Phosphoproteomic analysis of TMEM16A knockdown and EGF treatment on pancreatic cancer line AsPc-1

June 15, 2018

Motivation

The calcium-activated chloride channel (TMEM16A), also known as anoctamin 1 (ANO1), is over-expressed in many cancers. The transmembrane protein is an attractive cancer target owing to its ability to activate many signaling pathways, including the epidermal growth factor receptor (EGFR) signaling, to support cell growth and migration. However, the effect of TMEM16A on cellular phenotype appears to be cell- and tumor-specific (Molecular Cancer 2017). For instance, Wu et al. found that TMEM16A over-expression promotes proliferation in ER-positive, PR-positive, and HER2-negative MCF-7 cells, but inhibits proliferation in ER-negative, PR-negative, and HER2-negative MDA-MB-435S cells (Oncotarget 2017). This suggests that further validation of TMEM16A as a viable target is needed on a per-cell or -tumor basis.

Pancreatic ductal adenocarcinoma (PDAC) is the most common type of pancreatic cancer, and TMEM16A is often over-expressed in PDAC cells. Given that TMEM16A directly interacts with EGFR, which in turn can trigger downstream growth signaling, it was hypothesized that TMEM16A over-expression drives the aggressive cell growth in PDAC. However, a study by Sauter et al. showed that TMEM16A silencing by siRNA did not reduce proliferation in PDAC cell lines (Eur J Physio 2015). In order to dissect the mechanism behind the sustained growth, we aim to probe the phosphoproteome of the PDAC cell line AsPc-1 by mass spectrometry.

Workflow

1. David cultured three biological replicates of AsPc-1 cells.
    - Cell conditions:
        - shControl (shControl_CTRL)
        - shControl + EGF (shControl_EGF)
        - shTMEM16A (shTMEM_CTRL)
        - shTMEM16A + EGF (shTMEM_EGF)
    - Cells were washed twice in cold PBS and freshly lysed in buffer containing 8M urea, 0.1M Tris pH8.5, 40mM 2-CAA, 10mM TCEP, and 1X HALT protease/phosphatase inhibitor.
    - Cell lysates were snap frozen and stored in -80°C.
2. Processed samples through the phosphoproteomics pipeline.
    - Sonicated cell lysates.
    - Digested 1mg lysate overnight with trypsin at 1:50 (w/w) enzyme-to-substrate ratio.
    - Desalted and enriched for phosphopeptides using Fe^{3+}-IMAC columns.
    - Dried down and stored phosphopeptides in -80°C.
    - *Note: Replicate 1 processed on 02/14/2017 and replicates 2 and 3 processed on 06/01/2018.*
3. Analyzed samples on LC-MS/MS.
    - Reconstituted peptides and injected 1μg of each sample onto the instrument.
    - Data acquisition with LC-MS/MS on a 4-hour gradient.
    - Samples were queued up by replicates and sample conditions.
        - First run: **shControl_CTRL_bR1**, **shControl_EGF_bR1**, **shTMEM_CTRL_bR1**, **shTMEM_EGF_bR1**.
        - Second run: **shControl_CTRL_bR2**, **shControl_EGF_bR2**, **shTMEM_CTRL_bR2**, **shTMEM_EGF_bR2**.

    - Third run: **shControl_CTRL_bR3**, **shControl_EGF_bR3**, **shTMEM_CTRL_bR3**, **shTMEM_EGF_bR3**.
  - *Note: Replicate 3 not yet analyzed by mass spectrometry.*
4. Processed MS data by MaxQuant using Phospho (STY) settings and match-between-runs.
5. Data analysis and visualization in R.
   - **Filtering**
     - Remove contaminants and reverse hits.
     - Retain confident phosphosite identifications.
     - Retain sites where at least 50% of replicates are quantified in one condition.
   - **Normalization**
     - Transform data by taking $\log_2$ of phosphosite intensities.
     - Quantile normalization to remove technical variabilities.
   - **Imputation**
     - Hybrid imputation to fill in missing values with either a low value when the signal in replicates is low or a value assigned by maximum likelihood estimation when signal is medium or high.
   - **Statistical Testing**
     - Welch's two-sample t-test to compare any two conditions.

Results

Data Tables

The *Phospho (STY)Sites* MaxQuant output on phosphosite intensity was used in the following analyses. Please see *TMEM16A-EGF-phospho-analysis.xlsx* for the data tables. Table descriptions are outlined below.

- **raw**: Raw *Phospho (STY)Sites* output from MaxQuant.
- **shControl_EGF-vs-shControl**: Effect of EGF on phosphorylation status. The *Log2FC* column shows the $\log_2$-transformed fold change in phosphorylation after EGF treatment.
  - Green highlight = Greater than 3-fold enrichment after EGF versus no treatment.
  - Red highlight = Greater than 3-fold depletion after EGF versus no treatment.
- **shTMEM_EGF-vs-shTMEM**: Effect of EGF on phosphorylation status of TMEM-knockdown cells. The *Log2FC* column shows the $\log_2$-transformed fold change in phosphorylation after EGF treatment on TMEM-depleted cells.
  - Green highlight = Greater than 3-fold enrichment after EGF treatment on shTMEM cells.
  - Red highlight = Greater than 3-fold depletion after EGF treatment on shTMEM cells.
- **shTMEM-vs-shControl**: Effect of TMEM knockdown on phosphorylation status. The *Log2FC* column shows the $\log_2$-transformed fold change in phosphorylation after TMEM16A depletion.
  - Green highlight = Greater than 3-fold enrichment after TMEM16A knockdown.
  - Red highlight = Greater than 3-fold depletion after TMEM16A knockdown.
- **shTMEM_EGF-vs-shControl**: Combined effect of TMEM knockdown and EGF treatment on phosphorylation status. The *Log2FC* column shows the $\log_2$-transformed fold change in phosphorylation after the treatment.
  - Green highlight = Greater than 3-fold enrichment after TMEM16A knockdown + EGF.
  - Red highlight = Greater than 3-fold depletion after TMEM16A knockdown + EGF.

