Supplemental Figure Legends

Supplemental Figure 1. Principal component analysis of expression estimates for pathway perturbation data. A. The samples are colored by pathway: ATM (red), BCL2 (blue), GPX1 (purple), IKBKB (orange), NOS2 (pink), SIRT1 (gray), and GFP (black). B. The samples are colored by batch: batch 1 (red), batch 2 (blue), batch 3 (black), batch 4 (orange). Samples in batch 3 were eliminated from the analyses in the paper.

Supplemental Figure 2. Pathway activation predictions for each control dataset. A. ATM pathway activation is increased in H1299 and transiently decreased and then restored in H460 cells exposed to 2 Gy of ionizing radiation (IR). RNA from the cells was sampled pre-exposure (n=12) and at various time points post-exposure (n=30) and hybridized to Affymetrix Gene 1.0 ST microarray (GSE20549). B. GPX1 pathway activation is decreased in response to a selenium deficient diet. Rats were fed selenium deficient diets (n=3) or supplemented with sodium selenite up to 5ug/g for 28 days (n=15). Rat livers were harvested and RNA was isolated and hybridized to Affymetrix Rat Genome 230 2.0 microarrays (GSE23895). C. IKBKB pathway activation is increased after retrovirus infection with the IKBKB. H929 and SACHI cells lines were infected with retroviruses containing MAP3K14 (n=6), IKBKB (n=6) compared with controls were no infection was conducted (n=6). RNA was isolated and hybridized to Affymetrix HGU133 Plus 2.0 microarrays (GSE18047). D. SIRT1 pathway activation is increased in A549 cells treated with resveratrol. A549 cells treated with an ethanol control (n=1) or 25 microM of resveratrol (n=3) for 48h and RNA was hybridized to Affymetrix HGU133 Plus 2.0 microarrays (GSE9008). E. SIRT1 pathway activity is decreased in the SIRT1 knockdown compared to the control. MCF-7 cells were treated with two retroviral constructs, a shRNA to knockdown SIRT1 (n=3), and a shRNA directed against Luciferase (n=3) and RNA was hybridized to Affymetrix HGU133A 2.0 microarrays (GSE13459). F. SIRT1 pathway activation was increased in cells treated with tobacco smoke condensate (TSC). A549 cells or Calu-6 cells were cultured with and without TSC and hybridized to Affymetrix HGU133 Plus 2.0 microarrays (GSE13309). For each cells type, there were 3 controls and 9 samples treated with TSC.

Supplemental Figure 3. The effect of modulating the number of genes in the pathway signatures on pathway activation predictions in each control dataset. The number of genes in each signature is along the x-axis and the different control datasets from Supplemental Figure 2 are shown along the y-axis. A. ATM Pathway, GSE20549. B. GPX1 pathway, GSE23895. C. IKBKB pathway, GSE18047. D. SIRT1 pathway, GSE9008. E. SIRT1 pathway, GSE13309. F. SIRT1 pathway, GSE13459.

Supplemental Figure 4. SIRT1 pathway activation in lung tumor tissue datasets with smoking-status or different tumor subtypes. A. Lung tumor tissue (n=60, 56 adenocarcinomas, 1 squamous, 3 bronchioloalveolar carcinoma) and paired adjacent normal lung tissue (n=60) (GSE19804) from non-smoking women in Taiwan hybridized to Affymetrix HGU133 Plus 2.0 microarrays. There was no different between SIRT1 pathway activation between adjacent normal and tumor tissue. B. Top left. Adjacent normal lung tissue and paired lung tumor tissue (n=42 pairs, n=22 adenocarcinoma, n=9 large cell carcinoma, and n=11 squamous cell carcinoma) (GSE19188) hybridized to Affymetrix HGU133 Plus 2.0 microarrays. SIRT1 pathway

