Task

1. Clustering – only use sample from dilution series to avoid large interfere from other samples - two
   1. Direct clustering
      1. Clustering with high dimensional data – HCA, scikit learn methods -- Ximin
      2. Clustering with reduced dimensional data – *t-sne/PCA* + clustering method from a --David
   2. Trend clustering – Nozomi Derek
      1. Similar to alignment, compare each feature pairwise to check if they have similar/different trends, merge the similar trend features into cluster and dynamically update the cluster information as comparison going on
2. Assessment - two
   1. Each cluster from step 1 needs to be checked either use scatter plot/heatmap
      1. if the noise is properly detected
      2. if the trends are good clustered
      3. if the clusters too much
      4. etc.
   2. After clustering only from dilution series, add filters using other sample to verify if the features are from source or from background
3. Modeling – T the table – 1\*4
   1. Modeling using all clustered features together
   2. Modeling on separate clusters and merge all the features in the end while giving each model different weight