Analysis of the transcriptomes of wild-type and mutant *C. elegans*

Andrew Christopher Nelson

This dissertation is submitted for the
degree of Doctor of Philosophy
September 2008

Corpus Christi College
University of Cambridge

The Wellcome Trust Sanger Institute
Hinxton
Cambridge, UK

Declaration

I hereby declare that my dissertation contains material that has not been submitted for a degree or diploma or any other qualification at any other university. This thesis describes my own work and does not include work that has been done in collaboration, except when specifically indicated in the text.

Andrew C. Nelson

29/09/2008

Abstract

A key question in biology is how genotype can inform us of phenotype. For model organisms, most phenotypes reported have been at the level of the morphology and behaviour of the whole organism. However, recent advances in technology allow gene expression to be assessed on a genome-wide scale and pioneering work in yeast has shown that such expression profiles can be used as high density, quantitative phenotypes. I wanted to test whether expression profiles can also serve as useful phenotypes of whole animals rather than single cells. More specifically I sought to test whether the expression profiles resulting from perturbations of genes in one pathway looked more like those of other perturbations of the same pathway than another pathway. To do this I used two-colour DNA expression microarrays to survey gene expression in the nematode *Caenorhabditis elegans*. Expression profiles were produced for a number of different worm strains with mono-genic perturbations in different pathways involved in germline development. Clustering of the resulting expression profiles rediscovered the known pathways. This then allowed me to query perturbations of candidate modulators of EGF signalling against the compendium of expression profiles. I conclude that, as in yeast, expression profiles serve as reliable high-density phenotypes that allow meaningful biological comparisons to be drawn.

The quality of an expression microarray can only be as high as the gene annotations on which it is based. I therefore sought to evaluate how well characterised the transcribed genome of *C. elegans* is. To do this I used a combination of whole genome tiled microarrays and ultra-high density sequencing to assess the transcript complement of whole animals throughout development. We found that the vast majority (~95%) of expression is genic but the combinations and numbers of splice sites used are greater than previously predicted, suggesting that current annotations are largely complete, but that our knowledge of splice variation across development is still far from finished.

Whilst surveying transcripts in wild-type animals yields valuable data, it is known that there are many transcripts that are produced and subsequently degraded by the nonsense-mediated mRNA decay pathway (NMD). To identify these transcripts we compared the transcripts of wild-type animals to those of mutants of the NMD pathway. We find that ~13% of endogenous genes are NMD targets. The majority of these transcripts have upstream start codons in the 5' UTR or are alternatively spliced leading to a premature in-frame stop codon. Finally, we find that ~10% of all gene expression changes throughout development require NMD and thus that NMD is a bona fide regulator of gene expression.

List of Figures

List of Tables