SUPPLEMENTARY METHODS

Construction of EGFRvIII-Y845F mutant

To generate EGFRvIII containing tyrosine 845 to phenylalanine mutation, recombinant PCR using Pfu Turbo (Stragene) was performed. Briefly, two separate PCR amplifications were performed on pSKWCER (pBluescript SK (Stragene) + EGFRvIII) using primers EGFR1: atc ttg aag gaa act gaa ttc and Y845Fas: gcc tcc ttc tgc atg **aaa** ttc ttt ctc ttc cgc for reaction #1 and EGFR2: tcg ggc cat ttt gga gaa ttc and Y845Fs:gcg gaa gag aaa gaa **ttt** cat gca gaa gga ggc for reaction #2. Equal molar amounts of gel purified reaction 1 and 2 products were subsequently combined and reamplified using outer EGFR1 and EGFR2 primers. The resultant recombinant PCR product containing the Y845F mutation was next digested with EcoR1 and ligated with pSKWCER from which the corresponding wild type sequence was removed by EcoR1 digestion. Transformed colonies containing the ligated recombinant fragment were screened for correct orientation and presence of mutation by sequencing with EGFR1 and EGFR2. A positive colony containing the EGFRvIII-Y845F mutation (pSKWCER-Y845F) was next subcloned into pLRNL as a three-piece ligation. Briefly, 1.6- and 4.6-kilobase pair vector bands generated from Sca1/Xba1 and Sca1/Sal1 digestions of pLRNL respectively were gel purified and ligated with a 3.1-kilobase pair mutation-containing fragment generated from Xba1/Sal1 digestion of pSKWCER-Y845F. Transformed colonies were screened by PCR using primers EGFR1 and M13-20, which generated a 1.5-kilobase pair product from positive clones.

Retroviral infection and generation of MEK expressing cell lines

U87MG cell expressing EGFRvIII and mutant receptors were transfected with the retroviral vectors containing control pBABE-puro, constitutively active MEK (CA) pBABE-MANE, or dominant negative MEK (DN) pBABE-LIDA (previously described in [1]. Viruses were produced by seeding 293-GP cells at $3x10^6$ cells per 10cm plate and transfected with FuGene 6 transfection reagent (Roche Applied Sciences) with 10 μg of pBABE retroviral plasmid and 5 μg of VSV-G plasmid. After 16 hours, cells were washed with PBS and replaced with 10 ml of fresh media. Viruses were harvested from the media 48 hours later, filtered with a 0.45 μm SFCA filter and used to infect the U87MG cell lines. Stable populations were obtained by selection in 2.5 μg/ml puromycin and expression of MEK mutants was confirmed by immunoblotting.

Immunoblotting Analysis

Cells were lysed in lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 2.5 mM sodium PPi, 1 mM -glycerophosphate) containing protease and phosphatase inhibitors after the indicated treatment. Protein concentration of cell lysates was determined using micro bicinchoninic acid assay (Pierce), according to the manufacturer's protocol. 50 μg of protein from the cell lysate was mixed with 4X sample buffer (250 mM Tris-HCl, pH 6.8, 40% glycerol, 0.04% bromophenol blue and 400 mM DTT) and loaded on either 7.5% or 10% SDS-PAGE gels, separated and developed as

previously described [2]. Blots were developed with the SuperSignal West Femto Maximum Sensitivity Substrate kit (Pierce) and scanned on a Kodak Image station 1000. Primary antibodies used were anti-phopho-Erk1/2 and anti-tubulin (Cell Signaling Technologies). Secondary antibody used was goat anti-rabbit antibody (Jackson Immunoresearch).

Cell viability assay

5000 cells were seeded per well in a 96 well plates. 24 hours later, 100µl of serum free media containing DMSO or U0126 (Promega) at the indicated dose was added to the cells. At various timepoints, cell viability was measured using the WST-1 reagent (Roche Applied Sciences). 10µl of WST-1 was added to each well and incubated at 37°C for one hour prior to measuring absorbance at 450nm in a spectrophotometer. All measurements were performed in biological triplicates.

Cell growth measurements

For cell growth measurements of Y-F mutant cell lines, 4×10^4 cells were plated per well of a 6-well plate in full growth media. After 24 hours, the cells were serum starved for an additional 24 hours. Cell counts were performed at the times indicated using the Nexcelom cellometer system. For MEK mutant cell growth measurements, 1×10^5 cells were plated per well in a 6-well plate. 24 hours later, the cells were serum starved. Cells were counted at 72 hrs after serum starvation using the Vi-cell cytometer system.

Calculation of growth rate constants

Quantification of growth requires both selection of growth models and selection of parameters from those growth models that are robust and representative of cell line growth. The simplest growth model assumes that growth is exponential, and the growth rate is contained in a single parameter, the growth rate constant. Preliminarily, cell growth measurements from 0-12 hours were fit to an exponential model. Although the fit was very good for all but DY3 and Y1173F this result led us to consider other growth models and fitting procedures.

The closely related logistic and Gompertz equations have both been used to empirically model the growth of cancer for both *in vivo* and *in vitro* contexts. These curves, which involve more parameters than the exponential, are capable of describing exponential growth following an initial lag phase and/or followed by a phase at which growth rate diminishes and cell population approaches a saturating, steady-state. The relative growth rate constants of DY3 and Y1173F appeared to be decreasing with time, and, thus, a more sophisticated modeling approach (logistic, Gompertz) seemed appropriate. The cell lines that did appear to grow exponentially were also fit to the more complicated models, as it has been suggested that fitting the full set of growth measurements (in this case, 0-24hrs.) as opposed to subset of time points (0-12hrs., assumed to be exponential) may yield more robust growth parameters [3].