Table 1: Summary statistics on the processed samples (after data filtering)

Samples	# sites detected	# sites quantified	% sites quantified
shControl_CTRL_bR1	5608	2879	51.3
shControl_CTRL_bR2	5608	3892	69.4
shControl_EGF_bR1	5608	2840	50.6
shControl_EGF_bR2	5608	4383	78.2
shTMEM_CTRL_bR1	5608	2589	46.2
shTMEM_CTRL_bR2	5608	4394	78.4
shTMEM_EGF_bR1	5608	2835	50.6
shTMEM_EGF_bR2	5608	3553	63.4

Figure 1 - Correlation between replicates after data filtering and normalization

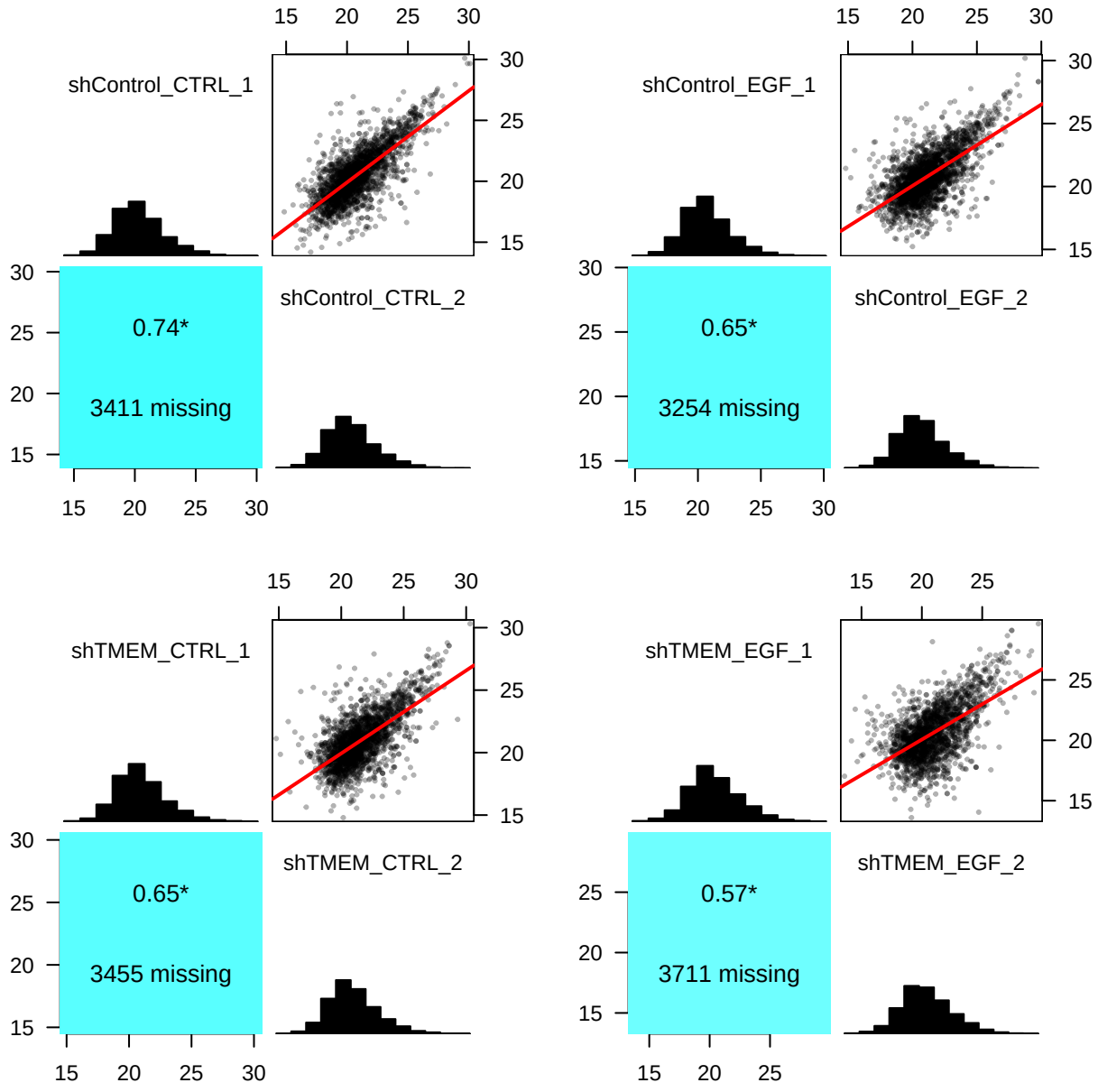
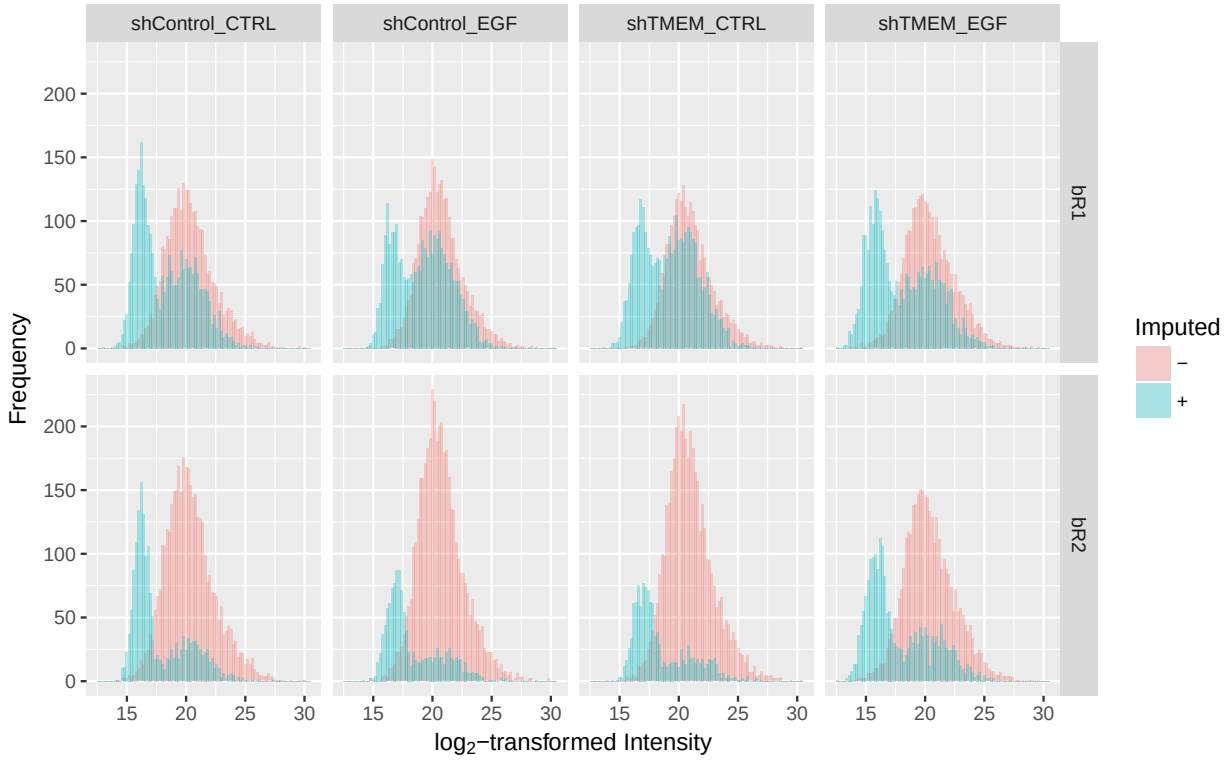


