Poster presentation

# Is the transcript from the overlapping region of two converging genes part of either gene?
Alastair Spence*

Address: University of Bath, Bath, UK

Email: Alastair Spence* - as@maths.bath.ac.uk

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

MYH11 (myosin heavy polypeptide 11) and NDE1 (Nude1) are transcribed from opposite strands of human chromosome 16 (see Figure 1). The null hypothesis is that microarray features representing a single gene will behave in the same way; estimating the confidence of our belief is an important step to identification of deviations from this model. One Affymetrix microarray probeset "227249_at", annotated as detecting NDE1, is designed for the region of
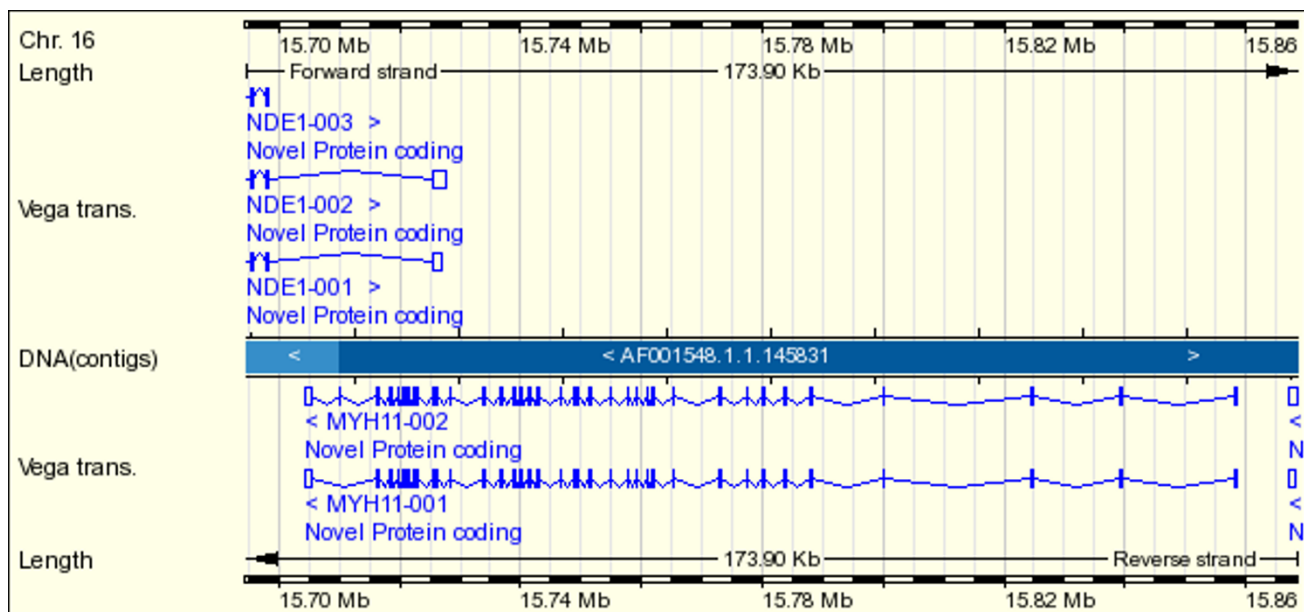


**Figure 1**
Transcription of MYH11 and NDE1

chromosome 16 where the two genes converge and is complementary to an identified part of MYH11 mRNA. The expression detected by this probeset is negatively correlated with some of the probesets for MYH11. We test the hypothesis that this region at the 3' end of NDE1 has a negative role in the expression of MYH11 and ask: "Is the RNA transcribed from here really part of the NDE1 or could it be an independent transcript?"

Microarrays continue to generate complex data for gene-expression. Clustering of both genes and samples is one of the most common analytical objectives – often achieved using spectral analysis of a matrix associated with the bipartite graph generated by the genes and samples and their corresponding links. Specifically, we first represent the activity of the $i$th gene in the $j$th sample as a positive value $w_{ij}$ and then store these values in a rectangular matrix W. Then clustering of both genes and samples may be achieved using the singular value decomposition (SVD) of the matrix W, with the singular vectors corresponding to the second largest singular values providing the information to implement the clustering. These clustering techniques are heuristic and it is natural to ask how reliable they are. Using techniques from numerical linear algebra and probability analysis, it is possible to provide a sensitivity measure of the robustness of clustering using SVD. We use this sensitivity analysis to provide an answer to the above question about the expression of MYH11 and NDE1.

The advent of microarrays for all exons leads to new possibilities in identifying alternative transcripts and changes in the composition of mRNA and proteins. With these possibilities comes the challenge of reliably identifying candidates for alternative splicing and where possible suggesting "clusters" of co-expressed exons which can then be tested in the laboratory. The mathematical techniques used in the above work can help in this process.