To select among the more complicated models, the method of Shnute [4] was employed. The Shnute model is a generalized four-parameter growth equation which reduces to one of several common models: Gompertz, Logistic, Richards, Linear, Quadratic, Exponential, etc. We used least squares fit in Matlab to characterize the cell lines' growth curves and found that most cell lines fit a Richards model best, but that no single model was selected by all cell lines.

Next we used cross-validation (CV) to gauge the robustness of growth parameters from the various models [5]. Each cell line's growth curve was a composite of four biological replicate timecourses. For each cell line, we sequentially left out an entire biological replicate (growth curve) and fit to Richards (0-24hrs.), Shnute (0-24hrs.), or an exponential curve (0-12hrs.) and calculated both the mean and maximal relative growth rates across the time interval 0-12hrs. Cross validation provided a mean and standard deviation for both mean and maximal relative growth rates. For an exponential curve, of course, the relative growth rate is constant and there was no difference between mean and maximal relative growth rates for a given CV dataset. For the Richards and Shnute equations, however, the maximal relative growth rate was much less robust than the average relative growth rate and the correlation between exponential and Shnute/Richards values for maximal growth rate was relatively small (Figure S6). However, the average growth rate constant of all three models correlated well, and so, with confidence, we quantified growth using the relative rate constants derived from the exponential model, as it appeared representative and robust across several models.

Statistical Significance Testing of Growth Rate Comparison

To calculate a p-value for EGFRvIII and Y1068F growing faster than the other cell lines, we used a one-sided t-test and corrected for multiple hypothesis testing using Boole's inequality Bonferroni correction. Correction for multiple hypothesis testing was necessary, because we decided to test this hypothesis after seeing the data, and, given that there were 7 cell lines, there were a total of 126 comparisons or tests that could have been made.

Statistical Analysis

Statistical analysis was performed using two-tailed student t-test.

Partial Least Squares Regression Analysis

PLSR models were built using a PLSR algorithm based on the nonlinear partial least squares (NIPALS) as described in [6] using MATLAB (Version 7.4.0.287, Mathworks Inc.). An NxP data matrix (**X**) was constructed from the mass spectrometry signaling dataset, where N is the number of cell lines and P is the number of predictors (phosphosites). A column vector (**y**) of length N contained the growth rate constants for each cell line, which was mean-centered and standardized to unit-variance. The PLSR

algorithm simultaneously decomposed **X** and **y** matrices in terms of the X-scores and Y-scores matrices, **T** and **U**:

$$\mathbf{X} = \mathbf{TP}' \text{ and } \mathbf{T} = \mathbf{X}(\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}) \quad (1)$$
$$\mathbf{y} = \mathbf{UC}' \sim= \mathbf{TBC}' = \mathbf{X}(\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1})\mathbf{BC}' \quad (2)$$

where **T** and **U** are chosen such that $[\text{Cov}(\mathbf{t},\mathbf{u})]^2$ is maximal. For data visualization, phosphosite and growth rate were plotted in the PLSR space defined by the X-scores using the coefficients contained in matrices **P** and **BC**.

The "full" regression model follows from equation (2). To select the number of principle components to use in the model, both the variance in growth captured by the model as well as associated predictive or Q^2 value were used as criteria. Q^2 was calculated using leave-one-out cross-validation (LOOCV) [7]. Variance explained was calculated as described in [8]. Based on this analysis, three principle components were used to build the model.

The PLSR space used in Figure 6A was constructed as described above, except the column vector **y** was replaced by an *Nx2* matrix, whose columns corresponded to mean-centered and variance normalized growth or CA MEK viability measurements.

The "reduced" regression model was constructed using variable importance for projection (VIP) as criteria to determine which phosphosites to incorporate into the model. [7]. The thirteen phosphosites with greatest VIP score were selected for incorporation into the "reduced" model. In particular, thirteen were selected because thirteen phosphosites maximized both the model's predictive power and fit (Figure S7).

REFERENCES

1. S. Cowley, H. Paterson, P. Kemp and C. J. Marshall, *Cell*, 1994, **77**, 841-852.

2. P. H. Huang, A. Mukasa, R. Bonavia, R. A. Flynn, Z. E. Brewer, W. K. Cavenee, F. B. Furnari and F. M. White, *Proc Natl Acad Sci U S A*, 2007, **104**, 12867-12872.

3. M. H. Zwietering, I. Jongenburger, F. M. Rombouts and K. van 't Riet, *Appl Environ Microbiol*, 1990, **56**, 1875-1881.

4. J. Shnute, *Can. J. Fish. Aquat. Sci.*, 1981, **38**, 1128-1140.

5. L. Zhang, *Annals of Botany*, 1997, **79**, 251-257.

6. P. Geladi and B. R. Kowalski, *Analytica Chimica Acta*, 1986, **185**, 1-17.

7. N. Kumar, A. Wolf-Yadlin, F. M. White and D. A. Lauffenburger, *PLoS Computational Biology*, 2007, **3**, e4.

8. P. Geladi and B. Kowalski, *Analytica Chimica Acta*, 1986, **185**, 1-17.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Hierarchical clustering of tyrosine phosphorylation sites. Heatmap clustering of the 132 phosphorylation sites quantified in the phosphoproteomic analysis. Phosphorylation levels were normalized to U87MG-EGFRvIII control cell line. On the left are the profiles for each of the four clusters of phosphorylation sites.

