Cis regulated Nonsense-mediated decay in human genome

*Hoi Fung Wong and Anze Zupanic | Institute for Ageing and Health | E-mail : h.f.wong@ncl.ac.uk | Student number : 100424009 |
Programme of study : Biomedical Science

Introduction and Background

Nonsense-mediated decay (NMD) is a mRNA surveillance pathway that prevents translation of transcripts with premature stop codons (PTCs) into malfunctioning proteins. Depending on the location of the PTC in the transcript and the properties of the truncated protein NMD is mostly beneficial, however in certain case it can lead to disease.

For NMD to occur, a stop codon must be recognised as premature and therefore differentiated from a genuine stop codon. Currently there are two hypotheses for the mechanisms of NMD: the faux 3'UTR model and the exon-exon junction complex (EJC) model (Figure 1).

Aim

The aim of this project was to use bioinformatical tools and databases to identify all NMD targets that have so far been characterized in the literature in yeast, mouse and humans and then to compare the ribosomal profiles of the targets as measured in recent studies. We hypothesized that, since NMD occurs during translation, patterns of ribosomal profiles could be identified that would eventually lead to the identification of the mechanisms of NMD.

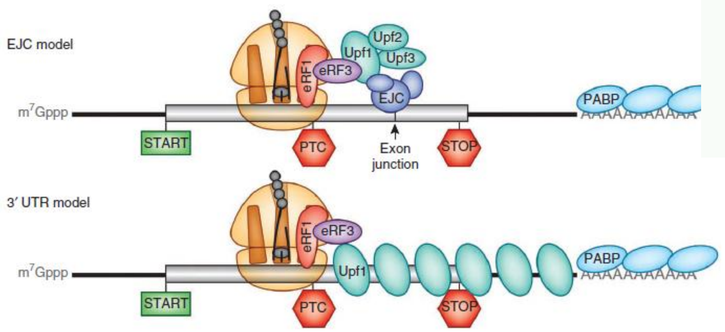

Figure 1. Graphical illustration of the two most common NMD models (1)

Results

About 500 yeast targets, 3000 mouse targets and 2700 human targets were compiled into individual list.

Over 2000 mouse NMD targets are analysed using the programme. The results are shown as a graphic which illustrate ribosome presence at different position of the mRNA. And these graphs were sorted into groups according to their features.

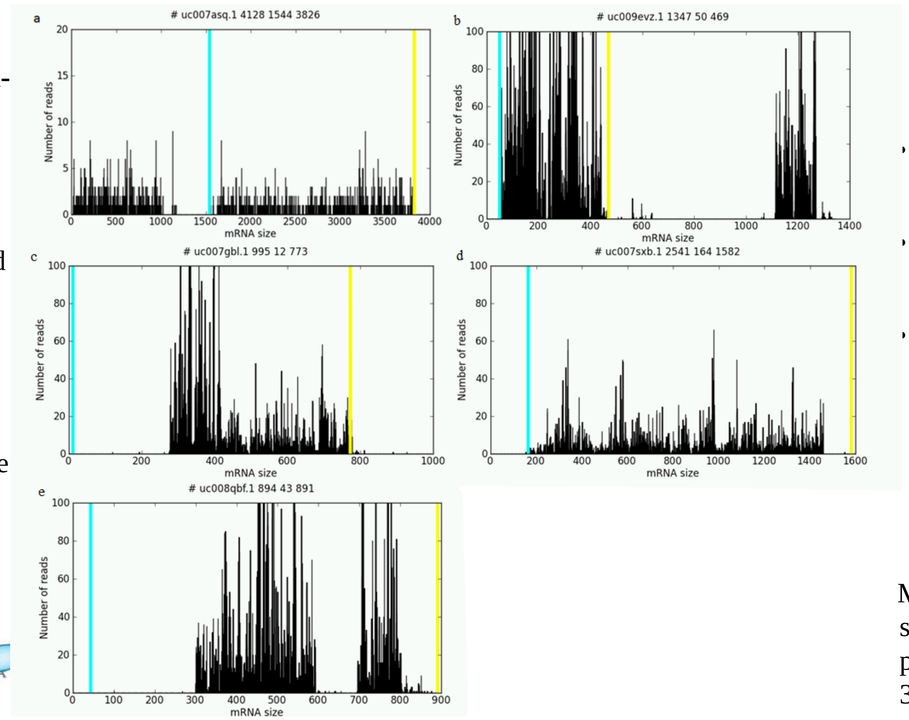

Figure 3. Ribosome profiles for different groups of NMD targets. The blue line represent the start of the ORF and the yellow one the end. a) abnormal reading on a uORF. b) has abnormal reading on 3'ORF. c) has a reading that does not start at the 5'ORF. d) has a reading that stops before the 3'ORF. e) has an abnormal pattern.

