Identification of the ISAv7 genomic expression profile in the 07/10 44K Liver Microarray data

Brad Davis, Postdoctoral Fellow in the Molecular Genetics Laboratory, PBS

December 7, 2011

In this document I describe the steps taken to identify the genomic signature associated with the presence and intensity of ISAv7 (Infectious salmon anemia virus segment 7) as identified using TaqMan qPCR (Plarre primers) on the BioMark Fluidigm system. As many of the liver samples that were analysed for ISAv7 had been run previously on 4x44K microarrays, I was able to perform a retrospective statistical analysis of the microarray to determine whether salmon were responding strongly to the presence of the virus. The strength of the host response is one way to gauge the potential that the virus could cause disease and mortality, as avirulent pathogens that are not causing damage will not elicit a strong host response.

Step 1: Remove any genes from the microarray database where we have gene expression data for less than 50% of the samples.

Step 2. Regress the individual ISAv7 CT values against each individual's gene expression values observed on the microarray for each of the remaining genes.

Step 3: Determine which genes show a statistically significant difference in gene expression across the Ct scores using False Discovery Rate.

Step 4. Use hierarchical clustering of samples and genes to produce a heatmap of the expression patterns for the top 50 and 1000 genes, by P-value.

Step 5: Perform a functional analysis of the resulting genelists in order to understand the causes, consequences, and biological processes underlying the differences in expression patterns

Using the original data, 54 genes where identified at a false discovery rate of 20% whose expression pattern varied as a function of the Ct values observed from the Taqman qPCR results. However, examination of the resulting heatmap (Figure 1, below) shows that a number of samples scored as ISAv7 negative (CT scores manually set to 50), clustered tightly with the positive samples, while no positive samples clustered with the negative samples. The potential for false negatives was high, as we knew that some ISAv8 primer sets amplified in some fish that did not amplify with the ISAv7 primer set, likely due to sequence variation in locations where the primers and/or probe attach. Consequently, I identified the 'negative' (putative **false negatives**) samples which appeared more similar to *positive* than *negative* samples and removed them from the data set and repeated the analysis (e.g. Step 2-4 above).