Figure S2. Cumulative variance explained by GBM cell growth models. The cumulative amount of growth variance explained as a function of principle components included in either the full or reduced model. Error bars represent standard deviation, determined by leave-one-out cross validation.

Figure S3. Phenotypic responses of EGFRvIII mutant cell lines to MEK inhibition. Cells expressing wtEGFRvIII were treated with 20μM or 40 μM MEK inhibitor U0126 for either (A) 24 hours or (B) 48 hours. Cell viability was assessed by WST-1 assay. Experiments were performed in biological triplicates.

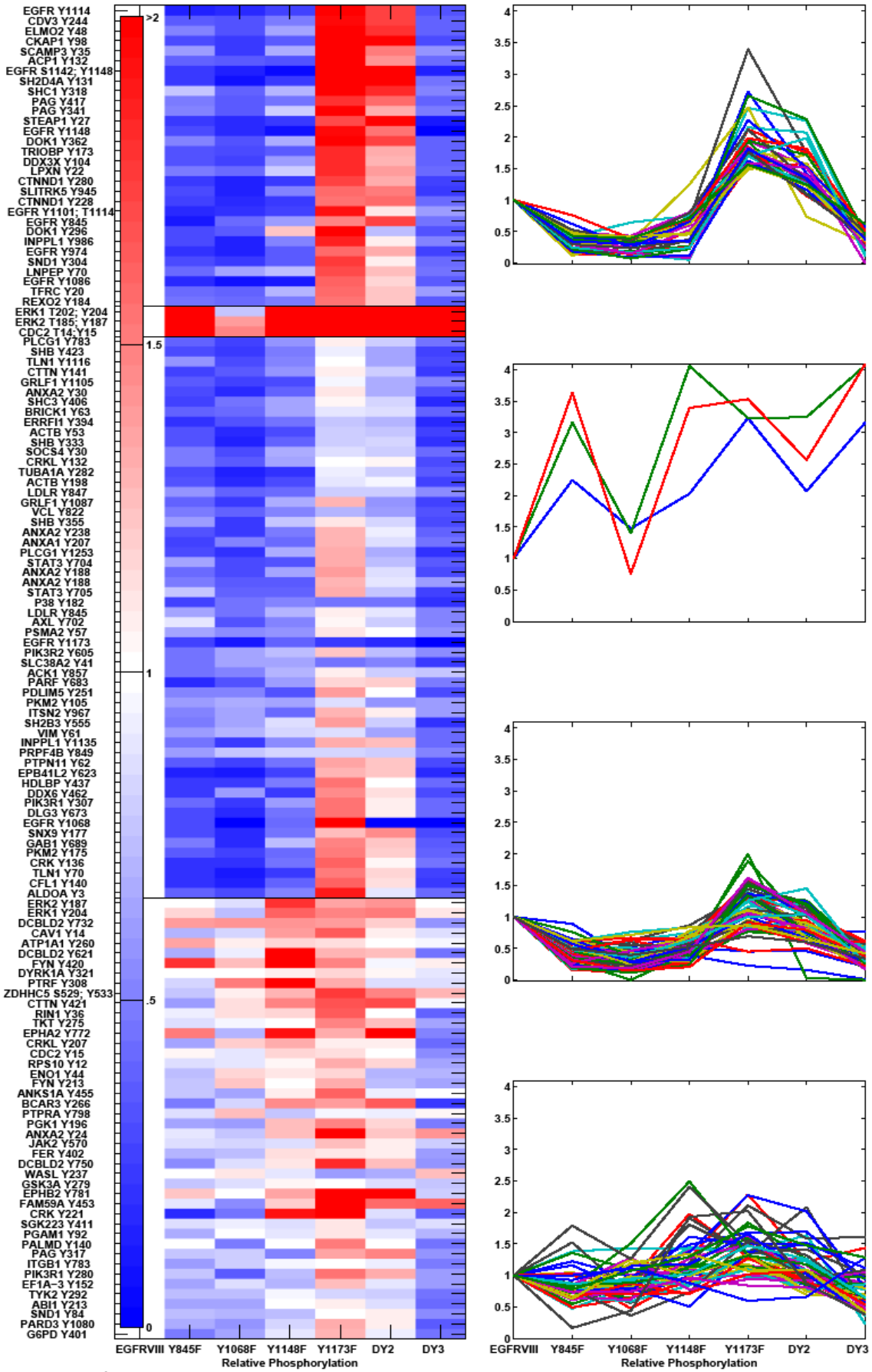
Figure S4. Phenotypic responses of EGFRvIII mutant cell lines to Erk hyperactivation. (A) Western blot of double phosphorylated Erk1/2 after transfection of EGFRvIII mutant cell lines with either pBabe-puro (Control) or constitutively activate MEK (CA). (B) Viability response of EGFRvIII mutant cells upon hyperactivation of the Erk pathway. Transfected cells were subjected to serum depravation conditions for 72 hours prior to counting viable cells by trypan blue exclusion method. Experiments were performed in 6 replicates normalized to control cells.

