

The Samples

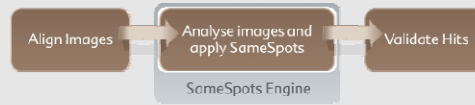


The two sample types are of *Haemophilus influenzae*. One treated with Actinonin (a natural occurring peptidyl-deformylase inhibitor) to provide a distinct treated sample.

20 technical replicates of each sample type were prepared under the same protocols.

The gels were post stained with Sypro Ruby.

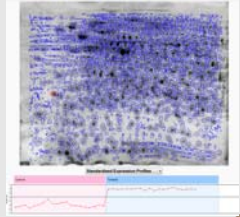
Samples and Analysis



The Analysis

The gels were manually cropped and then analysed fully automatically using Progenesis SameSpots. This produced a unified spot map of 2193 features. The normalised spot volumes for the 40 gels were then exported.

The resulting spot map is shown to the right.

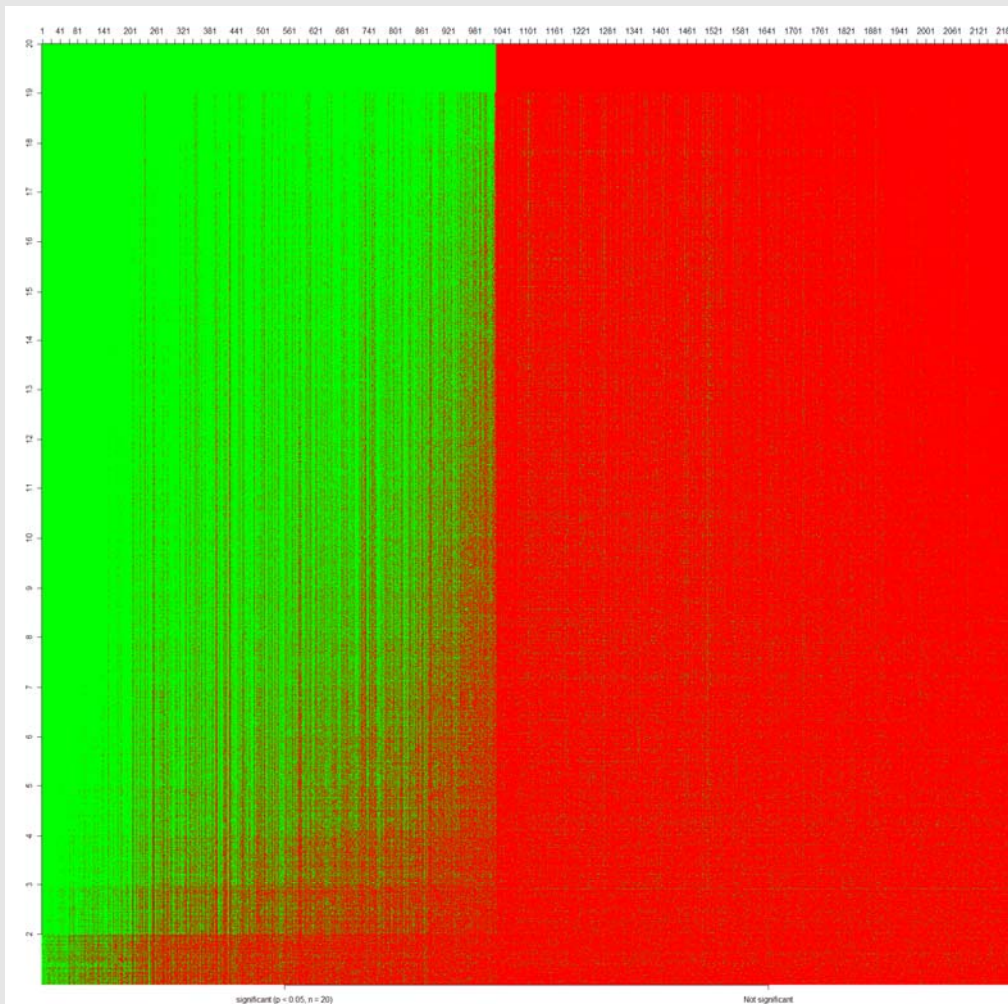


Results

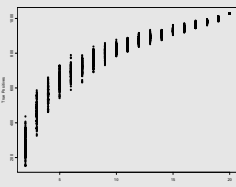
If we take the view that we have 20 different measures of the same sample we can use this large number of replicates to assess what the results of the statistical analysis would have been if we ran less replicates. To simulate an experiment where we ran 10 replicates we can randomly select 10 gels from the control group and then 10 gels from the treatment group and run a typical analysis. For this, we have simply run independent univariate ANOVA tests at each spot (we have not corrected for multiple testing as we wish to visually explore false discovery issues in this poster). We can run this sub-sampling procedure many times to get an idea of the stability of our statistical tests and also assess the impact the number of replicates has on the 'results' of our experiment.

As an analogy to how the sub-sampling works, imagine we had large volumes of 20 control and 20 treated samples. We decide we only want to run 10 gels so we randomly chose 5 control and 5 treated samples and completed the analysis. We have a lot of sample so we could do this again by simply randomly choosing more samples. Doing this to explore all of the possibilities would be an enormous task so as a shortcut we can do the sub-sampling after we have run the gels. All we need to do is assume that the replicates are reasonably independent of each other. This scheme allows us to create huge numbers of sample combinations for the lower replicate experiments. As the number of replicates increases the samples we take will obviously not vary as much (as they must contain the same samples more often) which can be seen as less variation in the high replicate experiments.