activation was decreased between adenocarcinomas and normal lung tissue (ADC, p=0.044) and between large cell carcinomas and normal lung tissue (LCC, p=0.002); however, SIRT1 pathway activation was significantly increased in squamous cell carcinoma compared to adjacent normal lung tissue (SCC, p=0.023). Top middle. Adenocarcinoma (ADC, n=40) and squamous cell carcinoma (SCC, n=18) lung tumor tissue (GSE10245) was hybridized to Affymetrix HGU133 Plus 2.0 microarrays. SIRT1 pathway activation was significantly higher in SCC versus ADC tumors ($p<0.001$). Top right. Adenocarcinoma (ADC, n=50), large cell carcinoma (LCC, n=20), and squamous cell carcinoma (SCC, n=28) lung tumor tissue (GSE28571) was hybridized to Affymetrix HGU133 Plus 2.0 microarrays. SIRT1 pathway activation was significantly higher in SCC versus ADC tumors ($p<0.001$). Bottom left. Adenocarcinoma (ADC, n=58) and squamous cell carcinoma (SCC, n=53) lung tumor tissue (GSE3141) was hybridized to Affymetrix HGU133 Plus 2.0 microarrays. SIRT1 pathway activation is significantly lower in ADC versus SCC tumors ($p<0.001$). Bottom middle. CCLE lung cancer cell lines (n=180) that include n=49 adenocarcinoma (ADC), n=5 bronchioloalveolar adenocarcinoma (BAC), n=14 large cell carcinoma (LCC) , n=4 mixed adenosquamous carcinoma (Mixed ADC/SCC), n=1 mucoepidermoid carcinoma (mucoepidermoid), n=24 non-small cell carcinoma (NSCLC), n=53 small cell carcinoma (SCLC), n=26 squamous cell carcinoma (SCC), n=1 undifferentiated carcinoma (UnDiff), and n=3 Unknown.

Supplemental Figure 5. SIRT1 and PI3K pathway activation in airway epithelial cells from current and formers smokers with and without lung cancer. CEL files from the training and test set samples in GSE4115 (n=129 samples, n=29 current smokers with lung cancer, n=22 current smokers without lung cancer, n=31 former smokers with lung cancer, n=26 former smokers without lung cancer) were processed as described in the methods. The pathway signature for PI3K was re-derived using the methods described above using a compendium of oncogenic signatures including BCAT, E2F3, MYC, p63, RAS, and SRC (GSE3151 and GSE12815). PI3K pathway activation probabilities for each sample were computed using the same methodology described for SIRT1. A. There is no significant difference in SIRT1 pathway activation between patients with and without lung cancer. SIRT1 is significantly activated in current smokers compared to former smokers ($p<0.001$). B. PI3K is significantly activated in patients with lung cancer ($p<0.001$).

Supplemental Figure 6. SIRT1 copy number changes in the Cancer Cell Line Encyclopedia (CCLE) dataset. The dataset consists of both gene expression (measured by Affymetrix HGU133 Plus 2.0 microarrays) and DNA copy number data (measured by Affymetrix SNP 6.0 microarrays) across 991 cancer cell lines. A. Across the 174 lung cancer cell lines, copy number values for SIRT1 in the CCLE dataset are in the insignificant range (dark gray, from -1 to 1), indicating that there is not a loss or amplification of the SIRT1 locus in lung tumors; in contrast, the values for MYC, range from -1 to 3 (light gray). B. SIRT1 and C. MYC pathway activation were predicted across the cancer cell lines using the SIRT1 signature and a MYC signature derived in human mammary epithelial cells (17). There is a significant Pearson correlation between MYC pathway activation and MYC copy number ($r=0.51$, $p<<0.001$) but not between SIRT1 pathway activation and SIRT1 copy number ($r=0.14$, $p=0.07$), suggesting that decreases in SIRT1 pathway activation are not due to copy loss at the locus. The results are derived by performing RMA as described in the manuscript using the ENTREZ Gene CDF file on each of the 14 batches of CCLE CEL files. Pathway predictions were made using either the SIRT1

or MYC signature across each batch of the CCLE dataset, the predictions from each batch were combined, and lung cancer cell lines were separated out for the analyses in B and C.