Figure S5. Full GBM cell growth model predictions. (A) Fit of the full PLSR U87 cell growth model to the exponential growth rate constants. Error bars for experimental and PLSR fits represent standard deviations as estimated by maximum likelihood or leave-one-out-cross-validation methods, respectively. (B) The cumulative amount of phenotypic (growth and CA-MEK viability) data variance explained as a function of principle components included in a full PLSR model. Error bars represent standard deviation, as determined by leave-one-out cross validation.

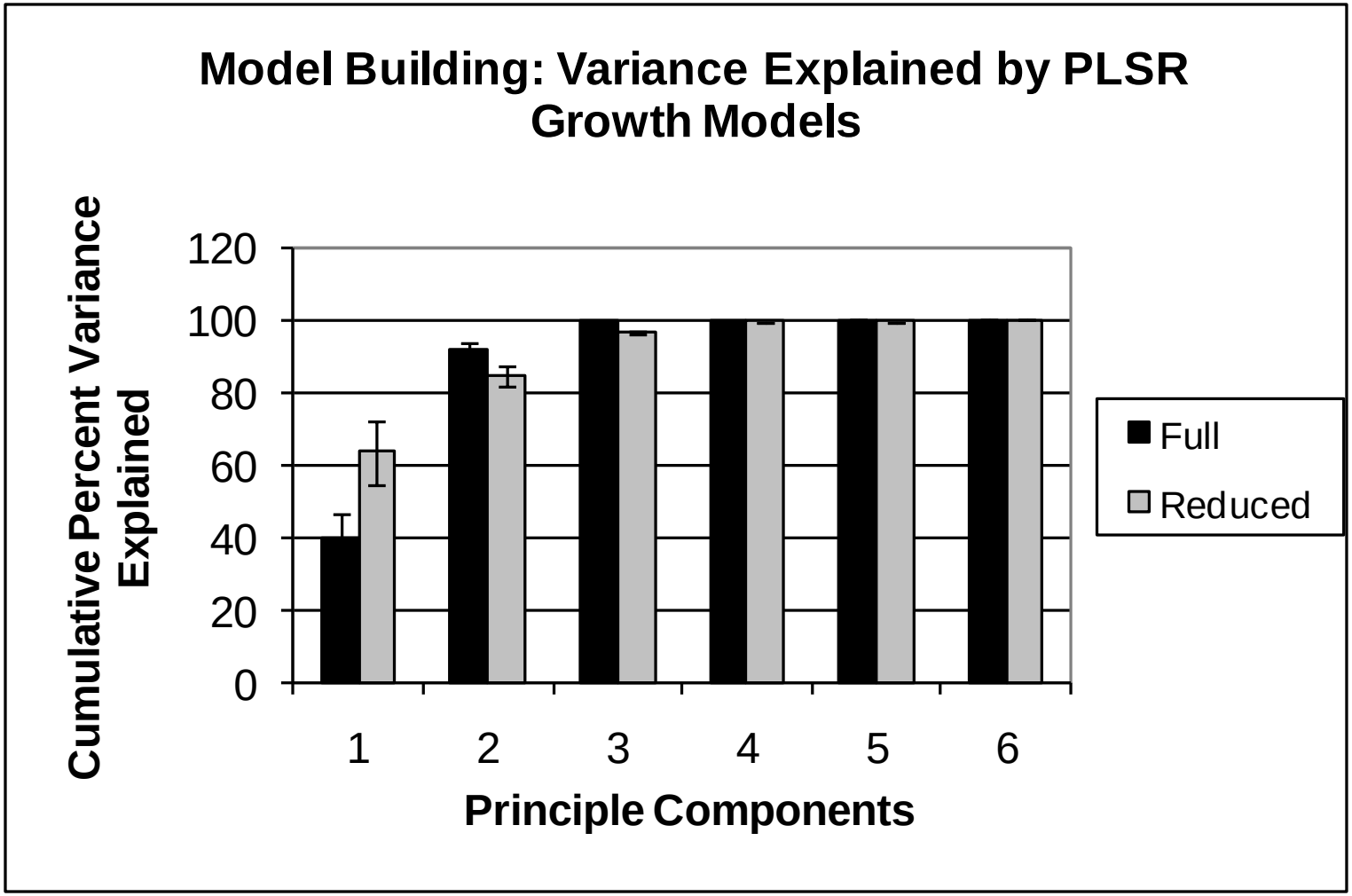
Figure S6. Growth rate constants determined using Richards, Shnute, and Exponential Models. (A) Mean and maximum relative rate constants, (B) goodness-of-fit R^2 values for the DY series cell lines using Richards, Shnute, and Exponential growth models; standard deviations are estimated by leave-one-out cross validation (see text). The robustness of the growth parameters, mean or max relative rate constants, and model fitting procedures is compared as the mean or max relative rate constant (R.R.C.) percent standard deviation, (C) and (D). In (C), the R.R.C. percent standard deviation is shown per cell line, and, in (D), the mean R.R.C. percent standard deviation across the cell lines is displayed. The max R.R.C. for Richards and Shnute models are less robust, because they have the highest percent standard deviation.

Figure S7. Predictive power and fit for reduced models of different sizes with either two or three principal components. Indicators of model fit and prediction, R^2 and Q^2

values, respectively, are plotted as a function of phosphosites included in a reduced model based on either two or three principle components. In general Q^2 values are higher in the two-principle-component models, suggesting that inclusion of three principle component leads to over-fitting. In the case of two principle components, models with less than thirteen components lack predictive power, while inclusion of more than about sixteen phosphosites decreases Q^2 value and predictive power.