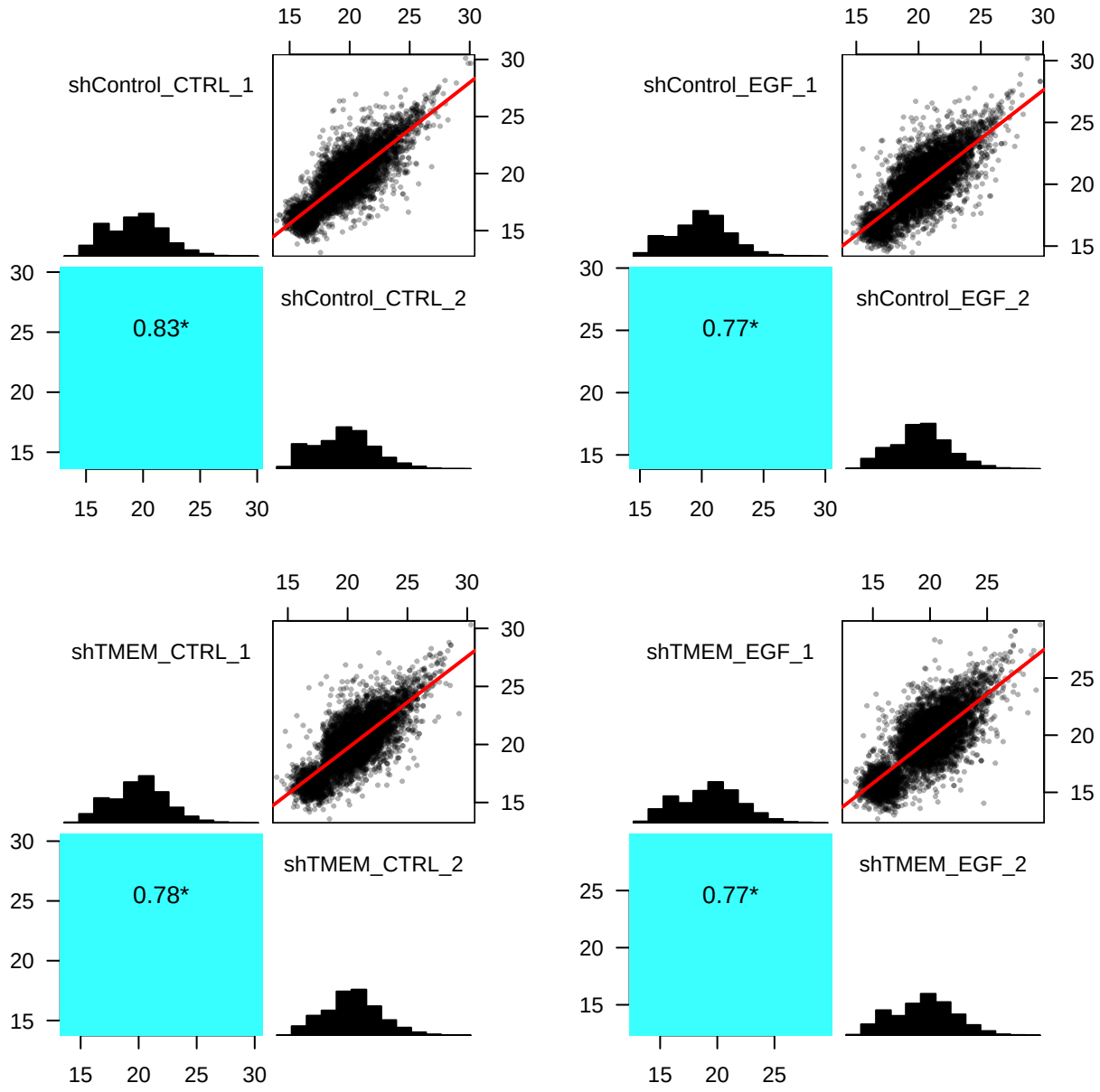
Figure 1 shows the relationship of phosphosite intensities between replicates. Each point in the scatter plot is a phosphosite quantified in both replicates. The line of best fit is drawn in red. The distribution of the normalized, $\log_2$-transformed phosphosite intensities are depicted as histograms along the diagonal. The Pearson correlation coefficients and the number of missing values present in the replicates are displayed in the bottom-left corner. These results suggest that a considerable amount of the variability exists between replicates, which is not surprising given the loosely conserved nature of phosphorylation marks and the stochasticity of acquisition method on the mass spectrometer. Additionally, batch effects could play a role since these samples were processed more than a year apart.

