Figure e4:
Correlation of experimental SF2 and predicted SF2 based on gene expression levels. Using leave-one-out cross-validation, we fitted 12 times a linear model for the experimental SF2 of 11 of the 12 cell lines using the gene expression levels of the top 10 genes, and predicted the SF2 of remaining cell line. The euclidean distance of the 36 predictions, as a measurement of error, ranges from 0.0023 to 0.1312 (mean = 0.061). The coefficient of determination for the 12 fitted models, which represents the goodness of fit for a linear model, ranges from 0.95 to 0.97 (mean = 0.95), with 1.0 indicating a perfect fit.