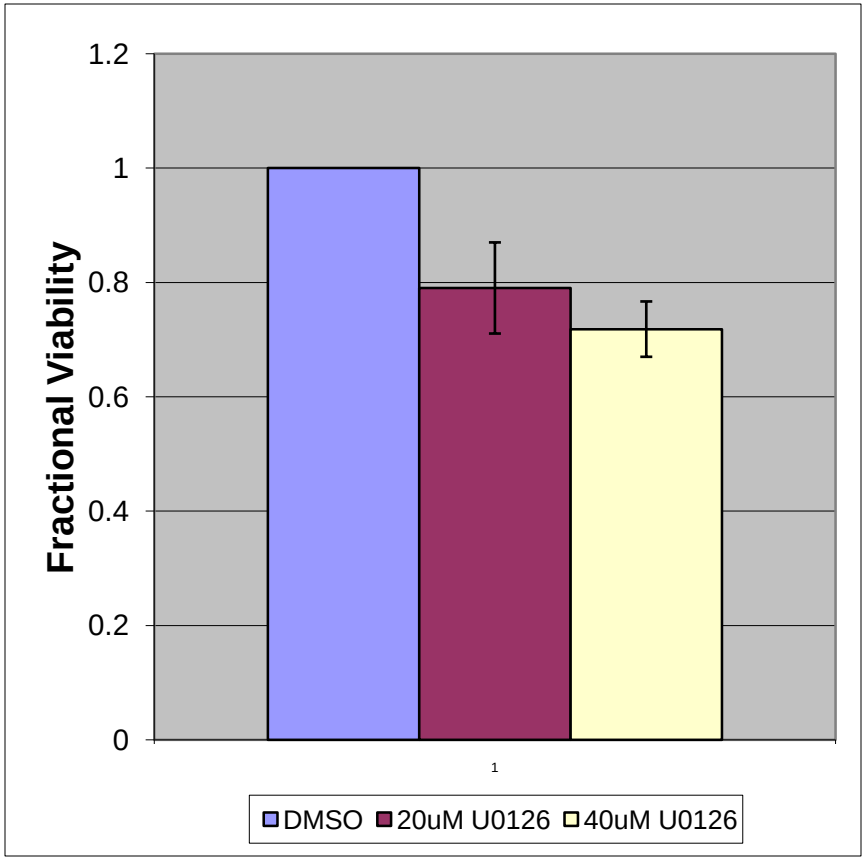


Supplemental Figure S1

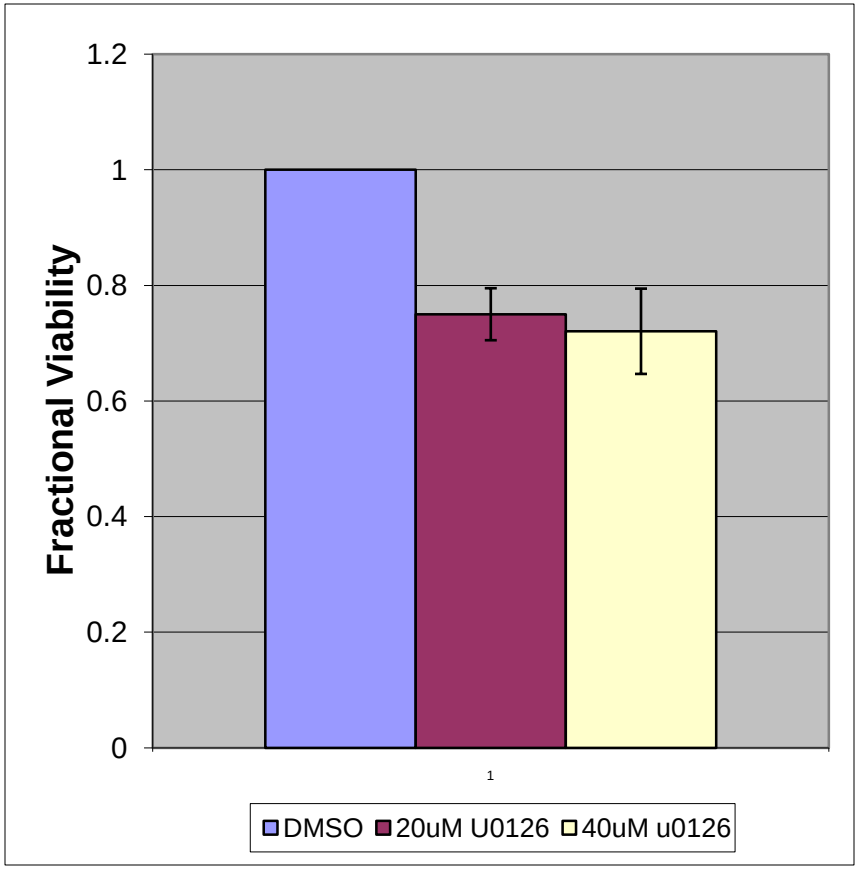


