGGCTT
CTAGTTCAGATGTAAATGGGAGCACAGAGTCGTTTGAAATGGTCATTGAG
GAAGCAAGTACAGATTTGGCGACAAAACCAAACTCTGGGGCCACAGCAGA
AGCAGATTGGGTTCCTCTTGAGCTGTGTTTTGGAATTCCTTTGTTCAGTT
CCGAATTAAACCGGAAAGTTTGTCAGAAAATAGCTACCCATGGCCTGTGC
AGGAAAGAGAGCCTTCAAAGTCTCTTACATTCCAGTAGAAAACTCTCCCT
GCAAGTCCTTAACTTTGTTCATTCATTCCAGGAAGGTGCTGCTACTCTGG
ATCTCCACGCGGAGCCTGGCTTCTCAAGTGTGCTTTCACAGTCGCCCTGC

GCTGACATGGGAGTTCCACTTCCGGCCAAAAACTTAATGTTTAAAGATGG
TGTCTTATCAGAGTGGAGCGGACGCTCACCTTCCTCCCTGCTCATCGCTA
GTCTAAGCTTGTAG

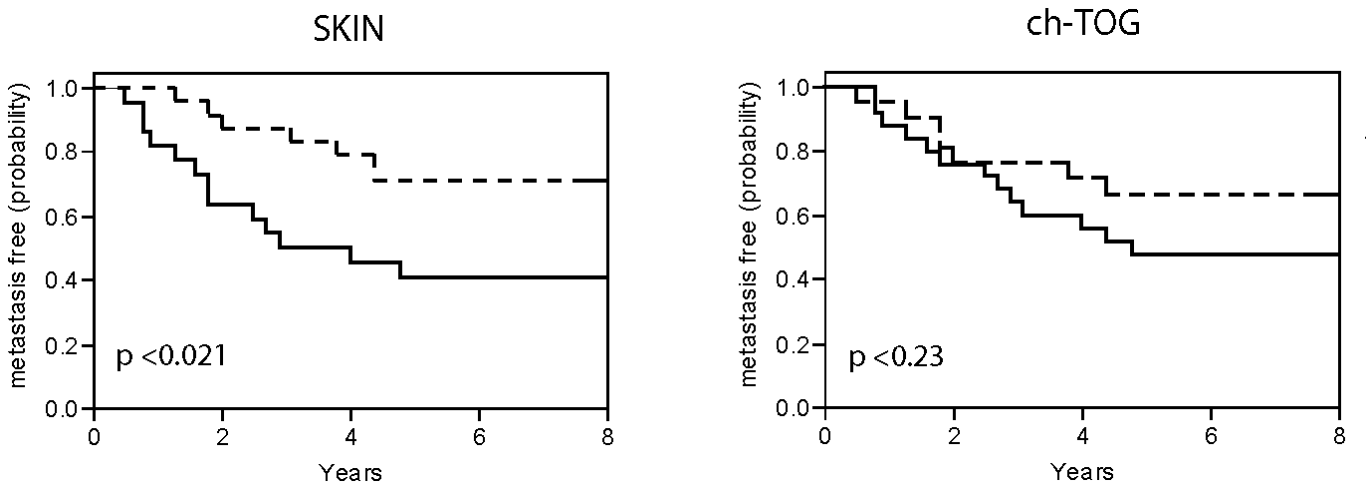
REFERENCES TO SUPPLEMENTAL METHODS

1. Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M.J., Sauter, G., and Kallioniemi, O.P. 1998. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4:844-847.
2. Rugarli, E.I., Lutz, B., Kuratani, S.C., Wawersik, S., Borsani, G., Ballabio, A., and Eichele, G. 1993. Expression pattern of the Kallmann syndrome gene in the olfactory system suggests a role in neuronal targeting. *Nat Genet* 4:19-26.
3. van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530-536.
4. Dudoit, S., Fridlyand, J., and Speed, T.P. 2002. Comparison of discrimination methods for the classification of tumors using gene expresssion data. *Journal of the American Statistical Association* 97:77-87.
5. Molinaro, A.M., Simon, R., and Pfeiffer, R.M. 2005. Prediction error estimation: a comparison of resampling methods. *Bioinformatics*.

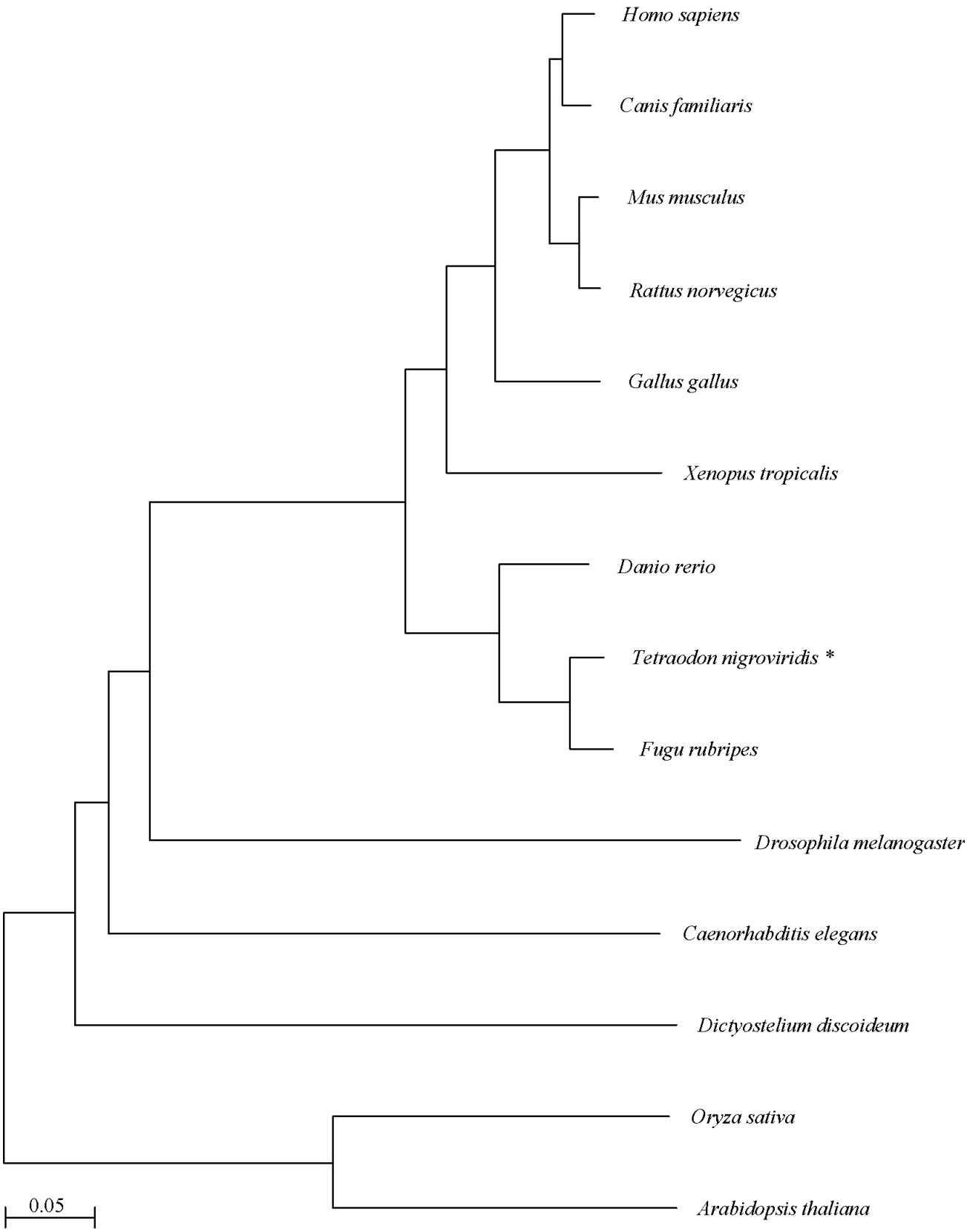


Supplemental Figure 1: Construction and screening of an E1A-induced gene library. TD C2C12 myotubes were infected with either the adenovirus dl520 (expressing only the 12S mRNA of E1A) or the control adenovirus dl312 (expressing no E1A). About 70% of myotubes infected with dl520 displayed S-phase re-entry (as measured by BrdU incorporation) 48 h post-infection (p.i.); whereas dl312-infected myotubes did not re-enter S phase. Total RNAs were harvested at various time points after infection (14, 15, 16, 18, 20, 24 h), pooled (in equimolar ratios) and subjected to polyA+ isolation, by standard methods. Two micrograms of polyA$^+$ RNA from dl520 (tester) and dl312 (driver) infected myotubes were then used as starting material for cDNA retro-transcription (Invitrogen). The subsequent subtraction procedures were performed with a commercial kit (BD Biosciences-Clontech), according to the manufacturer's specifications. In particular, in the first subtraction, adaptor-free driver cDNA (dl312) was subtracted, at a 100:1 ratio, to each of the two adaptor-linked tester cDNA population (dl520), in order to enrich for tester-specific (dl520) differentially expressed sequences. In the second subtraction, the previously subtracted tester (dl520) cDNA populations were mixed together and amplified according to the manufacturer's protocol (BD Biosciences-Clontech) to equalize the rare and the abundant transcripts. Secondary PCR amplification products were cloned into the PCR II vector (Invitrogen). These procedures yielded a library of ~ 800 clones that was further screened by Reverse Northern. Parallel filters contained all the cloned sequences, as single purified PCR bands, and some controls (DNA ladder IX as negative control, adenoviral cDNA and NP95 sequence as positive control, GAPDH as internal standard). The duplicate filters were then hybridized with two labeled cDNAs obtained from the same mRNA pools used as tester and driver (from dl520- and dl312-infected myotubes cDNA, respectively). This allowed for the identification of E1A-induced clones (circled in black in "clone identification". The clones circled in white are positive and negative controls). Clones were then picked, sequenced and analyzed with the BLAST algorithm, for further grouping (frequently more than one clone represented the same gene).