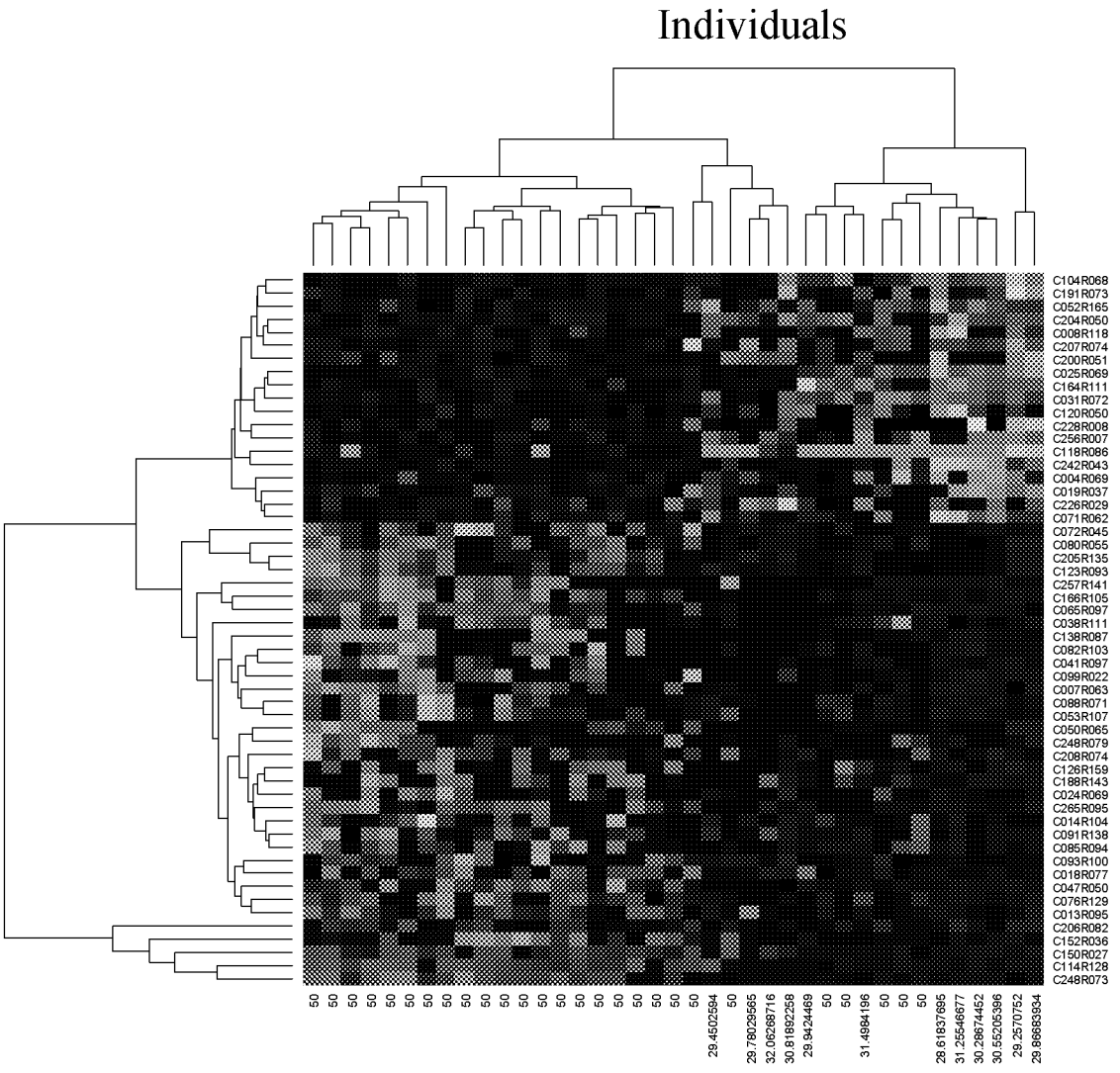

Figure 1: Heatmap produced by clustering the 54 genes identified from the original data analysis (includes all samples). Individuals are clustered on the x-axis, genes on the y-axis.

Following the removal of the putative **false negative** samples and repeating the data analysis, 5120 genes were identified at a false discovery rate of 5%, a substantially enhanced signal. Figures 2 and 3 (below) show the corresponding heatmaps for the top 50 and 1000 genes (as identified by P-Values) from the identified 5120 genes.

Note that all of the negative samples (e.g. Ct score = 50) cluster on the left side of the figures while all the positive samples are clustering on the right side (Figure 2).

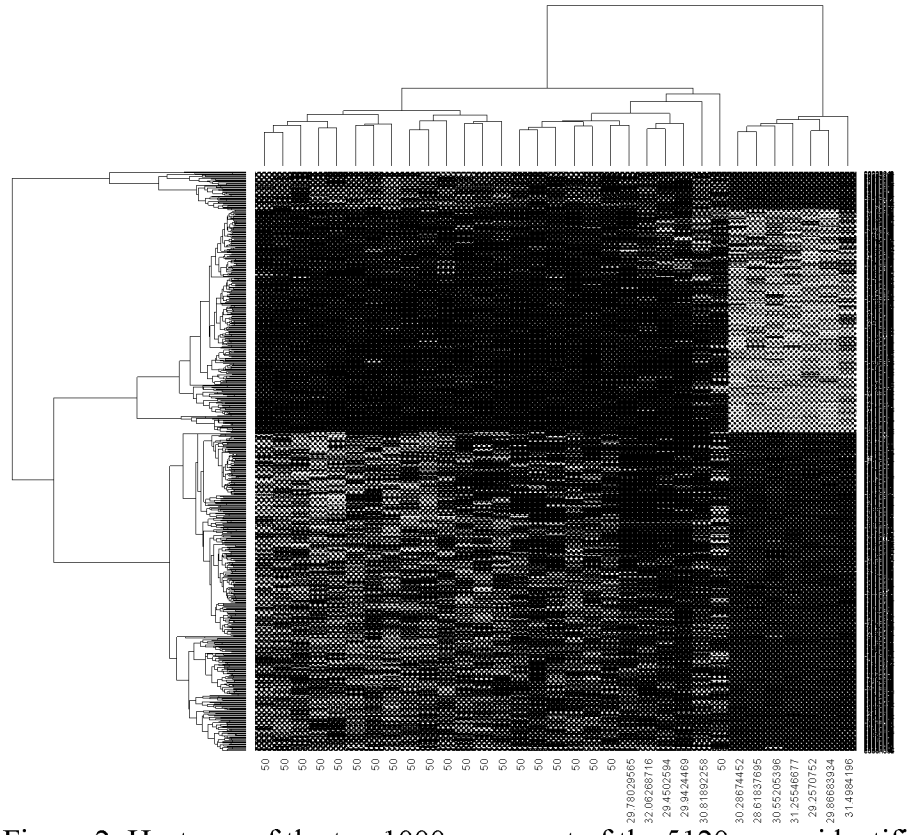

Figure 2: Heatmap of the top 1000 genes out of the 5120 genes identified after removing the putative false negative samples from the data. Individuals are clustered on the x-axis, genes on the y-axis.

A functional analysis of the total gene list (5120 genes) allows us to ascertain which biological processes and pathways are differentially affected. The program DAVID (a public genomics functional analysis package) identified "Influenza Infection" as the most enriched pathway (i.e. there was an over-representation of genes known to be associated with Influenza Infections in mammals), with 32 genes within this pathway identified at a false discovery rate (FDR)< 0.0001. This is a very strong signal indicating that fish positive for ISAv7 are responding to the virus similarly as mammals would respond to other influenza infections. As ISAV is an influenza virus, this result not only validates

the ability of microarrays to identify strong biologically relevant signals associated with health and condition on wild-caught migrating salmon, but it suggests that the virus is causing enough damage to elicit a strong response in the salmon.