Figure 2 - Histogram of phosphosite intensities by sample



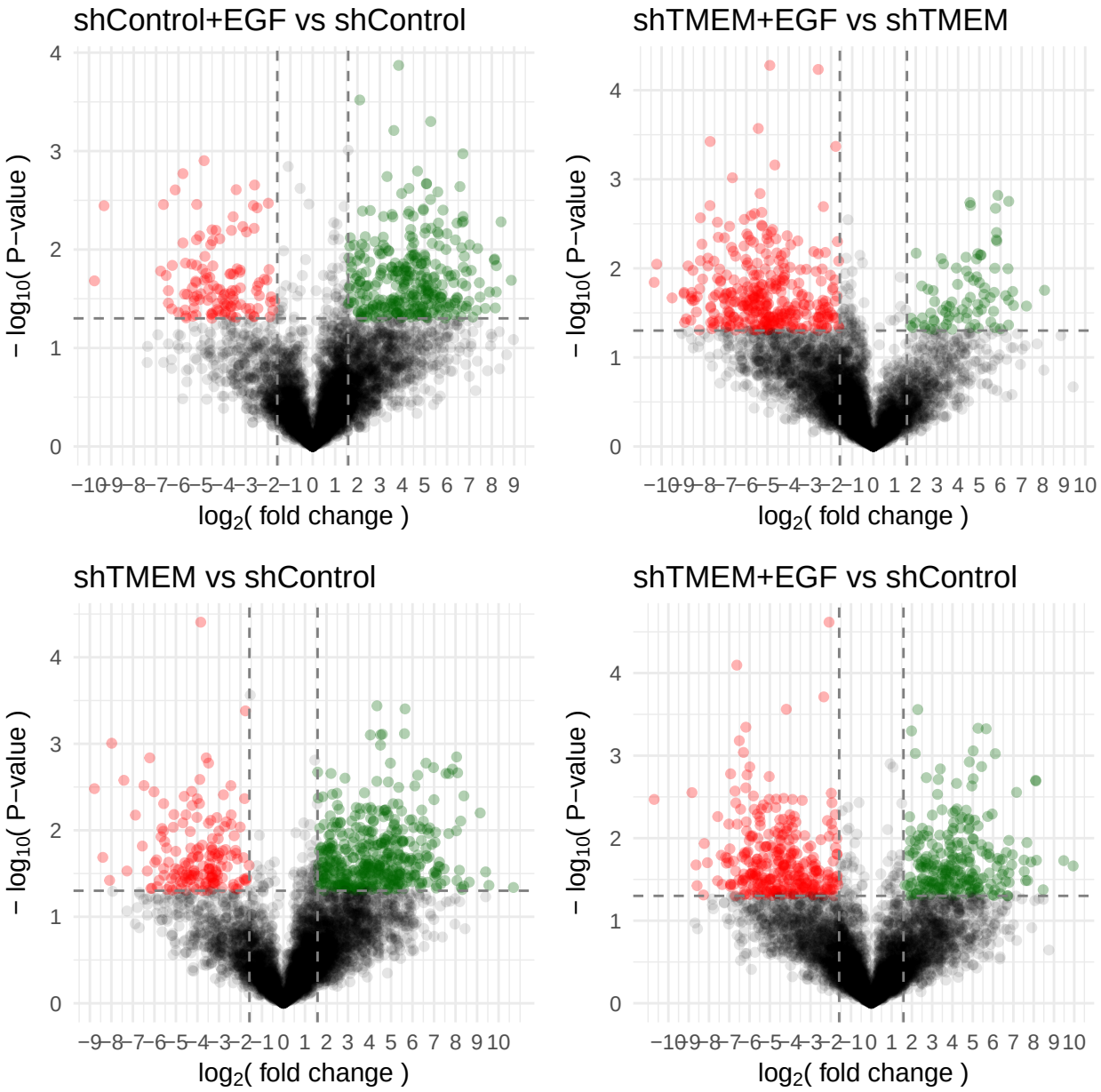
Distribution of the measured phosphosite intensities and the imputed intensities are superimposed. Clearly, replicate 1 suffers more heavily from missing values than replicate 2. Nonetheless, the hybrid imputation method appears to have compensated for this loss by filling in more substitute values along the expected unimodal distribution. With a third replicate, this imputation strategy will become more reliable at assigning proper figures for missing data.