The final validation was by Q-RT-PCR on total RNAs from dl520- and dl312-infected myotubes. Only those genes displaying reproducible and higher than two-fold induction in at least two independent experiments were selected for further studies (twenty-eight genes, globally defined as E1A-induced genes, see details in Supplemental Tables 1 and 2).

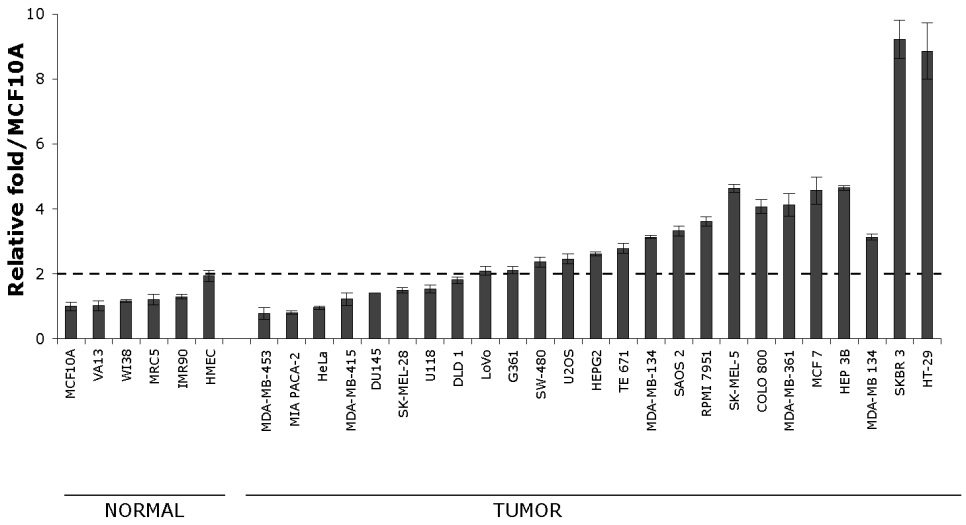


Supplemental Figure 2: Individual covariates of the analysis presented in Figure 5A. SKIN or ch-TOG were used individually as predictors of prognostic outcome on our dataset of 46 N0 patients. Data are shown as the probability of remaining free of metastatic relapse, in a Kaplan-Meier plot, as a function of a "favorable" (dashed line), or "unfavorable" (continuous line) signature. The shown p values were calculated with the log-rank test; of note, in the 2-gene model (Figure 5A), p value was < 0.005.



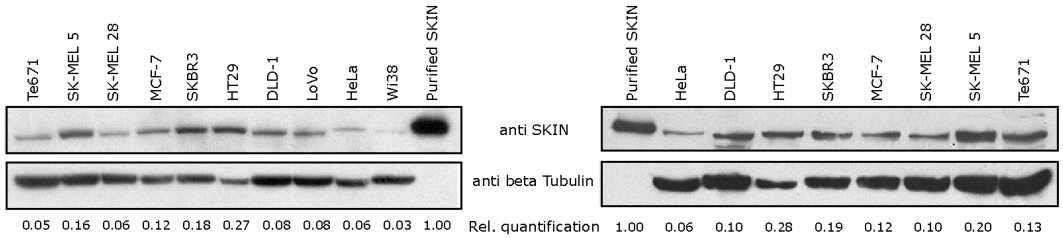
Supplemental Figure 3: Phylogenetic tree of SKIN proteins from different species. Protein sequences were retrieved either from GenBank or predicted from genomic assemblies from various organisms and aligned using the Clustalw program. The phylogenetic trees were then drawn after the alignments were manually adjusted. Genomic databases were searched with the BLAT program at the UCSC Genome Browser (http://genome.ucsc.edu/), and when needed the entire locus was reconstructed with the GeneWise software (http://www.ebi.ac.uk/Wise2/). Xenopus Laevis, for which the entire CDS cannot be recovered (exon 19 is missing), is marked by an asterisk.

A

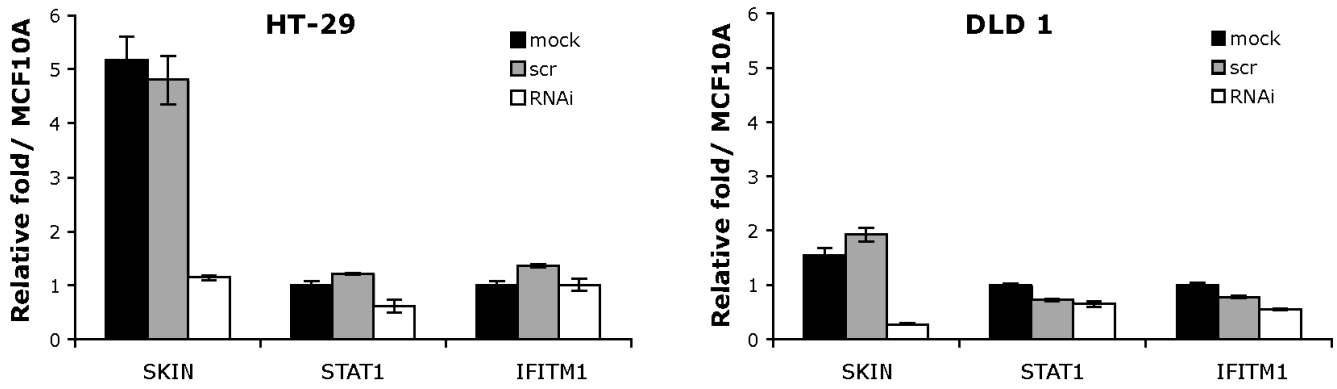


CELL TYPE	TISSUE TYPE	CELL LINE	RNA level
NORMAL	BREAST EPITHELIAL	MCF10A	1.00 ± 0.1
NORMAL	LUNG FIBROBLAST	VA13	1.02 ± 0.2
NORMAL	LUNG FIBROBLAST	WI38	1.17 ± 0.0
NORMAL	LUNG FIBROBLAST	MRC5	1.21 ± 0.2
NORMAL	LUNG FIBROBLAST	IMR-90	1.30 ± 0.1
NORMAL	BREAST EPITHELIAL	HMEC	1.94 ± 0.2
TUMOR	BREAST CARCINOMA	MDA-MB-453	0.80 ± 0.2
TUMOR	PANCREATIC CARCINOMA	MIA PaCa-2	0.81 ± 0.1
TUMOR	CERVIX CARCINOMA	HeLa	0.96 ± 0.1
TUMOR	BREAST CARCINOMA	MDA-MB-415	1.23 ± 0.2
TUMOR	PROSTATIC CARCINOMA	DU145	1.41 ± 0.0
TUMOR	MELANOMA	SK-MEL-28	1.50 ± 0.1
TUMOR	GLIOBLASTOMA	U118	1.55 ± 0.1
TUMOR	COLON CARCINOMA	DLD-1	1.81 ± 0.1
TUMOR	COLON CARCINOMA	LoVo	2.09 ± 0.1
TUMOR	MELANOMA	G361	2.11 ± 0.1
TUMOR	COLON CARCINOMA	SW-480	2.37 ± 0.1
TUMOR	OSTEOSARCOMA	U2OS	2.46 ± 0.2
TUMOR	HEPATOCARCINOMA	HEPG2	2.62 ± 0.2
TUMOR	RHABDOMIOSARCOMA	TE671	2.79 ± 0.2
TUMOR	BREAST CARCINOMA	MDA-MB-134	3.13 ± 0.1
TUMOR	OSTEOSARCOMA	SAOS2	3.32 ± 0.1
TUMOR	MELANOMA	RPMI 7951	3.61 ± 0.1
TUMOR	MELANOMA	SK-MEL-5	4.63 ± 0.1
TUMOR	MELANOMA	COLO 800	4.07 ± 0.2
TUMOR	BREAST CARCINOMA	MDA-MB-361	4.12 ± 0.2
TUMOR	BREAST CARCINOMA	MCF7	4.57 ± 0.4
TUMOR	HEPATOCARCINOMA	HEP3B	4.65 ± 0.1
TUMOR	BREAST CARCINOMA	MDA-MB-134	3.13 ± 0.1
TUMOR	BREAST CARCINOMA	SKBR3	9.22 ± 0.6
TUMOR	COLON CARCINOMA	HT-29	8.86 ± 0.8

B



Supplemental Figure 4: SKIN mRNA levels in normal and tumor cell lines. A. Q-RT-PCR analysis of SKIN mRNA in various normal and tumor cell lines. Ten ng of cDNA from each cell line were used in the amplification. Values were normalized to those detected in MCF-10A cells (assumed as 1), to allow for internal comparison. All normal cells displayed low levels of SKIN transcripts. Tumor cell lines were arbitrarily considered as overexpressors when they displayed ≥ 2-fold increase, as compared to MCF-10A (indicated by a dashed line). **B.** Western blot of lysates from representative cell lines with an anti-SKIN antibody. Two independent blots are shown. Beta-tubulin is shown as a loading control. In the lane "purified SKIN" 1.0 microgram of purified SKIN (purified from 293 cells transfected with a His-SKIN expression vector) were loaded as a reference. A quantitation of SKIN protein levels is also shown, after normalization for beta-tubulin levels, and is reported relative to the "purified SKIN" standard, assumed as 1. As shown, relative protein levels match reasonably well the mRNA expression levels (shown in A).



Supplemental Figure 5: additional controls for SKIN KD experiments.
The Figure shows the effects of the KD procedures on SKIN mRNA and on two PKR-induced genes, as a control for non-specific PKR-induced effects. HT-29 or DLD 1 cells were treated with 100 nM of either a SKIN-specific siRNA (RNAi) or a control scrambled oligo (scr) or mock-treated (mock). 24 h after the siRNA treatment, the levels of SKIN and of two PKR-induced genes (STAT1 and Interferon induced transmembrane protein-IFITM1) were measured by Q-RT-PCR to verify the KD of SKIN and the absence of non-specific siRNA induced effects. Values are expressed relative to those detected in growing MCF10A cells, to allow for comparison among cell lines.

