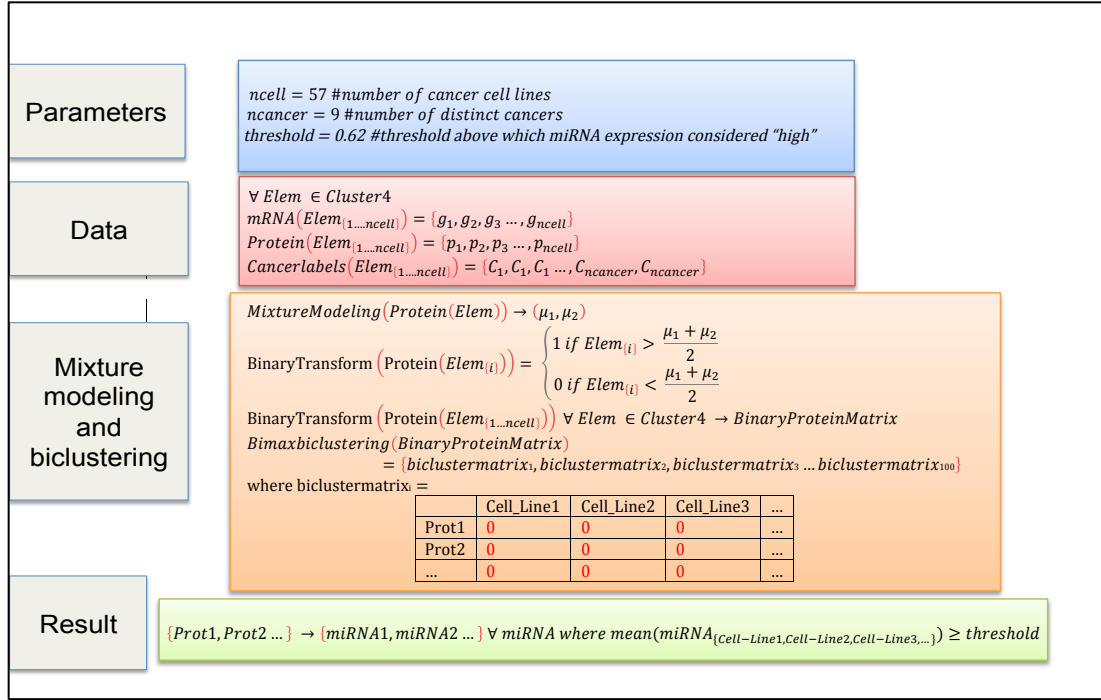Supplementary Appendix 2 (Biclustering). Modeling potential post-transcriptional regulation with mixture modeling, bi-clustering and miRNA integration.

Step1: Mixture modeling and bi-clustering - Since the defining characteristic of the fourth cluster were extremely high protein abundances for some cell lines and extremely low abundances for the rest, we employ mixture Gaussian modeling of protein abundances in all cell lines, for each transcript-protein relationship. The mixture modeling endeavors to fit two Gaussian distributions to the given data. The two Gaussian distributions report back two means ($\mu1$, $\mu2$) for each of the two Gaussians. The mid-point between $\mu1$ and $\mu2$ served as our linear separator to quantify the cell line data points as having "high" or "low" protein abundance. Thus, any data point of the corresponding protein, which reported expression larger than ($\mu1 + \mu2$)/2 was quantified as "high" protein expression (1) and the rest were quantified as "low" protein expression (0). Further, quantifying each data point as high (1) or low (0) converted the protein abundance matrix to a binary matrix, which was used for bi-clustering using the bi-max algorithm. The mathematical formulation for this method can be found in **Supplementary Appendix 2 Equation 1**.

Step2: miRNA integration - We hypothesize that, subgroups of low protein abundances across subgroups of cell lines, isolated from the bi-clustering analysis could shed light on potential miRNAs that serve as post-transcriptional regulators. To recover these gene-miRNA mappings, we search for highly expressed miRNAs in bi-clusters that reported as low abundance in the previous step. We generate these mappings by calculating the average of each miRNA's expression in cell lines that form the low abundance bi-clusters, and extract the miRNAs that report high expression. All average miRNA expression associated with the top 100 bi-clusters reported that the top 25% of the data was contained between normalized expression levels of 0.62 and 8.4. Thus, we use this threshold to label average miRNA expression as "high" if it is reported to be above 0.62.

Parameters	$ncell = 57$ #number of cancer cell lines $ncancer = 9$ #number of distinct cancers $threshold = 0.62$ #threshold above which miRNA expression considered "high"
Data	$\forall\, Elem \in Cluster4$ $mRNA\big(Elem_{\{1\dots ncell\}}\big) = \{g_1, g_2, g_3 \dots, g_{ncell}\}$ $Protein\big(Elem_{\{1\dots ncell\}}\big) = \{p_1, p_2, p_3 \dots, p_{ncell}\}$ $Cancerlabels\big(Elem_{\{1\dots ncell\}}\big) = \{C_1, C_1, C_1 \dots, C_{ncancer}, C_{ncancer}\}$

Mixture modeling and biclustering

$MixtureModeling\big(Protein(Elem)\big) \rightarrow (\mu_1, \mu_2)$

$$BinaryTransform\Big(Protein\big(Elem_{\{i\}}\big)\Big) = \begin{cases} 1 \; if \; Elem_{\{i\}} > \dfrac{\mu_1 + \mu_2}{2} \\[2mm] 0 \; if \; Elem_{\{i\}} < \dfrac{\mu_1 + \mu_2}{2} \end{cases}$$

$BinaryTransform\Big(Protein\big(Elem_{\{1\dots ncell\}}\big)\Big) \; \forall\, Elem \in Cluster4 \rightarrow BinaryProteinMatrix$

$Bimaxbiclustering(BinaryProteinMatrix)$
$\qquad = \{biclustermatrix_1, biclustermatrix_2, biclustermatrix_3 \dots biclustermatrix_{100}\}$

where $biclustermatrix_i =$

	Cell_Line1	Cell_Line2	Cell_Line3	...
Prot1	0	0	0	...
Prot2	0	0	0	...
...	0	0	0	...

Result

$\{Prot1, Prot2 \dots\} \rightarrow \{miRNA1, miRNA2 \dots\} \; \forall\, miRNA \; where \; mean(miRNA_{\{Cell-Line1, Cell-Line2, Cell-Line3, \dots\}}) \geq threshold$

Supplementary Appendix 2 Equation 1. miRNAs to gene set mapping of Cluster 4 members using mixture modeling and bi-clustering of protein matrix