Methods

- NMD targets were identified from Cho *et al.*(2), Weischenfeldt *et al.*(3), Zhang *et al.*(4) and from several studies in the database Array Express. These targets were identified by comparing mRNA abundance in normal cell and in Upf1 and Upf2 knockouts.

  NMD targets' sequence information were obtained from Ensembl Biomart

```
#open file
l = open('C:/Users/Ivan Wong/Desktop/Placement/Lists of targets/Mouse
#read file
x = l.readlines()
#giving definition to the length of the list
listlength = len(x)
#Remove /n at the end of each line
stripped_x = map(lambda s: s.strip(), x)
#From the start till the end of the list
n = 0
while (n<listlength):
    y = stripped_x[n]+'_nt_counts.txt'
    #Open file
    f = open('C:/Users/Ivan Wong/Desktop/Placement/fp_mesc_nochx/'+y,
    #Read file into a variable
    list = f.readlines()
    #Close file
    f.close
    #Delete the 1st item in the variable
    del list[:1]
    #Remove the symbol at the end of the string
    list = map(lambda s: s.strip(), list)
    #Convert strings into intergers
    interger = map(int, list)
    sum = 0
    for i in range(len(interger)):
        sum = sum + int(interger[i])

    s = sum
    print y
    print s
    if s >=10000:
        import numpy as np
        import pylab as P
        import matplotlib.pyplot as PP
        #Open file
        q = open('C:/Users/Ivan Wong/Desktop/Placement/fp_mesc_nochx/
```

Figure 2. One of the Python codes written to create graphical results

- Ribosome profiling data are from Ingolia *et al.* (5) and GEO database with accession number GSE30839
- Algorithms for the analysis of ribosome profiling reads and position of codons were written in Python (Figure 2)
- Manual sorting of results into groups according to their features. These groups are abnormal reading on uORF (Figure 3a), abnormal reading on 3'ORF (Figure 3b), reading does not start at 5'ORF (Figure 3c), pre-mature stop (Figure 3d) and abnormal pattern (Figure 3e).

Conclusion and Discussion

Most of the results showed a general pattern expected to see in a NMD target e.g. a sudden drop of reads after encountering a stop codon. There are some strange patterns that cannot be explained with the current knowledge on NMD e.g. figure 3e.

These results alone cannot verify the precise NMD mechanism but with more experimental data or Bioinformatics analysis, it is possible to establish a model that explain NMD in-depth.

Reference

(1) C.J. Shoemaker & R. Green, 'Translation drives mRNA quality control' , Nature structural & molecular biology Vol 19, no. 6, June 2012

(2) H. Cho, K.M. Kim, S. Han, J. Choe, S. G. Park, S.S. Choi, and Y.K. Kim, 'Staufen1-Mediated mRNA Decay Functions in Adipogenesis', Molecular Cell Vol 46, 495–506, May 2012

(3) J. Weischenfeldt, J.E Waage , G. Tian, J. Zhao, I. Damgaard, J.S Jakobsen, K. Kristiansen, A. Krogh, J. Wang, B.T Porse, 'Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns', Genome Biology Vol 15 Issue 5, May 2012

(4) Z. Zhang, L. Zhou, L. Hu1, Y. Zhu1, H. Xu, Y. Liu, X. Chen, X. Yi, X. Kong, L. D Hurst, 'Nonsense-mediated decay targets have multiple sequence-related features that can inhibit translation', Molecular Systems Biology 6:442, Oct 2010

(5) N.T. Ingolia, L.F. Lareau, J.S. Weissman, 'Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes', Cell, 2011