Supplemental Table 1: E1A-induced genes.

ACC. N° (mouse)	NAME AND DESCRIPTION (mouse)	HUMAN HOMOLOG	ACC. N° (human)	SHORT NAME
NM_009830	CCNE2:cyclin E2	CCNE2	NM_004702	CCNE2
NM_008567	MIS5/MCM6 minichromosome maintenance deficient 6	MCM6	NM_005915	MCM6
NM_010931	Np95: nuclear protein 95	UHRF1	NM_013282	NP95
NM_007633	CCNE1:cyclin E1	CCNE1	NM_001238	CCNE1
NM_008568	cdc47/MCM7: minichromosome maintenance deficient 7	MCM7	NM_005916	MCM7
NM_008565	cdc21/MCM4: minichromosome maintenance deficient 4	MCM4	NM_005914	MCM4
NM_178683	XTP1: HBxAg transactiv. Prot.1	XTP1	NM_018369	XTP1
NM_016777	Nasp (somatic): nuclear autoantigenic sperm prot.	NASP(somatic)	NM_152298	NASPs
NM_026115	Hat1: histidine aminotransferase 1	HAT1	NM_003642	HAT1
NM_133815	Lbr: lamin B receptor	LBR	NM_002296	LBR
NM_007891	E2f1: E2F transcription factor 1	E2F1	NM_005225	E2F1
NM_177784	RIKEN cDNA C130068N17	MGC22679	NM_144711	MGC22679
NM_009104	Rrm2: ribonucleotide reductase M2	RRM2	NM_001034	RRM2
NM_176972	C330008N13Rik: ubiquitin spec. protease 37	KIAA1594	NM_020935	K1594
XM_130287	4930432B04Rik/mKIAA0097	KIAA0097/ch-TOG	NM_014756	Ch-TOG
XM_358357	9030416H16Rik/mouse KIAA0648	KIAA0648	NM_015200	K0648
NM_031179	SAP150: splicing factor 3b, subunit 1	SF3B1	NM_012433	SF3B1
NM_026149	4921532K09Rik/mouse CML_66	CML66	NM_032869	CML66
NM_171826	1110019C08Rik/mouse C3Orf_4	C3orf4	NM_019895	C3orf4
NM_145959	BC033609 unknown/mouse SKIN	FLJ23790: SKIN	NM_144963	SKIN
NM_011816	G3BP2: Ras-GTPase-activating protein	G3BP2	AB014560	G3BP2
XM_129997	Taf3: TAF3 RNA polymerase II	TAF3	XM_291729	TAF3
NM_019550	Ptbp2:polypyrimidine tract binding protein 2	PTBP2	NM_021190	PTB2
NM_181278	RIKEN cDNA B230219D22	FLJ37562	NM_152409	FLJ37562
NM_019693	Bat1a:HLA-B-associated transcript 1A	BAT1	NM_004640	BAT1
NM_019666	Syncrip: synaptotagmin binding	NSAP1	NM_006372	NSAP1
NM_019828	Trpc4ap: transient receptor pot. Cat. channel 4 ass. Prot.	TRPC4AP	BC013144	TRPC4AP
NM_021535	RIKEN cDNA 2610203K23	SMU-1	BC002876	SMU-1

Supplemental Table 1: E1A-induced genes. The Table shows all twenty-eight E1A-induced genes. Shown are, the short name of each gene (used throughout the descriptions of the main text and in all Tables and Figures), followed by its accession number (mouse), its full name and description (mouse), the name of the human homologue and the accession number of the human homologue.

Supplemental Table 2: Validation of E1A-induced genes.

	E1A-C2C12 (TD)		Myobl. C2C12		E1A-MSC (TD)		Myobl. MSC	
SHORT NAME	**Induction**	**p value**	**Induction**	**p value**	**Induction**	**p value**	**Induction**	**p value**
CCNE2	24.4 ± 1.8	<0.01	7.4 ± 0.0	<0.01	84.5 ± 4.4	<0.01	11.3 ± 1.2	<0.01
MCM6	16.9 ± 2.0	<0.01	11.3 ± 0.6	<0.01	70.3 ± 1.0	<0.01	46.7 ± 0.9	<0.01
NP95	19.1 ± 0.8	<0.01	22.3 ± 1.2	<0.01	54.2 ± 3.4	<0.01	47.2 ± 2.5	<0.01
CCNE1	11.1 ± 0.2	<0.01	3.7 ± 0.0	<0.01	29.6 ± 2.6	<0.01	18.6 ± 1.0	<0.01
MCM7	10.8 ± 0.0	<0.01	4.4 ± 0.0	<0.01	33.2 ± 3.6	<0.01	16.0 ± 0.6	<0.01
MCM4	8.8 ± 0.4	<0.01	8.7 ± 0.1	<0.01	21.7 ± 1.9	<0.01	20.9 ± 1.0	<0.01
XTP1	7.9 ± 1.2	0.01	14.3 ± 1.7	<0.01	12.7 ± 0.2	<0.01	10.5 ± 1.7	0.02
NASPs	7.3 ± 0.6	<0.01	8.9 ± 0.2	<0.01	16.9 ± 0.7	<0.01	11.0 ± 0.2	<0.01
HAT1	6.8 ± 0.1	<0.01	3.1 ± 0.0	<0.01	15.3 ± 0.5	<0.01	7.5 ± 0.4	<0.01
LBR	9.4 ± 1.5	0.02	3.6 ± 0.6	0.02	15.8 ± 1.0	<0.01	10.5 ± 0.8	<0.01
E2F1	6.7 ± 0.5	<0.01	2.2 ± 0.0	<0.01	57.1 ± 2.8	<0.01	28.5 ± 0.9	<0.01
MGC22679	10.4 ± 1.0	<0.01	1.7 ± 0.2	0.05	4.6 ± 0.1	<0.01	2.1 ± 0.1	<0.01
RRM2	3.0 ± 0.2	<0.01	2.5 ± 0.0	<0.01	19.8 ± 0.6	<0.01	16.5 ± 0.8	<0.01
K1594	6.0 ± 0.7	0.01	6.6 ± 0.1	<0.01	7.2 ± 0.6	<0.01	5.2 ± 0.2	<0.01
Ch-TOG	5.7 ± 0.3	<0.01	2.3 ± 0.0	<0.01	4.9 ± 0.2	<0.01	5.7 ± 0.1	<0.01
K0648	5.3 ± 0.4	<0.01	2.4 ± 0.2	<0.01	3.4 ± 0.3	<0.01	3.7 ± 0.4	0.01
SF3B1	3.5 ± 0.4	0.01	1.0 ± 0.0	* 0.43	7.4 ± 0.6	<0.01	3.7 ± 0.2	<0.01
CML66	3.6 ± 0.1	<0.01	1.8 ± 0.0	<0.01	4.2 ± 0.4	<0.01	4.2 ± 0.3	<0.01
C3orf4	2.3 ± 0.2	0.01	1.6 ± 0.1	<0.01	4.5 ± 0.2	<0.01	4.1 ± 1.0	0.04
SKIN	2.0 ± 0.2	0.02	2.8 ± 0.4	0.02	3.7 ± 0.0	<0.01	7.9 ± 2.0	0.04
G3BP2	3.3 ± 0.3	<0.01	1.5 ± 0.1	0.01	3.3 ± 0.4	<0.01	2.2 ± 0.1	<0.01
TAF3	2.5 ± 0.6	* 0.06	1.1 ± 0.1	* 0.20	3.8 ± 0.1	<0.01	11.6 ± 4.6	* 0.07
PTB2	3.3 ± 0.4	0.02	2.3 ± 0.2	<0.01	2.9 ± 0.2	<0.01	2.7 ± 0.3	0.02
FLJ37562	2.7 ± 0.1	<0.01	0.8 ± 0.4	* 0.34	3.6 ± 0.2	<0.01	3.0 ± 0.5	0.03
BAT1	2.6 ± 0.2	0.01	2.7 ± 0.0	<0.01	3.8 ± 0.1	<0.01	5.4 ± 0.1	<0.01
NSAP1	3.7 ± 0.2	<0.01	2.4 ± 0.0	<0.01	2.2 ± 0.0	<0.01	3.1 ± 0.1	<0.01
TRPC4AP	2.0 ± 0.0	<0.01	0.9 ± 0.3	* 0.38	3.0 ± 0.1	<0.01	1.9 ± 0.3	0.05
SMU-1	2.0 ± 0.0	<0.01	1.4 ± 0.0	<0.01	2.3 ± 0.0	<0.01	2.7 ± 0.1	<0.01

Supplemental Table 2: Validation of E1A-induced genes. The results of several validations, performed by quantitative PCR, on both C2C12 cells and primary mouse satellite cells (MSC) are shown. In detail:
- E1A-C2C12(TD): C2C12 TD myotubes were infected for 24 h with either the dl520 strain (carrying E1A) or the control dl312 strain. Results of Q-RT-PCR are reported as fold increase ("Induction", average of two independent experiments ± SD) in dl520-infected- over dl312-infected TD C2C12.
- Myobl.-C2C12: C2C12 undifferentiated myoblasts were analyzed by Q-RT-PCR for the all of the indicated E1A-induced genes. Results are reported as fold increase ("Induction", average of two independent experiments ± SD) in undifferentiated myoblasts over dl312-infected TD C2C12 (which represent a TD myotube control).
- E1A-MSC(TD): MSC TD myotubes were infected with either the dl520 strain (carrying E1A) or the control dl312 strain. Results of Q-RT-PCR are reported as fold increase ("Induction", average of two independent experiments ± SD) in dl520-infected- over dl312-infected TD MSC.
- Myobl.-MSC: MSC undifferentiated myoblasts were analyzed by Q-RT-PCR for the all of the indicated E1A-induced genes. Results are reported as fold increase ("Induction", average of two independent experiments ± SD) in undifferentiated myoblasts over dl312-infected TD MSC (which represent a TD myotube control).