Figure 3 - Correlation between replicates after data imputation



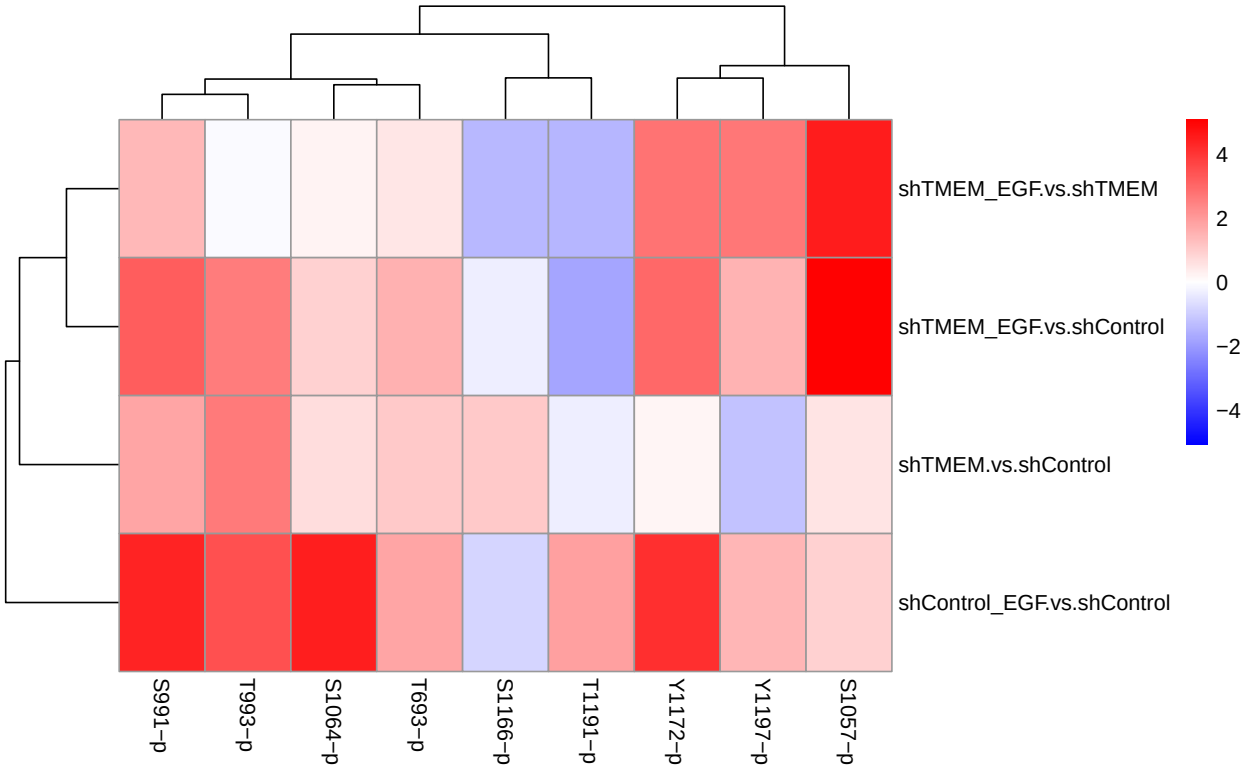
Here are the same sets of plots as **Figure 1** after data imputation has been performed. The imputation method resolves the missing data problem by inferring values that closely reflect replicate measurements. Therefore, it is perhaps not surprising that the correlation coefficients have improved upon imputation.

Figure 4 - Volcano plots on phosphorylation response to perturbation



Volcano plots show the effect of perturbation on phosphorylation levels and the statistical significance of the change along the x- and y-axis, respectively. Each point represents a phosphosite, and the thresholds are drawn at a fold change of 3 (or 1.58 in $\log_2$) and P-value of 0.05. EGF stimulation appears to up-regulate phosphorylation in shControl cells but, interestingly, exerts the opposite effect in shTMEM cells. Additionally, the shape of the shTMEM-vs-shControl scatter resembles that of shControl+EGF-vs-shControl, weakly suggesting that TMEM16A knockdown and EGF stimulation may act on the cells in similar ways. To tease out any functional patterns among proteins whose phosphosites were differentially regulated, a GO enrichment analysis was performed on a list of proteins associated with the colored points above. A list of unregulated proteins identified in the phosphoproteomic analysis was supplied as a reference set. Unfortunately, no GO terms were significantly enriched.

Figure 5 - Heatmap of EGFR phosphorylation status under a combination of shTMEM and EGF treatment



To examine the changes in phosphorylation under different conditions at a finer granularity, phosphosites on EGFR were isolated and visualized in a heatmap. The units are $\log_2$ fold change in intensity. According to PhosphoSitePlus (PTM database focused on curating protein modifications and their functions), phosphorylation on **Y1172** and **Y1197** are associated with induced EGFR activity. Taken together, these results show that EGFR is active upon EGF treatment even when TMEM16A is depleted. In contrast, TMEM16A depletion alone does not appear to induce EGFR activity. This finding, however, does not preclude the possibility that baseline growth signaling is somehow maintained or even increased in the absence of TMEM16A.