Analysis of the genes that were down regulated in the samples positive for ISAv7 showed that mRNA processing, mRNA splicing, and genes associated with beta-interferon (a powerful mediator of anti-viral responses) were severely down-regulated. These patterns are highly consistent with influenza infections and are partially controlled by the virus itself. Thirty-eight genes connected to TGF-beta were also differentially affected, again consistent with known mechanisms used by influenza viruses to disrupt processes leading to cell death. Sialic acid is the receptor that influenza-viruses use to gain entry into the host cell. Five genes related to sialic acid where found to be up-regulated, and these tend to be associated with sialyltransferase (enzymes that transfer sialic acid residue he nascent oligosaccide) activity or neuraminidase activity (aka sialidase activity). Importantly, neuroaminidase activity is required for the influenza virion to be released from the host cell.

Analysis of genes up-regulated (stimulated) in samples shown to be positive for ISAv7 revealed that genes involved in the *negative* regulation of RNA processing and splicing are up-regulated (supporting the down regulation of these processes). There was also a strong immune response signal that included up-regulation of lymphocyte and leukocyte activity, activation of the humoral immune response, positive regulation of the innate immune response, viral transcription and replication pathways, and genes involved in host-virus interactions. There were also a significant number of apoptosis-related (cell death) genes up-regulated (156), again consistent with known host response patterns induced by influenza infections. .

Next we tried re-analyzing the data again, this time including the putative **false negative** samples, but manually changing their status to being positive samples and repeating the analysis. If these samples are truly **false negatives** then re-analyzing the data should produce a similar or stronger signal to the one found when the samples were removed.

Reintroducing "**false negatives"** increased the number of genes identified at a 5% False Discovery Rate from 5120 to 9302. Figure 5 (below) was very similar to Figures 2 and 3 (above). A functional analysis on the larger gene list of 9302 also results in Influenza being the top hit providing support for the hypothesis that these samples were actually positives that the qPCR tests were unable to identify.

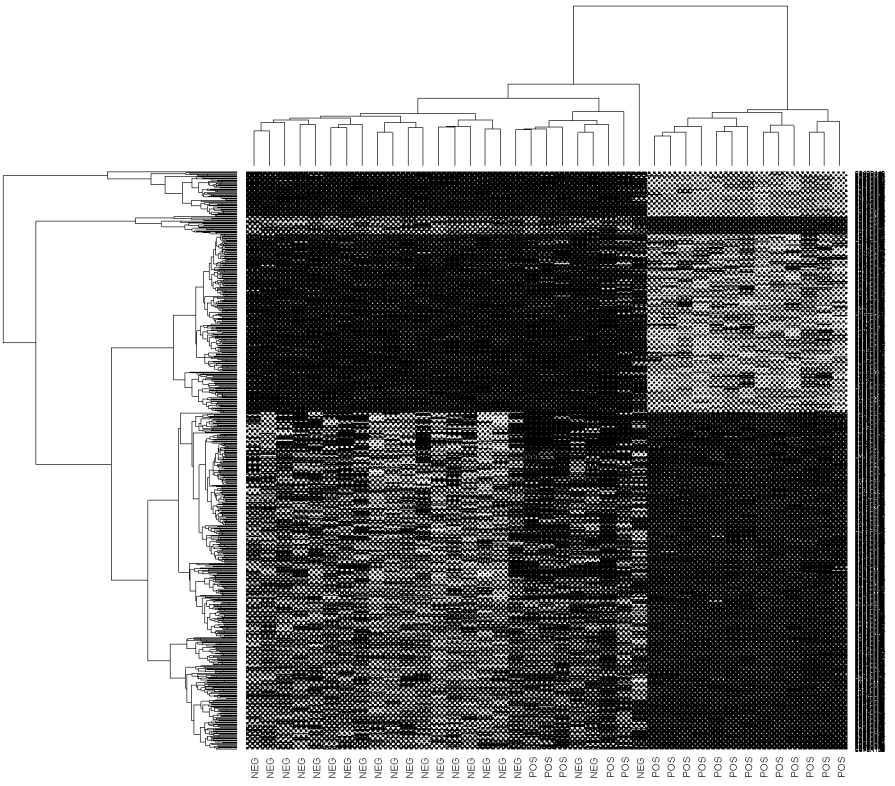

Figure 5: Heatmap of the top 1000 genes out of the 9302 genes identified after manually assigning the previously removed false negative samples to positives and re-running the analysis.

Note: A cursory functional analysis of the enhanced gene list (e.g. 9302) showed very similar patterns to those observed with the 5120 genes. A more detailed functional analysis on the longer list has not been attempted.

What we can take from these analyses is that salmon infected with the BC wtISAv orthomyxovirus are responding quite strongly and in a manner that is similar to responses to influenza viruses in mammals. Therefore, we cannot at this point assume that this virus does not cause disease in these fish. Follow up controlled laboratory challenge work is warranted.