In all Q-RT-PCRs, the mouse GAPDH gene was used as the internal standard. Statistical significance (p values) was evaluated with unpaired parametric t-test. Asterisks mark non-significant ($p>0.05$) values.

Supplemental Table 3: E1A-induced genes and their timing of induction.

	E1A 24 h		E1A 36 h		Timing		
SHORT NAME	**Induction**	**p value**	**Induction**	**p value**	**Ratio 24h/36h**	**Code**	**p value E/L**
CCNE2	55.6 ± 22.7	0.01	113.5± 41.1	<0.01	0.56 ± 0.19	EARLY	0.05
MCM6	28.9 ± 9.3	<0.01	44.0 ± 7.8	<0.01	0.77 ± 0.04	EARLY	<0.01
NP95	34.9 ± 20.1	0.04	48.3 ± 20.4	0.01	0.75 ± 0.25	EARLY	0.01
CCNE1	18.9 ± 7.1	0.01	20.2 ± 4.8	<0.01	1.02 ± 0.55	EARLY	0.02
MCM7	12.3 ± 3.2	<0.01	15.4 ± 2.3	<0.01	0.84 ± 0.19	EARLY	<0.01
MCM4	13.7 ± 5.2	0.01	22.5 ± 1.3	<0.01	0.60 ± 0.21	EARLY	0.05
XTP1	12.5 ± 1.5	<0.01	30.3 ± 2.8	<0.01	0.43 ± 0.01	EARLY	<0.01
NASPs	6.3 ± 1.3	<0.01	10.4 ± 1.1	<0.01	0.46 ± 0.05	EARLY	0.03
HAT1	6.3 ± 1.6	<0.01	10.4 ± 7.2	* 0.06	0.54 ± 0.31	EARLY	* 0.20
LBR	5.4 ± 2.8	0.04	7.0 ± 2.9	0.01	0.48 ± 0.04	EARLY	0.01
E2F1	6.6 ± 1.6	<0.01	11.5 ± 7.6	0.04	0.67 ± 0.26	EARLY	0.04
MGC22679	9.5 ± 2.1	<0.01	14.9 ± 3.0	<0.01	0.68 ± 0.04	EARLY	<0.01
RRM2	4.9 ± 1.5	0.01	11.0 ± 2.4	<0.01	0.49 ± 0.03	EARLY	<0.01
K1594	3.9 ± 1.4	0.02	5.7 ± 1.8	<0.01	0.47 ± 0.06	EARLY	0.04
Ch-TOG	2.8 ± 2.0	* 0.19	4.9 ± 1.9	0.01	0.19 ± 0.09	LATE	<0.01
K0648	2.8 ± 1.7	* 0.14	5.8 ± 2.8	0.03	0.27 ± 0.17	LATE	* 0.08
SF3B1	2.0 ± 1.0	* 0.18	5.5 ± 0.6	<0.01	0.07 ± 0.04	LATE	<0.01
CML66	2.1 ± 1.0	* 0.12	2.7 ± 0.3	<0.01	0.35 ± 0.03	LATE	0.01
C3orf4	1.6 ± 0.5	* 0.08	3.4 ± 1.7	0.04	0.21 ± 0.08	LATE	<0.01
SKIN	1.5 ± 0.4	* 0.07	2.3 ± 0.3	<0.01	0.27 ± 0.12	LATE	0.04
G3BP2	2.1 ± 0.8	* 0.07	3.5 ± 0.1	<0.01	0.26 ± 0.04	LATE	<0.01
TAF3	1.9 ± 0.6	* 0.06	2.8 ± 0.4	<0.01	0.36 ± 0.12	LATE	* 0.25
PTB2	1.9 ± 1.0	* 0.19	4.4 ± 1.3	<0.01	0.10 ± 0.19	LATE	0.01
FLJ37562	1.9 ± 0.6	* 0.06	3.4 ± 0.1	<0.01	0.19 ± 0.20	LATE	0.04
BAT1	1.7 ± 0.4	* 0.06	3.2 ± 1.7	0.05	0.33 ± 0.02	LATE	<0.01
NSAP1	1.8 ± 1.3	* 0.32	2.3 ± 0.3	<0.01	0.02 ± 0.05	LATE	<0.01
TRPC4AP	1.7 ± 0.4	* 0.06	2,7 ± 0.6	<0.01	0.30 ± 0.10	LATE	0.04
SMU-1	1.4 ± 0.3	* 0.16	2.1 ± 0.1	<0.01	0.28 ± 0.06	LATE	<0.01

Supplemental Table 3: E1A-induced genes and their timing of induction. Columns "24h and 36 h": TD C2C12 myotubes were infected with either the dl520 or the dl312 strain and harvested for RNA extraction 24 h post-infection (the E1A protein starts to accumulate at 16 h) and 36 h post-infection (immediately before S-phase entry). The mRNA levels of all E1A-induced genes were measured by Q-RT-PCR, and are expressed ("Induction" columns) as fold increase (average of three independent experiments ± SD for 24 h analyses, and of two independent experiments ± SD for 36 h analyses) in dl520-infected- over dl312-infected TD C2C12. Statistical significance (p values) was evaluated with unpaired parametric t-test. Asterisks mark non-significant ($p>0.05$) values. Note how the induction of all "early" genes (see below) is significant already at 24 h, while the induction of "late" genes only becomes significant at 36 h.

Columns "Timing". The ratio between the 24 h and the 36 h time points (Ratio 24h/36h) is reported, and was calculated on two independent experiments (average ± SD) in which the two conditions (24 h and 36 h) were analyzed in parallel. "Code" was defined by assigning arbitrarily the code "early" and "late" to ratios of > 0.4, and < 0.4, respectively. The statistical significance of the attribution to the early or late class, with respect to the arbitrary cut-off of 0.4, was evaluated with unpaired parametric t-test (p value E/L). Asterisks mark non-significant ($p>0.05$) values.

Supplemental Table 4: Details of Figure 1A of the main text.

	YH47 24 h		E2F 24 h		RB -/- 48 h	
SHORT NAME	**Ratio**	**Code**	**Ratio**	**Code**	**Ratio**	**Code**
Class A						
XTP1	0.07 ± 0.14		0.53 ± 0.13	*	0.43 ± 0.05	*
CCNE1	0.04 ± 0.03		0.69 ± 0.19		2.99 ± 0.29	
RRM2	-0.05 ± 0.04		0.82 ± 0.06		1.53 ± 0.19	
CCNE2	0.03 ± 0.04		0.73 ± 0.15		1.49 ± 0.15	
E2F1	0.03 ± 0.05		N.A.		1.09 ± 0.11	
Class B						
HAT1	0.07 ± 0.04		0.24 ± 0.04		0.91 ± 0.22	
LBR	0.20 ± 0.10		0.34 ± 0.11	*	2.06 ± 0.17	
NP95	0.01 ± 0.01		0.17 ± 0.07		1.34 ± 0.17	
MGC22679	0.18 ± 0.13		-0.03 ± 0.05		2.39 ± 0.22	
K1594	0.12 ± 0.07		0.02 ± 0.02		0.82 ± 0.05	
NASPs	0.07 ± 0.09		0.10 ± 0.07		0.90 ± 0.11	
MCM7	0.02 ± 0.02		0.07 ± 0.03		1.16 ± 0.17	
MCM4	0.03 ± 0.02		0.11 ± 0.02		0.73 ± 0.05	
MCM6	0.04 ± 0.02		0.14 ± 0.05		0.64 ± 0.07	

	YH47 36 h		E2F 36 h		RB rem. 48 h	
Class C						
C3orf4	0.30 ± 0.14	*	0.07 ± 0.04		1.27 ± 0.08	
G3BP2	0.40 ± 0.13	*	0.06 ± 0.04		0.71 ± 0.06	
TAF3	0.44 ± 0.18	*	0.15 ± 0.03		0.98 ± 0.13	
CML66	0.53 ± 0.04		0.13 ± 0.04		0.60 ± 0.05	
K0648	0.60 ± 0.03		0.41 ± 0.06	*	2.97 ± 0.51	
NSAP1	0.60 ± 0.17		0.30 ± 0.11	*	2.13 ± 0.23	
FLJ37562	0.61 ± 0.11		0.10 ± 0.02		0.75 ± 0.09	
BAT1	0.62 ± 0.28		0.04 ± 0.03		0.80 ± 0.17	
PTB2	0.70 ± 0.10		0.14 ± 0.09		1.65 ± 0.10	
Class D						
SMU-1	0.45 ± 0.15		0.15 ± 0.11		0.29 ± 0.10	
SF3B1	0.47 ± 0.12		0.07 ± 0.03		0.48 ± 0.25	*
ChTOG	0.51 ± 0.16		0.14 ± 0.06		0.28 ± 0.13	
SKIN	0.68 ± 0.31		0.19 ± 0.06		0.01 ± 0.01	
TRPC4AP	0.81 ± 0.26		0.06 ± 0.02		0.18 ± 0.11	

Supplemental Table 4: Details of Figure 1A of the main text. Reported in this Supplemental Table are the actual numbers of the data presented in Figure 1A of the main text. For each gene, results of Q-RT-PCR are reported as ratios ("Ratio", average of three independent experiments ± SD for YH47 and E2F experiments; average of two experiments ± SD for Rb removal) between the tested conditions, and the value obtained with E1A (infection with dl520). In detail, "Ratios" were calculated as follows:

- For YH47 experiments: (YH47 - dl312)/(dl520 - dl312).
- For E2F experiments: (E2F - dl312)/(520 - dl312).
- For Rb -/- experiments: (RB removal – dl312)/(E1A – dl312).

"Code", a color code (red/blue, identical to that shown in Figure 1A) was assigned as follows: For the YH47 and E2F conditions: red/blue, > or < 35% vs. the E1A-induced condition, respectively. For the Rb -/- condition: red/blue, > or < 50% vs. the E1A-induced condition, respectively. Details on how the arbitrary thresholds for classification were set are in Supplemental Tables 5-7. The statistical significance of the attribution to red or blue classes, with respect to the arbitrary cut-offs, was evaluated with unpaired parametric t-test, and asterisks mark non-significant ($p>0.05$) values.