Supplementary Figure S2

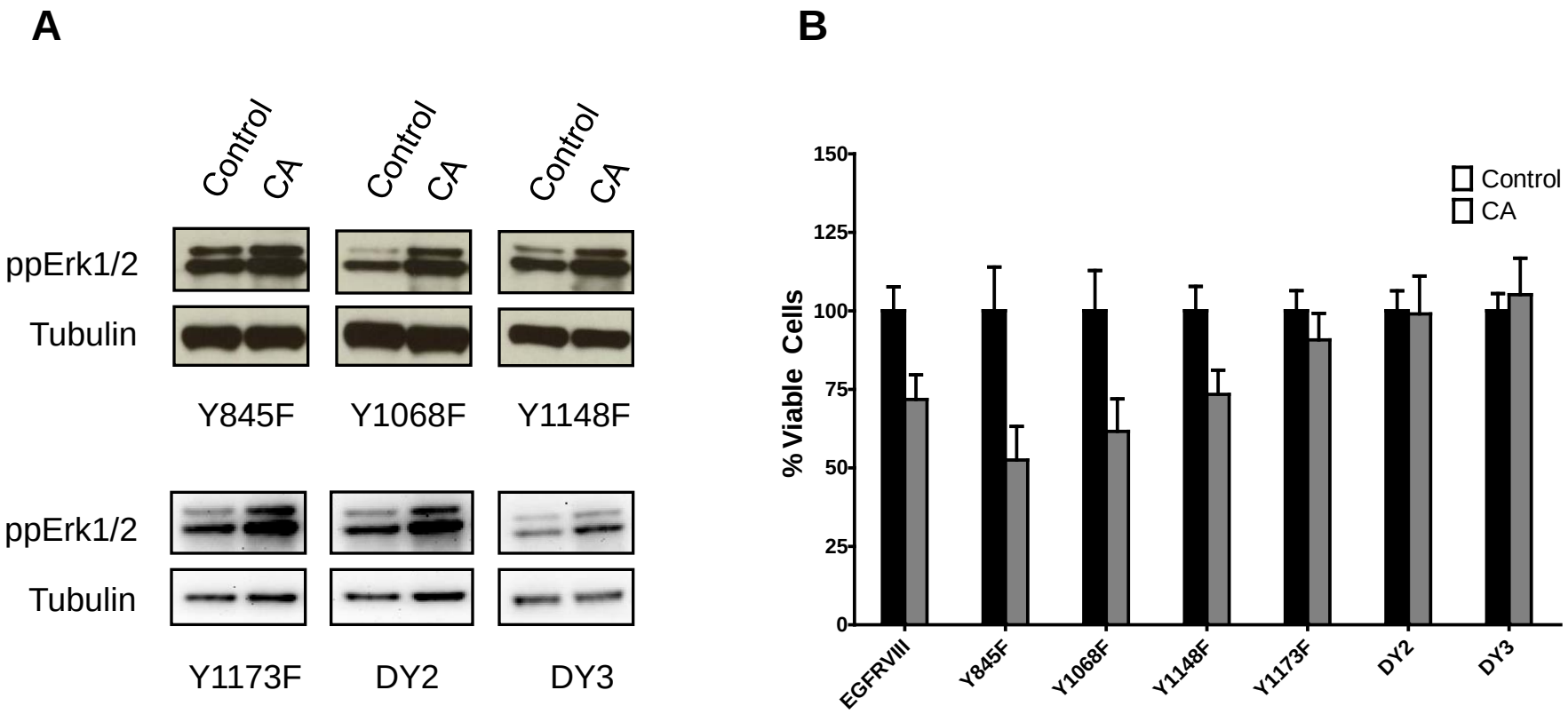
A



B

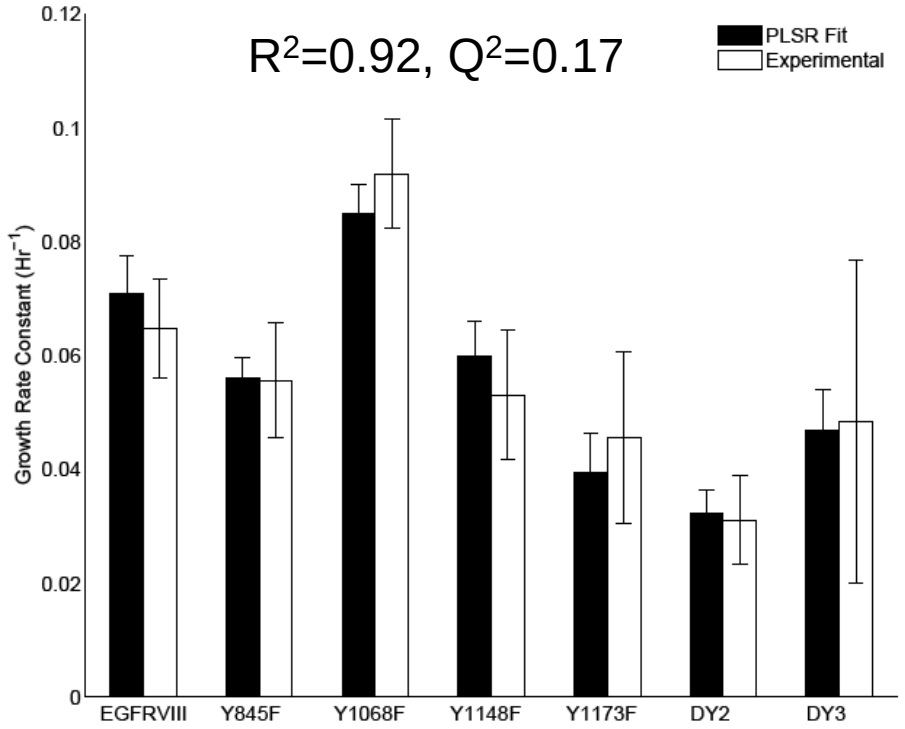


Supplementary Figure S3