Figure 6a - Kinase Activity Prediction: shControl+EGF vs shControl

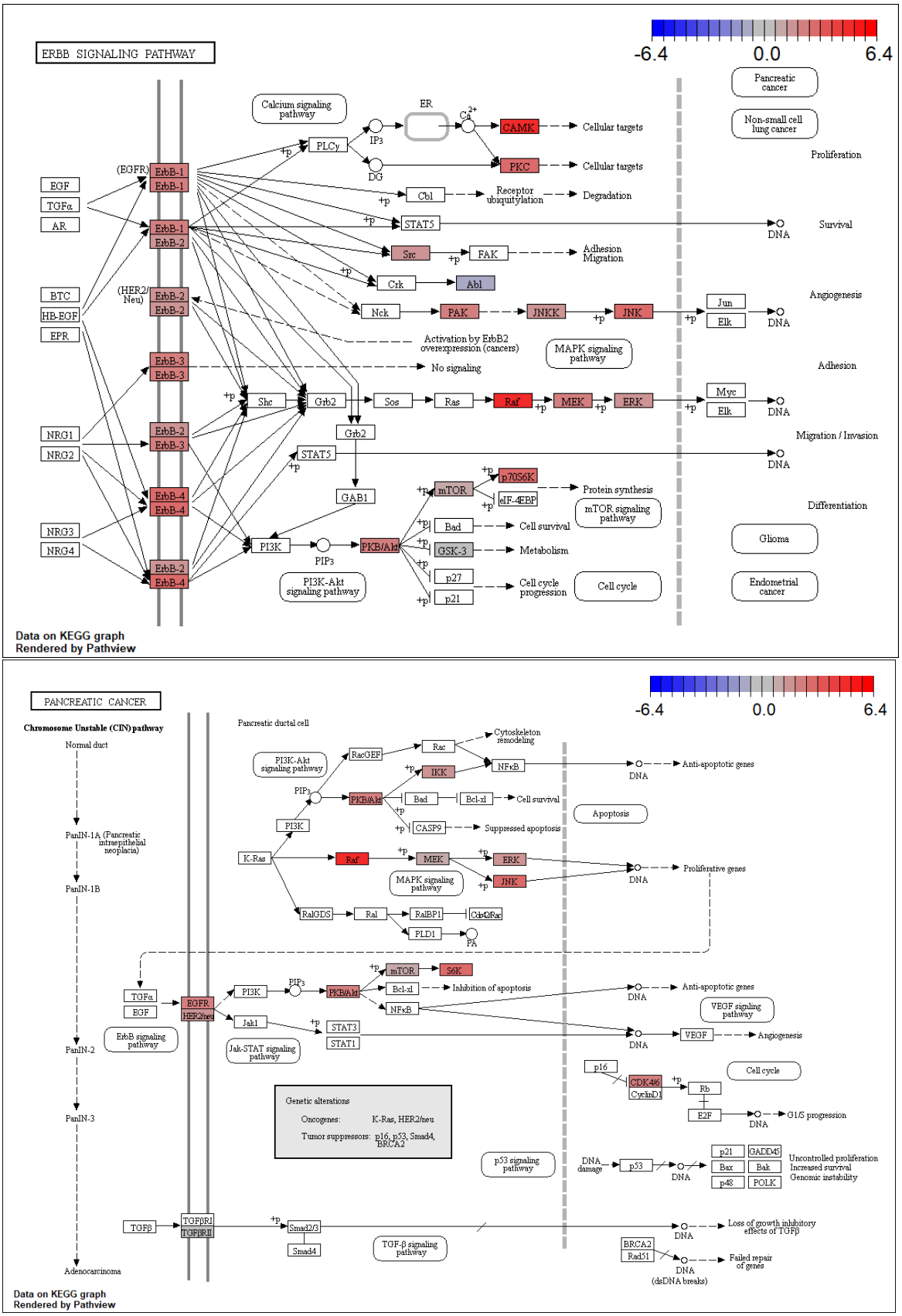


Figure 6b - Kinase Activity Prediction: shTMEM+EGF vs shTMEM