Data reported in this Table are at 24 h for "early" genes, and at 36 h for "late" genes. Complete time kinetics are to be found in Supplemental Tables 5-7.

Note that some genes (XTP1, C3orf4, K0648) are of uncertain classification (and display non-significant p values for classification). Their attribution to the various classes reflects the preponderant features, including "timing" of induction.

N.A., Not applicable, induction of E2F1 was not tested, since E2F1 was ectopically transfected in the cells.

Supplemental Table 5: Details of analyses with the YH47 mutant.

	YH47 24 h		E1A 24 h		Ratio 24 h			YH47 36 h		E1A 36 h		YH47/E1A 36 h		
SH. NAME	**Induc.**	**p**	**Induc.**	**p**	**Ratio**	**p**	**Code**	**Induc.**	**p**	**Induc.**	**p**	**Ratio**	**p**	**Code**
Class A														
XTP1	1.8 ± 0.5	0.05	10.1 ± 4.2	0.02	0.07±0.14	0.02		3.8 ± 1.3	0.02	36.7±13.0	0.01	0.08±0.03	<0.01	
CCNE1	1.6 ± 0.6	* 0.16	21.6 ± 5.8	<0.01	0.04±0.03	<0.01		6.3 ± 3.1	0.04	28.2±10.3	0.01	0.19±0.05	<0.01	
RRM2	0.8 ± 0.0	<0.01	3.9 ± 1.4	0.02	-0.05±0.04	<0.01		1.0 ± 0.1	* 0.65	8.5 ± 0.9	<0.01	-0.01±0.02	<0.01	
CCNE2	3.0 ± 2.0	* 0.15	66.7±10.7	<0.01	0.03±0.04	<0.01		24.8 ± 9.7	0.01	146.1±63.5	0.02	0.17±0.01	<0.01	
E2F1	1.1 ± 0.2	* 0.31	6.1 ± 1.1	<0.01	0.03±0.05	<0.01		2.1 ± 0.2	<0.01	11.3 ± 5.3	0.03	0.14±0.10	<0.01	
Class B														
HAT1	1.4 ± 0.2	0.04	6.1 ± 1.9	0.01	0.07±0.04	<0.01		2.5 ± 1.9	* 0.24	10.8 ± 5.1	0.03	0.12±0.11	<0.01	
LBR	1.6 ± 0.1	<0.01	4.0 ± 0.9	<0.01	0.20±0.10	0.03		2.5 ± 0.5	0.01	8.3 ± 3.0	0.01	0.19±0.06	<0.01	
NP95	1.8 ± 0.2	<0.01	40.2±20.9	0.03	0.01±0.01	<0.01		10.8 ± 6.7	* 0.06	54.8±18.3	0.01	0.17±0.09	<0.01	
MGC22679	2.7 ± 1.7	* 0.16	9.2 ± 2.4	<0.01	0.18±0.13	0.05		5.2 ± 0.3	<0.01	17.7 ± 2.4	<0.01	0.25±0.05	0.01	
K1594	1.3 ± 0.2	* 0.14	3.2 ± 0.5	<0.01	0.12±0.07	<0.01		2.5 ± 0.6	0.01	6.2 ± 1.6	<0.01	0.28±0.05	0.02	
NASPs	1.2 ± 0.1	0.02	5.1 ± 2.6	0.05	0.07±0.09	<0.01		2.6 ± 0.7	0.02	9.2 ± 2.2	<0.01	0.18±0.08	<0.01	
MCM7	1.2 ± 0.2	* 0.14	12.4 ± 2.9	<0.01	0.02±0.02	<0.01		3.5 ± 1.5	0.05	19.7 ± 5.2	<0.01	0.12±0.05	<0.01	
MCM4	1.3 ± 0.2	* 0.06	14.9 ± 4.5	0.01	0.03±0.02	<0.01		4.1 ± 1.1	0.01	21.6 ± 3.7	<0.01	0.16±0.07	<0.01	
MCM6	1.9 ± 0.9	* 0.19	32.9 ± 5.8	<0.01	0.04±0.02	<0.01		10.1 ± 4.5	0.02	49.4 ± 9.9	<0.01	0.19±0.10	0.01	
Class C														
C3orf4	1.2 ± 0.2	* 0.15	1.7 ± 0.4	0.04	0.31±0.13	* 0.28		2.2 ± 1.0	* 0.11	4.4 ± 2.4	* 0.07	0.30±0.14	* 0.26	
G3BP2	1.3 ± 0.0	<0.01	1.9 ± 0.3	0.01	0.37±0.10	* 0.31		2.0 ± 0.3	<0.01	3.6 ± 0.2	<0.01	0.40±0.13	* 0.23	
TAF3	1.5 ± 0.3	0.04	2.0 ± 0.1	<0.01	0.50±0.23	* 0.12		1.8 ± 0.4	0.03	2.9 ± 0.3	<0.01	0.44±0.18	* 0.20	
CML66	1.3 ± 0.1	0.02	1.9 ± 0.1	<0.01	0.34±0.14	* 0.42		2.0 ± 0.3	<0.01	2.9 ± 0.4	<0.01	0.53±0.04	<0.01	
K0648	1.5 ± 0.1	<0.01	2.4 ± 0.2	<0.01	0.35±0.07	* 0.49		4.9 ± 2.1	0.03	7.5 ± 3.5	0.03	0.60±0.03	<0.01	
NSAP1	1.2 ± 0.0	<0.01	1.4 ± 0.1	<0.01	0.45±0.15	* 0.12		1.8 ± 0.3	0.01	2.3 ± 0.2	<0.01	0.60±0.17	0.02	
FLJ37562	1.5 ± 0.2	0.02	1.9 ± 0.0	<0.01	0.57±0.26	* 0.07		2.5 ± 0.2	<0.01	3.4 ± 0.0	<0.01	0.61±0.11	<0.01	
BAT1	1.3 ± 0.1	0.01	1.8 ± 0.3	0.01	0.42±0.19	* 0.25		1.9 ± 0.2	<0.01	2.8 ± 1.3	* 0.08	0.62±0.28	0.05	
PTB2	1.4 ± 0.1	<0.01	2.0 ± 0.4	0.01	0.40±0.10	* 0.17		3.8 ± 0.2	0.02	4.9 ± 1.3	0.01	0.70±0.10	<0.01	
Class D														
SMU-1	1.4 ± 0.2	0.03	1.4 ± 0.1	<0.01	0.85±0.53	0.05		1.5 ± 0.2	0.04	1.9 ± 0.3	0.01	0.45±0.15	0.03	
SF3B1	1.6 ± 0.4	0.06	1.8 ± 0.2	<0.01	0.71±0.47	* 0.09		3.1 ± 0.6	<0.01	5.5 ± 0.4	<0.01	0.47±0.12	0.04	
Ch-TOG	1.5 ± 0.2	0.02	2.3 ± 0.5	0.01	0.43±0.21	* 0.23		2.6 ± 0.3	<0.01	4.3 ± 1.5	0.02	0.51±0.16	0.04	
SKIN	1.4 ± 0.3	* 0.07	1.5 ± 0.5	* 0.16	1.17±0.46	0.01		1.9 ± 0.3	0.01	2.4 ± 0.3	<0.01	0.68±0.31	0.04	
TRPC4AP	1.5 ± 0.3	* 0.06	1.7 ± 0.2	0.01	0.67±0.33	0.05		2.0 ± 0.0	<0.01	2.3 ± 0.3	<0.01	0.81±0.26	<0.01	

Supplemental Table 5: Details of analyses with the YH47 mutant TD C2C12 myotubes were infected with either the YH47 mutant, or the E1A wt dl520 strain, or the dl312 control strain, and harvested for RNA extraction either 24 h or 36 h post-infection. The mRNA levels of all E1A-induced genes were measured by Q-RT-PCR, and are expressed ("Induc." columns, average of three independent experiments ± SD) as fold increase of the condition of interest (either YH47 or E1A) over dl312-infected TD C2C12. Statistical significance (p values) was evaluated with unpaired parametric t-test. Asterisks mark non-significant (p>0.05) values.

"Columns Ratio": For each gene, results of Q-RT-PCR are reported also as ratios ("Ratio", average of three independent experiments ± SD), at the two time points, between the tested conditions, and the value obtained with E1A (infection with dl520). In detail, "Ratios" were calculated as follows: (YH47 - dl312)/(dl520 - dl312). An arbitrary cut off of 0.35 was used to classify the genes by a color code (code); blue, ratio < 0.35 (poorly induced by YH47); red, ratio > 0.35 (well induced by YH47).

The cut-off was set at 0.35 based on a series of control experiments (not shown) in which the induction by YH47 of genes known to be absolutely independent of pocket proteins (such as MHC or MyoD) never exceeded a ratio of 0.7 (70%) with respect to wt E1A (dl520 strain).

The statistical significance of the attribution to red or blue class, with respect to the arbitrary cut-offs, was evaluated with unpaired parametric t-test, and asterisks mark non-significant (p>0.05) values.

Supplemental Table 6: Details of analyses with E2F1 overexpression.

