Thema 7 Project Proposal

O. Gkourlias - 420172

BFV2

21-01-22

M. Kempenaar

# The study & paper.

This sequencing study focused on a certain frameshift mutation which changes the transcription rate of genes. The transcription factor in question is called SP1. The frameshift causes a heterozygous variant going by the name of c. 1995delA. This study has observed that this pathogenic variant may play an important role in causing a multitude of genetically inherited syndromes, or surely heightening the chance of acquiring a syndrome. This set of syndromes falls under a group named BMF, bone marrow failure. Inherited, in the case of this study.

There’s importance in recognizing the findings of this study, because BMF syndromes are very predisposing to oncogenesis, and the syndromes can even be life threatening on their own. By examining the variant with the ordinary samples of the SP1 gene, it was possible to determine how the gene expression is affected. In this case, how the variant causes defective haematopoiesis, which is characteristic of BMF.

The genes of a large family consisting of patients with hematologic abnormalities has been selected for this study. The discoveries previously stated about SP1 were recognized in the exomes of the family members. The variant was also present in the genes of a patient who passed away because of acute leukaemia.

To further study the effect of the variant, an experiment had been performed with mice specimen. Altering SP1 in these mice led to similar effects to those that were seen in the human family. There was a progressive impairment of haematopoiesis in these mice

# The Project plan.

After the examination of the data provided by this study, it’s possible to compare the different levels of gene expression between the human reference genome and the variant. To achieve this, Rstudio will be used in analysing the datasets. The significance of certain expression levels will be examined. Is a specific gene expressed more than it should be, or has it been altered because of the frameshift? It’ll be the goal of this project to answer such questions.

To answer the questions, a distinction must first be made as to which genes are abnormally expressed. To do this, the family members their transcription amounts will be compared to that of the most common alleles. When these genes have been selected, the most abnormal occurrences may be examined to determine whether it affects the final protein. The study had in fact detected a total of 1247 genes which were expressed differently.

# The data

The data provided consists of a raw counts TSV file. To gain further insight in what the data looks like, the first five rows will be pasted down here.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 4275 | 4277 | 4279 | 4280 | 4280a | 4281 |
| ENSG00000000003 | 23 | 30 | 11 | 43 | 31 | 8 |
| ENSG00000000005 | 0 | 0 | 0 | 0 | 2 | 0 |
| ENSG00000000419 | 778 | 910 | 838 | 911 | 1113 | 1051 |
| ENSG00000000457 | 378 | 438 | 441 | 772 | 738 | 389 |
| ENSG00000000460 | 44 | 51 | 58 | 61 | 65 | 28 |

The first row indicates which sample is being used. 4275, 4277, 4281 are the PS1 mutant types, while the others are wild type expression used in the comparison process. The differing amounts of expression of the genes, which are referenced by their ensemble id, are noted by raw and non-normalized numbers. Loading these numbers into RStudio allows for the comparisons to be made.