### We can distinguish growth phase, carbon sources and Mg concentrations but not for Na concentrations from mRNA data

Results indicate data has a strong signal associated with growth phase with a correct prediction rate of 94% for exponential, 92% for stationary and 69% for late stationary phase. For carbon sources data let us to do moderate predictions. With a 94% correct prediction rate glucose is the best predicted carbon source. Glycerol and lactate follows it with 77% and 66% correct prediction rates. 64% of the runs gluconate is predicted as glucose and only for 24% of the cases gluconate is correctly predicted. There is also a strong signal associated with Mg levels we can predict low magnesium 81%, base magnesium and high magnesium 71% of the time correctly. On the other hand, the signal associated with sodium levels is weak and our pipeline is not able to distinguish high Na data from base Na. Pipeline predicts base Na correctly %97 percent of the time and predicts high Na correctly only 56% of the time. [figure 3]

### Overall performance of using protein data in prediction is less than mRNA data with an exception of carbon sources

Due to the results protein concentration is a slightly weaker predictor compared to mRNA concentration with the exception of carbon sources. For different growth phases test results provide a correct prediction rate of 82% for exponential, 77% for stationary and 49% for late stationary phase. Most clean signal is related with carbon sources in protein data. Glucose, glycerol and lactate are correctly predicted 93% 92% and 100% percent of tests. The worst prediction results are associated with gluconate. Gluconate is predicted correctly 66% of the cases. When it is not predicted correctly it is mislabeled as either glucose or glycerol with 20% and 14% of the cases. Mg levels are one another variable that lost predictability for protein data compared to mRNA data. Algorithm mixes high and low Mg levels 30% percent of the cases. On the other hand, base Mg, which is predicted correctly 93% of the cases, is distinct from other two. As for mRNA concentrations protein concentrations cannot be used to distinguish base sodium samples from high sodium samples. Base sodium is predicted correctly 98% of the cases and high sodium is predicted correctly for only 48 % of the cases. [figure 4]

**Signals are getting weaker as samples translate from exponential to stationary state.**

We also analyze the data by dividing the data into two time frames exponential and stationary; the hope in here is to investigate the distinguishability of conditions through time. Overall trends indicate that the cells become more similar as time passes for both mRNA and proteins. Also in general protein concentrations look more similar to each other compared to mRNA concentrations. [figure 5]. The individual trends mRNA lactate samples become more similar to glucose as growth phase changes from exponential to stationary, similarly low Mg protein samples begin to look more like base Mg. As an exception predictability of Na concentrations based on mRNA levels increases from exponential to stationary phase. Supplementary figures [1-3]

**Predicting multiple conditions from a single model is also possible.**

We can also make predictions for all four variables for the same data which generate a prediction space composed of 16 distinct conditions. The pipeline is similar for individual parameters and includes weight factor normalization for different sample sizes in training data also cost and gamma values are assigned in the same way with individual predictions in previous section. The results indicate with 152 mRNA samples we have a prediction score of 0.58 and with 105 protein samples we have a prediction score of 0.51 [figures 6, 7]

### Combining models by combining datasets increases the prediction ability slightly.

Finding the complementary information between mRNA and protein concentrations is an important aspect of the project. We directly combine mRNA and protein data after calculating size factors and batch effects independently for both mRNA and proteins and provide combined data to SVM. To compare the gains compared to individual data we calculate prediction tables for intersection set mRNA and protein data [supplementary figure 4 - 5]. The results indicate for 102 intersecting samples proteins are better for predicting combined conditions compared to mRNA with prediction score of 0.43 and 0.49 respectively. The prediction power increases slightly compared to mRNA when we add two datasets together with a prediction score 0.52 [figure 8]. But the drops because of the decrease in data size is more dominant than the gain of combining mRNA and protein data together, which indicated the number of samples is the main limiting factor for making predictions [supplementary figure 6].

**We can make continuous predictions by using regression**