	E2F 24 h		E1A 24 h		Ratio 24 h			E2F 36 h		E1A 36 h		Ratio 36 h		
SH. NAME	**Induc.**	**p**	**Induc.**	**p**	**Ratio**	**p**	**Code**	**Induc.**	**p**	**Induc.**	**p**	**Ratio**	**p**	**Code**
Class A														
XTP1	5.1 ± 1.2	<0.01	9.9 ± 2.1	<0.01	0.53 ± 0.13	* 0.10		4.9 ± 1.9	0.01	17.7 ± 3.7	<0.01	0.23 ± 0.06	0.02	
CCNE1	11.3 ± 3.5	0.01	17.3 ± 5.4	0.01	0.69 ± 0.19	0.02		15.1 ± 2.1	<0.01	31.8±11.7	<0.01	0.48 ± 0.11	0.05	
RRM2	3.5 ± 0.6	<0.01	4.2 ± 1.0	<0.01	0.82 ± 0.06	<0.01		3.9 ± 0.1	<0.01	8.0 ± 0.8	<0.01	0.42 ± 0.06	0.05	
CCNE2	35.9±13.1	0.01	52.6±25.8	0.02	0.73 ± 0.15	0.01		38.6±14.8	<0.01	86.5±22.7	<0.01	0.43 ± 0.06	0.04	
E2F1	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Class B														
HAT1	2.5 ± 0.3	<0.01	7.0 ± 0.9	<0.01	0.24 ± 0.04	<0.01		2.7 ± 0.8	0.01	10.2 ± 6.8	0.03	0.21 ± 0.06	0.01	
LBR	2.8 ± 1.4	* 0.10	6.2 ± 2.8	0.03	0.34 ± 0.11	* 0.27		1.9 ± 0.8	* 0.06	6.9 ± 2.9	0.01	0.21 ± 0.23	* 0.16	
NP95	6.5 ± 1.9	0.01	40.4±20.7	0.03	0.17 ± 0.07	<0.01		8.0 ± 2.9	<0.01	47.3±19.0	<0.01	0.18 ± 0.14	0.05	
MGC22679	0.8 ± 0.3	* 0.27	8.8 ± 1.9	<0.01	-0.03 ± 0.05	<0.01		1.5 ± 1.5	* 0.55	14.9 ± 3.0	<0.01	0.02 ± 0.10	<0.01	
K1594	1.1 ± 0.2	* 0.19	4.3 ± 1.5	0.02	0.02 ± 0.02	<0.01		1.3 ± 0.4	* 0.22	5.6 ± 2.0	<0.01	0.09 ± 0.13	0.01	
NASPs	1.5 ± 0.5	* 0.19	6.7 ± 1.2	<0.01	0.10 ± 0.07	<0.01		1.4 ± 0.5	* 0.26	8.2 ± 2.0	<0.01	0.06 ± 0.08	<0.01	
MCM7	1.6 ± 0.2	0.01	12.8 ± 2.4	<0.01	0.07 ± 0.03	<0.01		4.0 ± 1.0	<0.01	18.7 ± 7.0	<0.01	0.17 ± 0.02	<0.01	
MCM4	2.1 ± 0.4	0.01	14.6 ± 5.0	0.01	0.11 ± 0.02	<0.01		3.8 ± 2.8	* 0.07	19.8 ± 2.6	<0.01	0.14 ± 0.13	0.03	
MCM6	3.7 ± 0.4	<0.01	25.4 ± 7.3	<0.01	0.14 ± 0.05	<0.01		4.6 ± 2.8	0.04	43.8 ± 7.5	<0.01	0.08 ± 0.05	<0.01	
Class C														
C3orf4	1.1 ± 0.1	* 0.15	1.9 ± 0.5	0.04	0.09 ± 0.07	<0.01		1.3 ± 0.4	* 0.21	4.9 ± 3.8	* 0.07	0.07 ± 0.04	<0.01	
G3BP2	1.1 ± 0.2	* 0.34	2.4 ± 0.8	0.04	0.05 ± 0.12	0.01		1.2 ± 0.1	* 0.26	3.6 ± 0.3	<0.01	0.06 ± 0.04	<0.01	
TAF3	1.1 ± 0.1	* 0.11	2.2 ± 0.3	<0.01	0.07 ± 0.04	<0.01		1.3 ± 0.1	0.02	3.1 ± 0.1	<0.01	0.15 ± 0.03	<0.01	
CML66	1.1 ± 0.1	* 0.07	2.5 ± 1.0	* 0.06	0.06 ± 0.04	<0.01		1.3 ± 0.1	0.05	3.1 ± 0.2	<0.01	0.13 ± 0.04	<0.01	
K0648	2.1 ± 0.8	* 0.08	3.3 ± 1.7	* 0.07	0.48 ± 0.16	* 0.12		3.8 ± 2.4	* 0.06	7.3 ± 4.9	0.04	0.41 ± 0.06	* 0.07	
NSAP1	1.4 ± 0.4	* 0.21	2.5 ± 1.1	* 0.07	0.19 ± 0.15	* 0.06		1.4 ± 0.2	* 0.07	2.2 ± 0.1	<0.01	0.30 ± 0.11	* 0.23	
FLJ37562	1.2 ± 0.2	* 0.27	2.2 ± 0.4	0.01	0.14 ± 0.23	* 0.09		1.2 ± 0.1	* 0.06	3.4 ± 0.1	<0.01	0.10 ± 0.02	<0.01	
BAT1	1.2 ± 0.2	* 0.37	2.3 ± 0.3	<0.01	0.06 ± 0.25	0.05		1.1 ± 0.1	* 0.97	3.3 ± 1.6	0.03	0.04 ± 0.03	<0.01	
PTB2	1.1 ± 0.2	* 0.31	2.3 ± 0.8	0.05	0.15 ± 0.15	* 0.08		1.5 ± 0.2	0.02	4.7 ± 1.7	0.01	0.14 ± 0.09	<0.01	
Class D														
SMU-1	1.1 ± 0.2	* 0.23	1.5 ± 0.4	* 0.10	0.20 ± 0.15	* 0.07		1.3 ± 0.1	* 0.08	2.2 ± 0.1	<0.01	0.15 ± 0.11	0.02	
SF3B1	1.2 ± 0.2	* 0.18	2.5 ± 0.9	0.04	0.19 ± 0.23	* 0.15		1.3 ± 0.1	0.03	5.8 ± 0.2	<0.01	0.07 ± 0.03	<0.01	
Ch-TOG	1.8 ± 0.4	0.02	3.7 ± 1.8	* 0.06	0.32 ± 0.06	* 0.19		1.6 ± 0.5	* 0.07	4.8 ± 2.0	<0.01	0.14 ± 0.06	<0.01	
SKIN	1.2 ± 0.2	* 0.33	2.0 ± 0.3	<0.01	0.15 ± 0.21	* 0.09		1.5 ± 0.1	< 0.01	2.6 ± 0.3	<0.01	0.19 ± 0.06	0.01	
TRPC4AP	1.1 ± 0.1	* 0.43	1.8 ± 0.3	0.01	0.02 ± 0.15	0.01		1.2 ± 0.1	* 0.15	2.9 ± 0.3	<0.01	0.06 ± 0.02	<0.01	

Supplemental Table 6: Details of analyses with E2F1 overexpression.TD C2C12 myotubes were infected with either Ad-E2F1, or the E1A wt dl520 strain, or the dl312 control strain, and harvested for RNA extraction either 24 h or 36 h post-infection. The mRNA levels of all E1A-induced genes were measured by Q-RT-PCR, and are expressed ("Induc." columns, average of three independent experiments ± SD) as fold increase of the condition of interest (either E2F or E1A) over dl312-infected TD C2C12. Statistical significance (p values) was evaluated with unpaired parametric t-test. Asterisks mark non-significant (p>0.05) values.

"Columns Ratio": For each gene, results of Q-RT-PCR are reported also as ratios ("Ratio", average of three independent experiments ± SD), at the two time points, between the tested conditions, and the value obtained with E1A (infection with dl520). In detail, "Ratios" were calculated as follows: (E2F - dl312)/(dl520 - dl312). An arbitrary cut off of 0.35 was used to classify the genes by a color code (code); blue, ratio < 0.35 (poorly induced by E2F); red, ratio > 0.35 (well induced by E2F).

The cut-off was set at 0.35 based on a series of control experiments (not shown) in which the induction by E2F1 of genes known to be absolutely dependent on E2F1 (such as CycE or CycD) never exceeded a ratio of 0.7 (70%) with respect to wt E1A (dl520 strain).

The statistical significance of the attribution to red or blue class, with respect to the arbitrary cut-offs, was evaluated with unpaired parametric t-test, and asterisks mark non-significant (p>0.05) values.

N.A., Not applicable, induction of E2F1 was not tested, since E2F1 was ectopically transfected in the cells.

Supplemental Table 7: Details of analyses with Rb removal in TD myotubes.