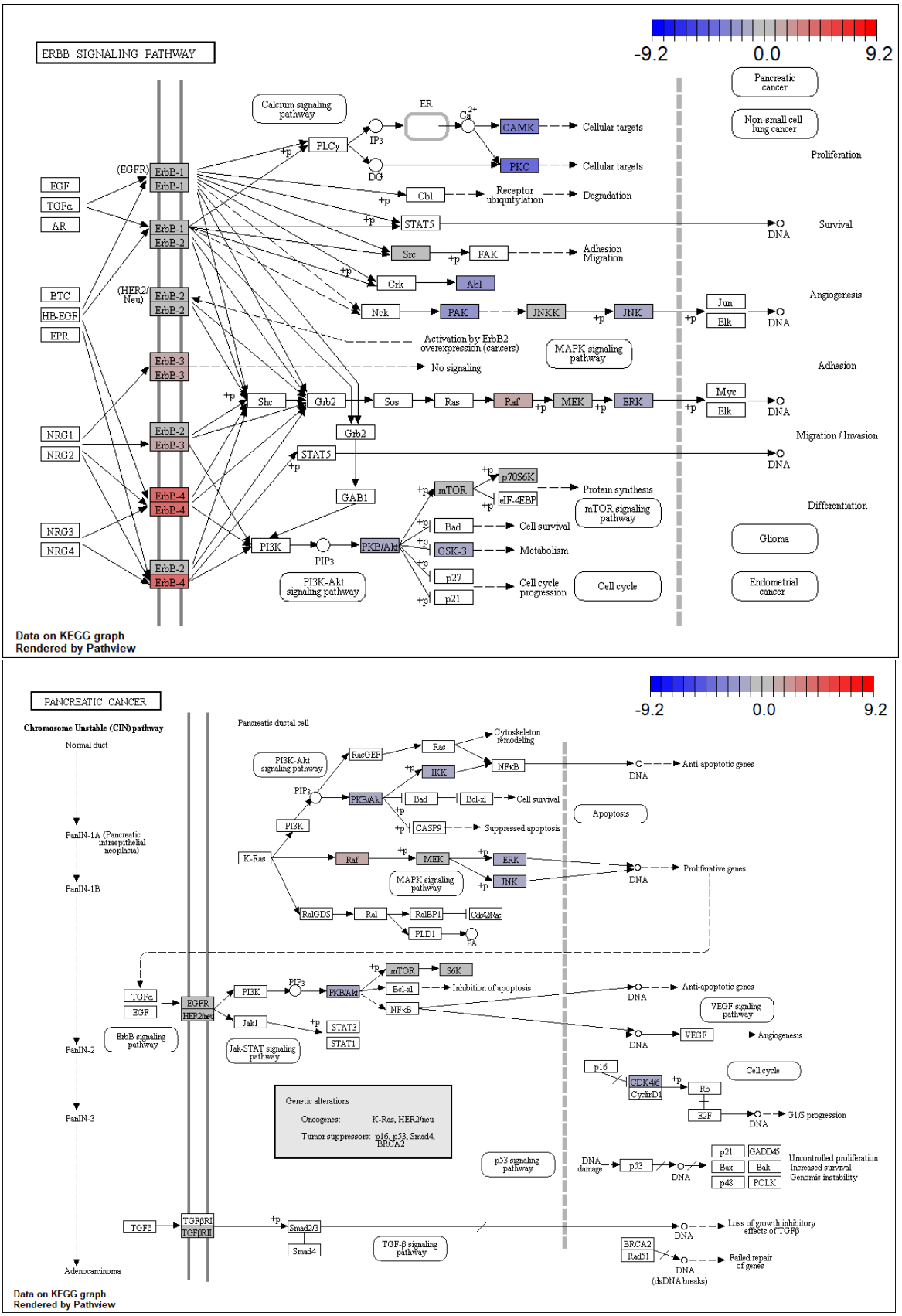


Figure 6c - Kinase Activity Prediction: shTMEM vs shControl

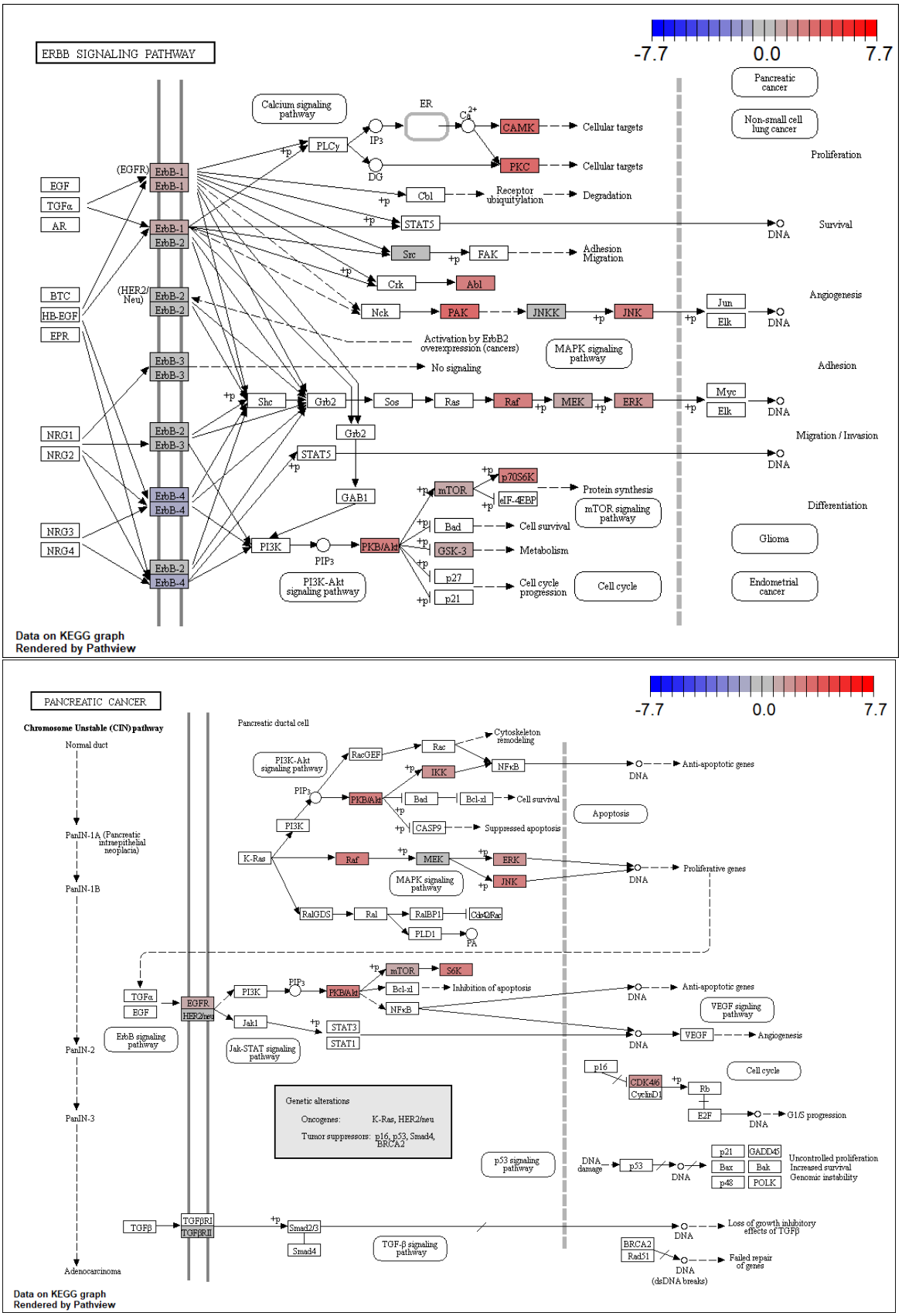
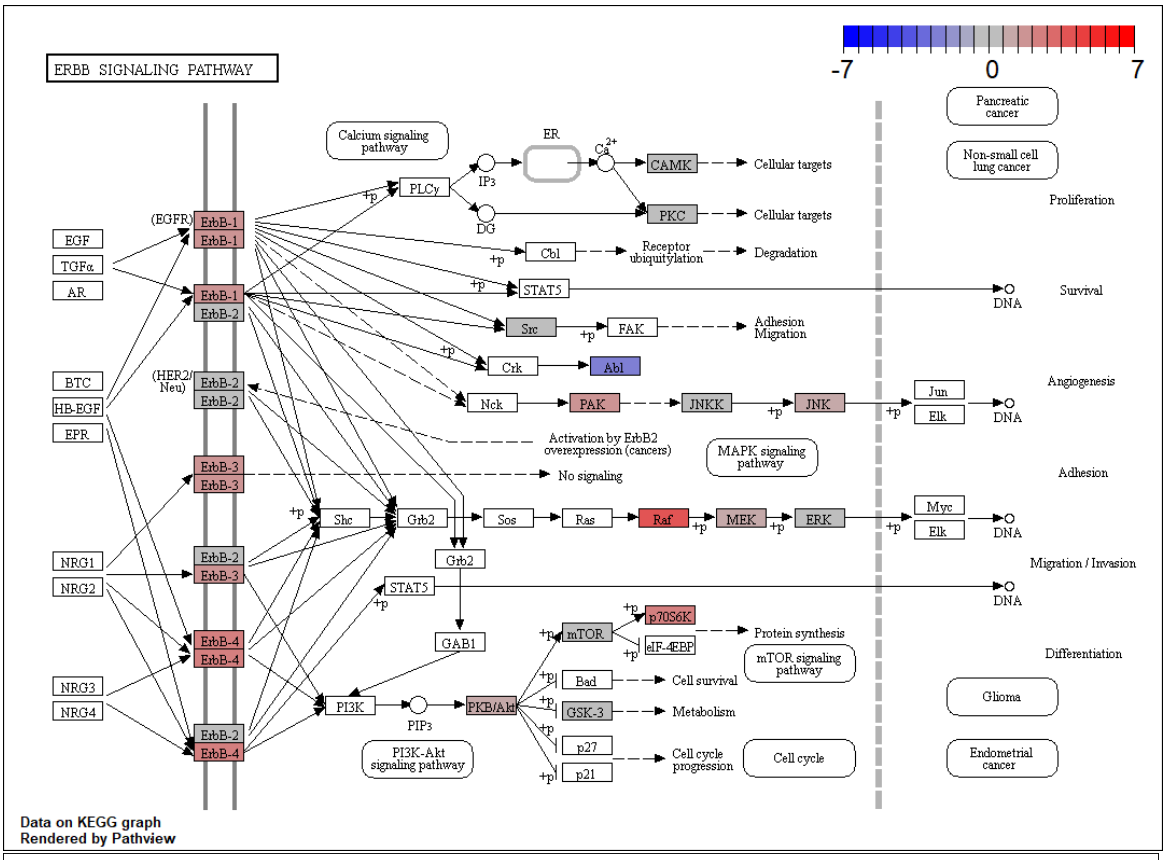