In the figure below 'green' means 'significant' in that uncorrected p was less than 0.05. 'Red' means 'insignificant' by the same criterion (i.e. $p \geq 0.05$). The numbers across the top correspond to the spot features detected. The 2193 spots were ranked by 'significant' (uncorrected $p < 0.05$) and 'non significant' and then by decreasing fold change. The sub-sampling analysis was run 100 times for each test replicate number (from 20 down to 2). As you can see the top 100 lines show no change - this is as expected and serves to show there are no issues introduced by the sampling procedure itself. The scale on the left shows how many replicates were in the sub-samples. By the bottom of the image we are comparing 2 controls to 2 treated. The ordering of the spots does not change so a red dot on the 'green' side shows a spot that was significant when we used 20 replicates that is not significant in the subset of samples selected. This can be classed as a False Negative with respect to the results from 20 replicates. Similarly a green dot in the 'red' side shows a spot calculated as 'significant' when it was not in the 20 replicate case. This is a False Positive.



Discussion and Conclusion



The results from this work show a powerful visualisation of why good experiment design and control of False Discovery rates is so important with proteomics data. It is even more striking that this large variation in the 'significant' spot lists comes from very well controlled technical replicates with no variation of spot measures. All we are doing is taking a subset of the data and applying exactly the same analysis as is relatively standard throughout proteomics. This sub-sampling analysis is relatively simple to perform on any data and can help to build confidence in the statistical results. This data has many very large expression differences so it is worth noting how many significant features (at the 20 replicate level) are simply missed with low numbers of replicates.

The graph to the left shows the True Positives (with respect to the results for 20 replicates) for each of the 1900 sub-sampling experiment runs. The graph on the right shows the same for False Positives.

It is worth noting that FDR control corrects for multiple testing by calculating a more stringent threshold p value that takes into account the fact we are running many statistical tests and provides a greater level of protection against the false positives we would expect from so many tests. It is obvious that this form of correction would potentially reduce the number of spots that are classed as significant. It will not change a spot that has been classed as 'non significant' to 'significant' so if we consider the left of the above figure we can see that any 'non significant' spots (red dots) are lost to us even if we apply post correction techniques.