	Rb -/-		E1A		Ratio Rb -/-/E1A		
SHORT NAME	**Induction**	**p value**	**Induction**	**p value**	**Ratio**	**p value**	**Code**
Class A							
XTP1	6.1 ± 0.7	0.02	12.7 ± 0.2	< 0.01	0.43 ± 0.05	* 0.10	
CCNE1	86.6 ± 3.8	< 0.01	29.6 ± 4.0	0.01	2.99 ± 0.29	< 0.01	
RRM2	29.7 ± 0.6	< 0.01	19.8 ± 2.0	0.01	1.53 ± 0.19	0.01	
CCNE2	125.4 ± 7.1	< 0.01	84.5 ± 13.4	0.01	1.49 ± 0.15	0.01	
E2F1	62.7 ± 3.1	< 0.01	57.1 ± 2.8	< 0.01	1.09 ± 0.11	0.01	
Class B							
HAT1	14.0 ± 1.4	0.01	15.3 ± 1.9	0.01	0.91 ± 0.21	0.05	
LBR	31.5 ± 2.5	< 0.01	15.8 ± 2.4	0.01	2.06 ± 0.17	< 0.01	
NP95	72.3 ± 6.0	< 0.01	54.2 ± 2.2	< 0.01	1.34 ± 0.17	0.01	
MGC22679	9.7 ± 0.2	< 0.01	4.6 ± 0.4	0.01	2.39 ± 0.22	< 0.01	
K1594	6.1 ± 0.2	< 0.01	7.2 ± 0.6	< 0.01	0.82 ± 0.05	< 0.01	
NASPs	15.2 ± 1.1	< 0.01	16.9 ± 0.8	< 0.01	0.90 ± 0.11	0.02	
MCM7	38.4 ± 2.8	< 0.01	33.2 ± 2.2	< 0.01	1.16 ± 0.17	0.01	
MCM4	16.1 ± 0.4	< 0.01	21.7 ± 1.9	< 0.01	0.73 ± 0.05	0.01	
MCM6	45.6 ± 0.2	< 0.01	70.3 ± 8.1	0.01	0.64 ± 0.07	0.05	
Class C							
C3orf4	5.5 ± 0.1	< 0.01	4.5 ± 0.3	< 0.01	1.27 ± 0.08	< 0.01	
G3BP2	2.7 ± 0.1	< 0.01	3.3 ± 0.4	0.01	0.71 ± 0.06	0.02	
TAF3	3.7 ± 0.1	< 0.01	3.8 ± 0.3	< 0.01	0.98 ± 0.13	0.02	
CML66	2.9 ± 0.1	< 0.01	4.2 ± 0.3	< 0.01	0.60 ± 0.05	0.05	
K0648	8.1 ± 0.1	< 0.01	3.4 ± 0.4	0.01	2.97 ± 0.51	0.01	
NSAP1	3.6 ± 0.1	< 0.01	2.3 ± 0.2	0.02	2.13 ± 0.23	0.01	
FLJ37562	3.0 ± 0.2	0.01	3.6 ± 0.1	< 0.01	0.75 ± 0.09	0.03	
BAT1	3.2 ± 0.2	< 0.01	3.8 ± 0.4	0.01	0.80 ± 0.16	0.05	
PTB2	4.1 ± 0.2	< 0.01	2.9 ± 0.2	0.01	1.65 ± 0.10	< 0.01	
Class D							
SMU-1	1.4 ± 0.1	* 0.06	2.3 ± 0.1	< 0.01	0.29 ± 0.10	0.05	
SF3B1	4.2 ± 1.6	* 0.11	7.4 ± 0.7	0.01	0.48 ± 0.25	* 0.46	
ChTOG	2.1 ± 0.2	0.02	4.8 ± 0.9	0.03	0.28 ± 0.11	0.05	
SKIN	1.1 ± 0.1	* 0.89	3.7 ± 0.1	< 0.01	0.01 ± 0.01	< 0.01	
TRPC4AP	1.4 ± 0.2	* 0.11	3.0 ± 0.2	0.01	0.18 ± 0.11	0.03	

Supplemental Table 7: Details of analyses with Rb removal in TD myotubes. Primary $Rb^{loxP/loxP}$ muscle satellite cells (MSC-$Rb^{loxP/loxP}$) were isolated from two-week-old mice of the $Rb^{loxP/loxP}$ genotype. After induction of terminal differentiation, the Rb gene was removed by infection with the Ad-Cre virus (Rb -/- condition). Parallel plates were infected with the E1A wt dl520 strain or the dl312 control strain. RNA was extracted 48 h after the infections. The mRNA levels of all E1A-induced genes were measured by Q-RT-PCR, and are expressed ("Induction" columns, average of two independent experiments ± SD) as fold increase of the condition of interest (either Rb -/- or E1A) over dl312-infected cells. Statistical significance (p values) was evaluated with unpaired parametric t-test. Asterisks mark non-significant (p>0.05) values.

"Columns Ratio Rb -/-/E1A": For each gene, results of Q-RT-PCR are reported also as ratios ("Ratio", average of two independent experiments ± SD), between the tested conditions, and the value obtained with E1A (infection with dl520). In detail, "Ratios" were calculated as follows: (Ad-Cre - dl312)/(dl520 - dl312). An arbitrary cut off of 0.5 was used to classify the genes by a color code (code); blue, ratio < 0.5 (poorly induced by Rb removal); red, ratio > 0.5 (well induced by Rb removal).

The cut-off was set at 0.5 based on a series of control experiments (not shown) in which the induction by Rb removal of genes known to be absolutely dependent on Rb (such as CycE or Cdc6) was shown to reach (or to exceed) a ratio of 1.0 (100%) with respect to wt E1A (dl520 strain).

The statistical significance of the attribution to red or blue class, with respect to the arbitrary cut-offs, was evaluated with unpaired parametric t-test, and asterisks mark non-significant (p>0.05) values.

Supplemental Table 8: Correlation between ISH and Q-RT-PCR for ErbB2 levels in breast cancer patients.

Case ID	ISH score	Q-PCR	Q-PCR score	Q-PCR range
1	0	1	0	
2	0	3	0	
3	1	3	0	**0<x<10**
4	0	4	0	
5	0	5	0	
6	0	5	0	
7	1	12	1	**10<x<20**
8	1	12	1	
9	1	13	1	
10	0	15	1	
11	1	16	1	
12	1	17	1	
13	1	19	1	
14	2	22	2	**20<x<100**
15	2	43	2	
16	2	74	2	
17	2	86	2	
18	3	213	3	**x>100**
19	3	259	3	
20	3	362	3	

Supplemental Table 8: Correlation between ISH and Q-RT-PCR for ErbB2 levels in breast cancer patients. This experiment was set up to demonstrate that there is good semi-quantitative correlation between the levels of expression detected by ISH-TMA, and those evidenced by a highly quantitative method, such as Q-RT-PCR. We compared the two methods on levels of ErbB-2 expression in breast cancer, since ErbB2 is overexpressed in breast cancers, over a wide range of levels.

A cohort of breast cancer patients for which Q-RT-PCR data on the levels of expression of ErbB2 were available was used. From this cohort (consisting of 69 cases), 20 cases were selected which displayed ErbB-2 levels ranging from 1 to >300 ("Q-PCR" column, values are normalized to those of Patient 1, assumed as 1.0). Cores from paraffin blocks, corresponding to these 20 cases, were then arrayed on a TMA (in duplicate) and subjected to ISH with an ErbB-2-specific riboprobe. ISH scores were assigned as described in Methods (column "ISH score"). Since data from ISH-TMA analysis are discontinuous, whereas those obtained by Q-RT-PCR analyses are continuous, we transformed the Q-RT-PCR values into discontinuous data (Ginestier, C. et al., Am J Pathol 2002, 161:1223–1233). The discontinuous Q-RT-PCR scores ("Q-PCR score" column) are shown along with the ranges of actual values used to attribute the score ("Q-PCR range" column). Pearson Correlation was calculated in Excel Software (Microsoft). The Pearson Correlation Coefficient or the ISH score vs. the Q-RT-PCR score was 0.95.

Supplemental Table 9: Engineering of the TMAs used in Figure 3A.

A. Composition of the TMAs employed in Figure 3A.

	Breast	Colon	Kidney	Larynx	Lung	Prostate	Stomach	Uterus	Melan.	Brain
TMA 1 T/N	19/9	14/11			10/2	25/25				
TMA 2 T/N			20/13	28/26	28/28		20/20			
TMA 3 T/N			8/8				13/8	28/0	17/0	19/0
TMA 4 T/N	12/11	14/14		18/15	15/14					
TMA 5 T/N	18/15	19/11			20/5		12/6			

B. Genes/TMAs

	SF3B1	Ch-TOG	SKIN	TRCP4AP	SMU-1
TMA 1	X	X	X	X	X
TMA 2	X			X	X
TMA 3	X	X	X	X	X
TMA 4		X	X		
TMA 5			X		

C. Total number of cases analyzed/gene (T/N, Tumor/Normal)

	Breast	Colon	Kidney	Larynx	Lung	Prostate	Stomach	Uterus	Melan.	Brain
SF3B1 (T/N)	19/9	14/11	28/21	28/26	38/30	25/25	33/28	28/0	17/0	19/0
Ch-Tog (T/N)	31/20	28/25	8/8	18/15	25/16	25/25	13/8	28/0	17/0	19/0
SKIN (T/N)	49/35	47/36	8/8	18/15	45/21	25/25	25/14	28/0	17/0	19/0
TRCP4AP (T/N)	19/9	14/11	28/21	28/26	38/30	25/25	33/28	28/0	17/0	19/0
SMU-1 (T/N)	19/9	14/11	28/21	28/26	38/30	25/25	33/28	28/0	17/0	19/0

D. Total number of cases scored/gene (tumor samples only) (S/A, Scored/Analyzed)

	Breast	Colon	Kidney	Larynx	Lung	Prostate	Stomach	Uterus	Melan.	Brain
SF3B1 (S/A)	13/19	13/14	13/28	18/28	31/38	20/25	26/33	19/28	12/17	15/19
Ch-TOG (S/A)	19/31	24/28	7/8	9/18	22/25	16/25	12/13	21/28	13/17	19/19
SKIN (S/A)	47/49	33/47	8/8	17/18	41/45	20/25	23/25	25/28	17/17	16/19
TRCP4AP (S/A)	16/19	13/14	25/28	23/28	36/38	19/25	30/33	18/28	8/17	13/19
SMU-1 (S/A)	15/19	13/14	21/28	27/28	31/38	21/25	30/33	17/28	8/17	15/19

Supplemental Table 9: Engineering of the TMAs used in Figure 3A.
A. Five different TMAs (TMA 1 through 5) were engineered. In each column, the number of cases, for each type of tumors and matched controls, deposited on individual TMAs is reported. In each box the first number indicates the number of tumor cases, and the second the number of normal matched samples (T/N). As indicated, normal counterparts were not always available. Each case was deposited in duplicate.
B. The Table shows which TMA was used for each of the indicated Class-D genes (indicated by Xs).
C. The Table shows the total number of cases (tumors and normal matched controls) analyzed per each indicated Class-D gene, and it is obtained combining information shown in **A** and **B**. In each box, the first number indicates the number of tumor cases, and the second the number of normal matched samples analyzed per each gene (T/N).
D. The Table shows the total number of cases scored per each gene. Frequently, during the manipulations for *in situ* hybridization, some cores detached from the slide and were lost. When one of the two duplicate cores (or both) was lost, we did not score the sample. Thus, the number of scored samples is actually lower than the number of analyzed samples. In each box the first number indicates the number of cases scored, and the second the number of cases analyzed (S/A). Numbers are provided for tumor samples only.

Supplemental Table 10: Details of Figure 3A of the main text.

