Gendelman et al. - Supplementary methods.

Real-Time Quantitative RT-PCR. Total RNA was extracted from cells at 24h and 72h post-siRNA transfection using RNeasy Micro kit (Qiagen). It was reverse-transcribed to cDNA and quantitative RT-PCR analysis using the Taqman assay (ABI) was performed at Genome Analysis Core Facility of Helen Diller Family Comprehensive Cancer Center, UCSF. PCR primers and TaqMan probes for CCND1, TRIB1, IER2, CDKN2C, NUAK1, C14ORF133, CCNE2, TBK1, EGR1, NPC1, SPRED2, KIAA0649, DR5, and YY1 were purchased from Applied Biosystems. hGUS was used as a normalization control. PCR was conducted in triplicate with 20 μL reaction volumes of 1X Taqman buffer (1X Applied Biosystems PCR buffer, 20% glycerol, 2.5% gelatin, 60nM Rox as a passive reference), 5.5 mM MgCl2, 0.5 mM each primer, 0.2 mM each deoxynucleotide triphosphate (dNTP), 200 nM probe, and 0.025 unit/uL AmpliTaq Gold (Applied Biosystems) with 5ng cDNA. PCR was conducted on the ABI 7900HT (Applied Biosystems) using the following cycle parameters: 1 cycle of 95° for 10 minutes and 40 cycles of 95° for 15 seconds, 60° for 1 minute. Analysis was carried out using the SDS software (version 2.3) supplied with the ABI 7900HT to determine the Ct values of each reaction. Ct values were determined for three test and three reference reactions (Histone 3.3) in each sample, averaged, and subtracted to obtain the ΔCt [ΔCt = Ct (test locus) – Ct (control locus)]. Relative percent expression was calculated for each sample as $2^{-\Delta Ct} \times 100$. Final relative transcript levels in each sample were obtained by normalization with NC-siRNA.

Assessment of model accuracy. To assess the accuracy of the model in learning the cell cycle distribution from gene expression values, the gene expression data used in training were used in Monte Carlo simulations to determine whether the network could recover the phenotypic values, in this case fraction of cells in the G1 phase of the cell cycle, from training data (Fig. S2A). The above procedure was then applied to every gene in the network. Specifically, Monte Carlo simulations were performed to determine whether the network could predict the gene expression intensity of a transcript based on all other nodes in the network. Fig. S2B shows predicted gene expression values versus observed gene expression values. The Pearson correlation coefficient of predicted and observed gene expression values for each gene were computed to assess how well the model learned and was able to predict that gene. A higher value of Pearson correlation coefficient indicates better learning. The distribution of Pearson correlation coefficients for all genes (Fig. S2C) suggests that the model learned well from the data and had excellent predictive power for the majority of genes.

Gene expression, copy number, and survival analysis of primary tumors. Differences in survival distributions for various subgroups based on the expression levels of relevant genes were assessed using Kaplan-Meier curves and the nonparametric log-rank test implemented in the *survival* package in the R statistical computing language. In order to visualize the survival effects for various single gene models in an exploratory analysis, the expression levels of several key genes were stratified into tertiles, and the predicted survival curves for individuals in the upper (66.66%) and lower (33.33%) groups were compared using Kaplan-Meier plots and the log-rank test. A total of 20 cases out of 1980 (1%) were lost to follow-up, and hence excluded from the survival analyses. We also examined survival as a function of continuous gene expression data for both single gene and multiple gene models based on the

panel of 38 TRIB1/TRAIL/NFκB associated genes by Cox regression using the *Design* package in R. Multivariable analysis was performed by building stratified Cox proportional hazards models, which included the most relevant clinical variables as covariates, namely, grade (numerical, linear), tumor size (numerical, spline 3 knots), number of lymph nodes positive for disease (numerical, spline 4 knots), and age (numerical, spline 3 knots) in order to test for differences in BCSS and OS. Restricted cubic spline terms were used for the more complex numerical variables as indicated above. Models were stratified by tumor bank (site) to account for differences in the basal hazard due to geographical location, as well as by ER status, which is known to violate the proportional hazards assumption. Separate models were also built for ER-positive and ER-negative cases to evaluate differences in outcome for these subgroups. The proportional hazards assumption is also likely to be violated for other predictors over prolonged survival, but as these analyses are exploratory in nature we do not consider time-dependent effects here. Moreover, we note that the effect of violating the proportional hazards assumption is loss in power to detect real differences, hence some effects may be underestimated, and estimates may be erroneous in certain time intervals. The Wald statistics, summary of effects of predictors, and 95% confidence intervals are reported for each model based on the *summary* function in the *Design* package.