

Details of each signal generation, included within the synthetic dataset, are enumerated below:

- a) **Unimpeded translation:** mRNA and protein abundances are highly correlated, implying efficient translation. NCI-60 proteogenomic dataset was sampled to randomly select paired mRNA expression and protein abundance with Spearman's correlation  $> 0.6$  (p-value  $<< 0.01$ ).
- b) **miRNA intervention:** A bimodal nature in the distribution of protein abundance across samples can result due to selective active miRNA targeting. To mimic this effect, the signals present a pattern, where protein levels in about 50% of the samples are non-zero (miRNA aided silencing in the remaining samples). NCI-60 proteogenomic dataset was sampled to randomly to select paired data points where mRNA expression varies across the entire dynamic range but approximately half of the data points correspond to no protein abundance.
- c) **Translation rate dependent on gene expression:** This trend shows groupings of samples at different values of mRNA and protein. A pattern such as this will exhibit varying translation rates and efficacies depending on different thresholds of gene expression. These may echo effects of degradation as well. The synthetic data corresponding to this pattern was generated by gridding the entirety of the 2D proteogenomic space (100 grids) and randomly selecting five grids from which paired data points (mRNA and protein abundance) were randomly sampled.
- d) **Random:** This trend consists of scatterplots with no distinguishable trend, essentially random values of mRNA expression and protein abundance paired together. Ideally we would expect these to be isolated and grouped as a separate cluster since there are no observable trends or indicative PTRs present.