A. Percent overexpressing tumors

	Breast	Colon	Kidney	Larynx	Lung	Prostate	Stomach	Uterus	Melan.	Brain
SF3B1	23%	77%	0%	6%	35%	0%	12%	21%	0%	7%
Ch-TOG	21%	71%	0%	44%	18%	0%	50%	19%	15%	32%
SKIN	21%	87%	0%	71%	29%	0%	39%	36%	88%	13%
TRCP4AP	56%	54%	4%	4%	28%	5%	37%	11%	63%	0%
SMU-1	27%	31%	0%	7%	13%	0%	10%	12%	0%	0%

B. Statistical analysis

	Breast	Colon	Kidney	Larynx	Lung	Prostate	Stomach	Uterus	Melan.	Brain
SF3B1	* 0.18	0.02	N.D.	* 0.31	0.03	N.D.	* 0.87	N.D.	N.D.	N.D.
Ch-TOG	0.05	<0.01	N.D.	0.04	* 0.12	N.D.	<0.01	N.D.	N.D.	N.D.
SKIN	<0.01	<0.01	N.D.	<0.01	<0.01	N.D.	0.02	N.D.	N.D.	N.D.
TRCP4AP	* 0.08	0.01	* 0.34	* 0.64	0.03	* 0.33	<0.01	N.D.	N.D.	N.D.
SMU-1	* 0.14	0.05	N.D.	* 0.22	* 0.14	N.D.	* 0.11	N.D.	N.D.	N.D.

Supplemental Table 10: Details of Figure 3A of the main text.
A. Reported in this Supplemental Table are the actual numbers of the data presented in Figure 3A of the main text. Overexpression was calculated as indicated in Methods and is reported as % of positive cases.
B. Statistical analysis was performed by Contingency Table analysis with Likelihood-Ratio Chi-squared (G^2) (JMP™ IN 5.1). Asterisks mark non-significant ($p>0.05$) values. N.D., Not done (in the case of uterus, melanoma and brain for lack of normal samples, in the other cases for lack of overexpressing samples).

Supplemental Table 11: Oncomine analysis of Class-D genes.

SHORT NAME	PUBMED ID	N (log2 median)	T (log2 median)	p value	site	N total	T total
SMU-1	*11742071*	-0.989	-0,614	0.02	multi-cancer	90	218
SF3B1	*11742071*	-0.371	1.163	0.04	pancreas	21	11
ch-TOG	*11707567*	-0.638	0.083	< 0.01	lung	17	139
ch-TOG	*12058060*	-1.156	-0.825	< 0.01	liver	76	104
ch-TOG	*12368205*	-0.678	0.277	< 0.01	salivary	6	16

Symbol	PUBMED ID	T (log2 median)	M (log2 median)	p value	site	T total	M total
SF3B1	*12154061*	-0.066	0.138	< 0.01	prostate	23	9
SF3B1	*11707567*	-0.520	-0.213	0.02	lung	123	16
ch-TOG	*11707567*	0.056	0.288	0.01	lung	123	16

Supplemental Table 11: Oncomine analysis of Class-D genes. The results of a meta-analysis regarding class D genes is reported. We used the ONCOMINE web tool (www.oncomine.org) to check for significant regulation of Class-D genes in published expression profile experiments. The genes which pass the statistical filter (p value < 0.05 with Bonferroni correction) are shown together with their $\log_2$ median value in every class considered, where: "N" stands for normal samples; "T" for primary tumors and "M" for distant metastasis.

SKIN (flj23790) analysis failed since the specific probeset is present only in some of the more recent array versions (Affymetrix HG-U133 chip B and HG-U95 chip B) therefore drastically reducing the database size. TRPC4AP did not reach statistical significance.

Supplemental Table 12: Correlation between overexpression of Class-D genes and various clinical and biological parameters.

A. Correlation between overexpression of Class-D genes in various tumors (data from Figure 3A) and Ki67 levels.

SHORT NAME	Ki 67 (p)
SF3B1	* 0.09
Ch-TOG	* 0.14
SKIN	* 0.19
TRCPA4	<0.01
SMU-1	* 0.12

B. Correlation between overexpression of Class-D genes in colon cancer (data from Figure 4A) and various biological and clinical parameters.

SHORT NAME	pN (p)	A/N (p)	T/N (p)	T/A (p)
SF3B1	* 0.60	0.01	< 0.01	* 0.95
Ch-TOG	* 0.20	* > 0.1	< 0.01	< 0.01
SKIN	* 0.27	* > 0.1	< 0.01	< 0.01
TRCPA4	* 0.60	* > 0.1	< 0.01	< 0.01
SMU-1	* 0.09	* > 0.1	< 0.01	< 0.01

Supplemental Table 12: Correlation between overexpression of Class-D genes and various clinical and biological parameters. In both **A** and **B**, statistical analysis was performed by Contingency Table analysis with Likelihood-Ratio Chi-squared (G^2) (JMP™ IN 5.1). pN, lymph node positivity. P vales (p) are also shown for overexpression of: adenomas vs normal tissues (A/N), tumors vs. normal tissues (T/N), and tumors vs. adenomas (T/N). Asterisks mark non-significant (p>0.05) values.

Supplemental Table 13: Details of the Affymetrix screening shown in Figure 5A.

A. Clinical data.

	Patient ID															
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Follow up	7.6	1.6	8.3	7.7	8.2	8.8	8.6	8.5	7.8	12	8.3	9.8	9.6	9.7	10	2.5
Age (years)	59	55	59	53	46	68	50	38	59	54	36	47	45	43	53	70
Metastasis	no	yes	no	no	no	no	no	no	no	no	no	no	no	no	no	yes

	17	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**
Follow up	1.8	7.8	1.3	9.4	0.5	2	2.9	4.8	0.8	0.9	1.3	2.7	4.4	4.4	0.8	8
Age (years)	52	60	60	54	63	59	78	62	62	67	61	78	39	51	69	57
Metastasis	yes	no	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no

	33	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**
Follow up	8	7.9	7.8	7.6	3.1	9.2	1.8	7.8	1.8	3.8	8	4	8	8
Age (years)	60	67	69	45	59	50	58	69	70	56	64	40	53	57
Metastasis	no	no	no	no	yes	no	yes	no	yes	yes	no	yes	no	no

B. Affymetrix data.

			patient ID											
Gene Symbol	**Probeset HG-U133**	**Repr. EST**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
SMU-1	233946_at	AL512690	1.684	0.649	2.09	1.859	1.66	1.998	2.619	1.713	1.498	1.161	2.89	2.002
SKIN	226294_x_at	AI245517	11.44	31.71	9.25	11.37	6.951	13.33	24.91	28.95	14.98	9.395	17.17	21.96
ch-TOG	212832_s_at	D43948	4.766	9.501	6.06	6.586	4.163	6.123	11.93	7.072	6.746	6.098	5.543	5.703
TRPC4AP	212059_s_at	AL096738	0.692	0.375	0.493	1.101	0.904	1.165	1.003	0.69	0.642	0.671	0.387	0.491
SF3B1	201071_x_at	AF054284	27.09	28.7	26.17	34.32	29.23	23.05	19.43	29.26	27.77	38.62	37.75	25.47

Gene Symbol	**Probeset HG-U133**	**Repr. EST**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
SMU-1	233946_at	AL512690	1.372	0.811	2.662	2.722	1.866	1.466	0.952	1.274	1.212	1.869	3.771	1.628
SKIN	226294_x_at	AI245517	40.12	9.187	31.08	28.76	27.9	13.95	11.63	16.52	18.83	11.19	22.64	34.55
ch-TOG	212832_s_at	D43948	6.651	2.666	3.657	8.296	4.306	4.559	4.893	2.419	5.099	4.532	6.23	6.243
TRPC4AP	212059_s_at	AL096738	0.192	1.383	1.063	0.527	0.585	0.666	0.819	0.429	0.567	0.908	0.888	0.333
SF3B1	201071_x_at	AF054284	35.44	32.66	20.51	27.37	35.1	38.04	34.33	37.29	28.09	30.71	39.52	40.77

Gene Symbol	**Probeset HG-U133**	**Repr. EST**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
SMU-1	233946_at	AL512690	1.801	1.949	0.894	2.091	0.949	2.075	1.581	0.967	0.645	0.823	0.722	1.601
SKIN	226294_x_at	AI245517	17.33	19.31	35.02	15.35	11.48	10.37	30.85	9.337	14.32	9.343	11.89	12.18
ch-TOG	212832_s_at	D43948	6.953	7.735	7.34	6.267	7.208	4.986	8.241	4.911	5.841	5.928	5.735	4.805
TRPC4AP	212059_s_at	AL096738	0.715	0.779	0.659	0.567	0.657	0.522	0.172	1.242	1.503	0.918	0.749	0.996
SF3B1	201071_x_at	AF054284	44.99	31.15	37.26	40.08	31.79	37.41	34.48	42.86	35.89	27.75	20.12	30.19

Gene Symbol	**Probeset HG-U133**	**Repr. EST**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**
SMU-1	233946_at	AL512690	0.162	1.75	1.225	1.225	0.836	0.8	1.169	1.532	0.782	2.871
SKIN	226294_x_at	AI245517	10.56	10.42	9.164	6.904	14.73	12.43	16.92	20.36	4.174	13.32
ch-TOG	212832_s_at	D43948	5.921	5.669	3.785	2.946	7.64	3.885	4.818	5.901	1.435	3.677
TRPC4AP	212059_s_at	AL096738	0.306	1.146	0.356	0.205	0.114	0.746	0.831	0.313	1.004	1.122
SF3B1	201071_x_at	AF054284	24.32	35.31	35.87	25.29	26.04	28.5	48.26	33.35	8.531	25.63

Supplemental Table 13: Details of the Affymetrix screening shown in Figure 5A. A. The relevant clinical information is provided for the 46 patients used in the study (all ER+). Follow up is in years. **B.** The Affymetrix data for all class D genes, in all 46 patients, are shown. Data are reported per chip-normalized.