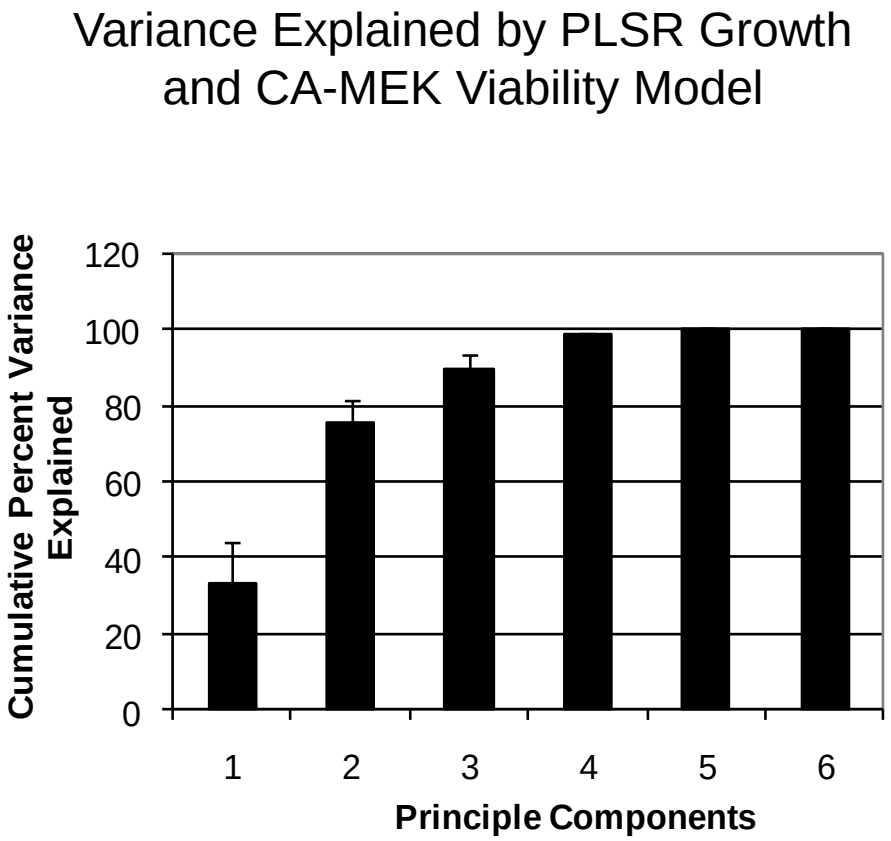


Supplementary Figure S4

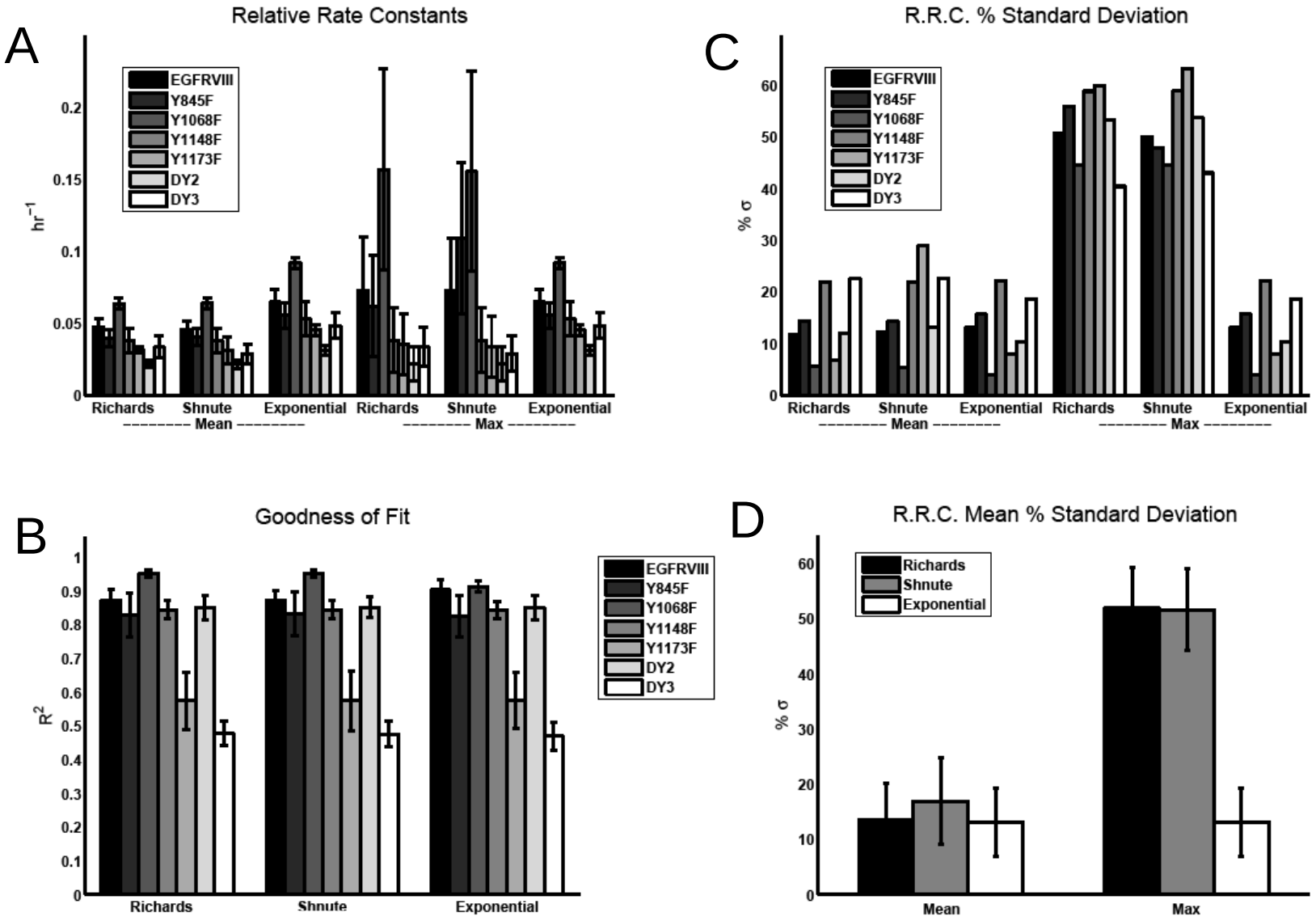
A