Figure 6d - Kinase Activity Prediction: shTMEM+EGF vs shControl



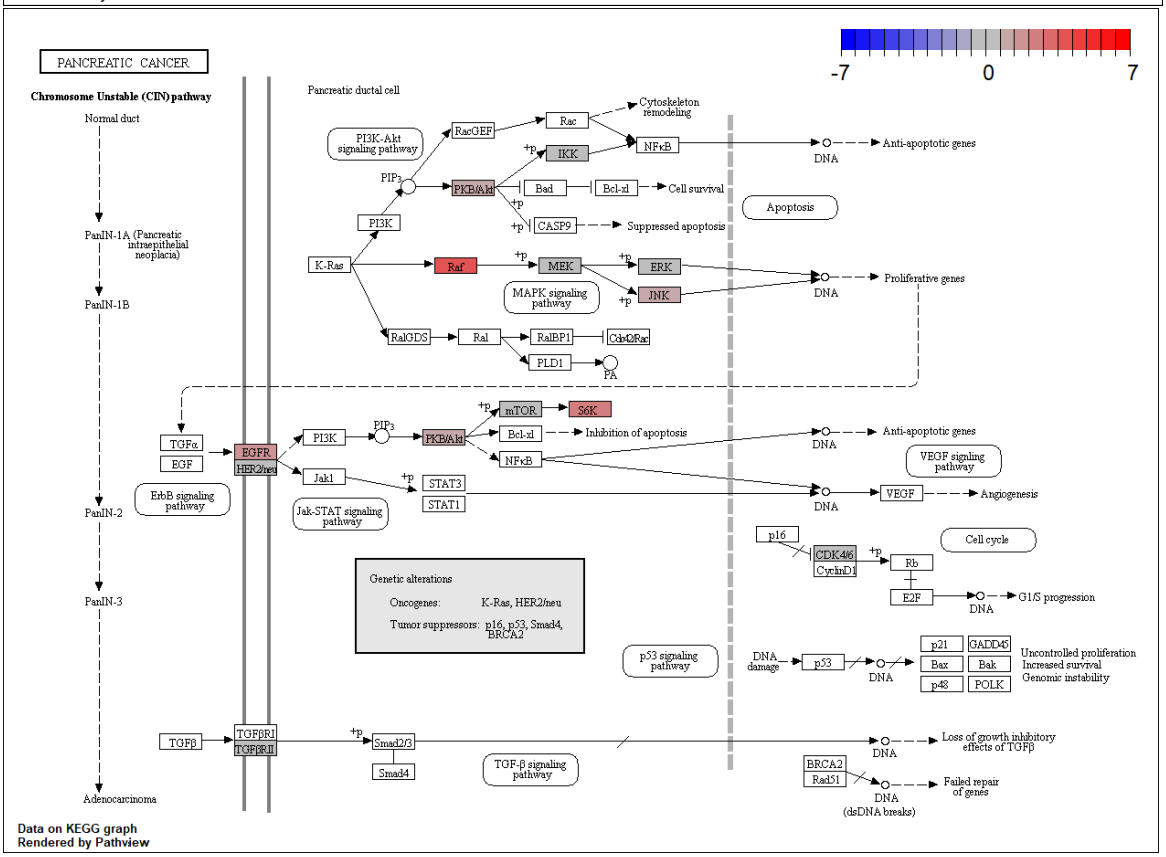


Figure 6 Discussion

To investigate the effects of TMEM16A depletion on EGFR signaling, kinase activities were inferred from our phosphoproteomic data. The kinase set enrichment analysis (KSEA) was applied as described by Casado et al. (Science Signaling 2013). The method operates under the notion that kinases are responsible for phosphorylating proteins and, therefore, their activities can be predicted from changes in the phosphorylation of their substrates when the system is perturbed. Once computed, these relative activities were mapped onto the ErbB signaling pathway and the pancreatic cancer model from KEGG. Kinases whose activities could not be inferred due to the absence of associated substrates have a white background. Expression of kinases in AsPc-1 was not considered in this analysis, meaning annotated kinases are not necessarily present in the cell.

Four different conditions were tested in **Figures 6a-d**. In **6a**, EGF stimulation turns on EGFR and activates downstream signaling as expected. In **6b**, EGF stimulation in shTMEM16A cells exhibits negligible effect on EGFR and, interestingly, down-regulates signaling activity. In **6c**, TMEM16A depletion produces a close version of EGF-induced signaling despite a weakly activated EGFR. In **6d**, the combination of TMEM16A knockdown and EGF treatment appears to activate EGFR but elicits a weaker downstream effect compared to TMEM16A knockdown alone.

In short, the KSEA analysis demonstrates that shTMEM16A cells mimic the increased EGFR signaling observed from EGF treatment in the control line. This is consistent with a previous study showing that TMEM16A depletion has no effect on proliferation. Perhaps more fascinating and novel is our finding that EGFR still responds to EGF even without TMEM16A, but the addition of EGF to shTMEM16A cells oddly negates the increase in shTMEM16A-induced EGFR response.