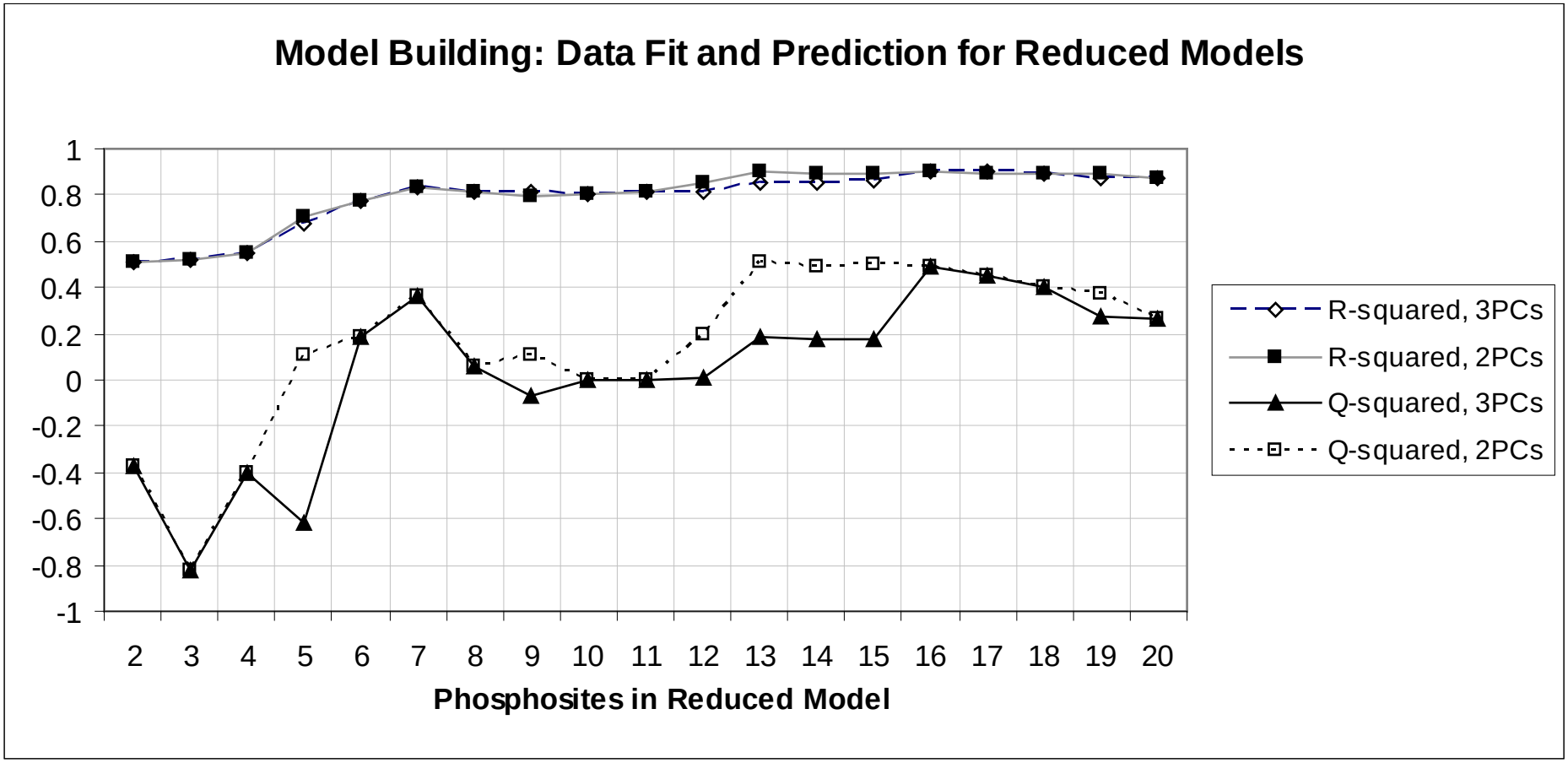
B



Supplementary Figure S5



Supplementary Figure S6



Supplementary Figure S7