**Bioinformatics Exercise 3| BIOL3120**

In this exercise, you will use what you have learnt in exercise 1 & 2 to identify and assess genetic variants. This exercise is based off research conducted in the Macquarie University Centre for Motor Neuron Disease Research.

**Learning objectives**

At the end of this exercise, you should be able to:

* Independently identify and name genetic variants
* Independently analyse predicted variant pathogenicity
* Interpret functional data from different variants

**The story so far…**

In 2016, Dr Williams et al., published evidence that mutations in *CCNF*, the gene encoding the Cyclin F protein, are linked to motor neuron disease (Williams et al., 2016, <https://www.nature.com/articles/ncomms11253>). However, at the time and in the years since, we needed to identify which novel variants are/were pathogenic and which were likely to be benign, naturally occurring SNPs in the general population. In this practical, you will work through some of our workflows to identify and analyse the *CCNF* variants (published by Dr Lee et al., <https://www.frontiersin.org/articles/10.3389/fnmol.2021.627740/full>).

**Part A - Variant identification**

This work requires you to use skills learned from bioinformatics exercise 1.

You will need to import the reference cDNA sequence in benchling and covert this to the coding sequence. See bioinformatics exercise 1 for details.

Sample *CCNF* sequences are provided on iLearn. Download these sequences, align these to your reference CCNF coding sequence and complete the table for each of your sequences, using standard variant naming rules.

|  |  |  |
| --- | --- | --- |
| ***CCNF sample sequence*** | **DNA variant** | **Effect on protein** |
| Sample A | None | None |
| Sample B | c.1861A>G | p.S621G |
| Sample C | c.583A>C | p.S195R |
| Sample D | c.1721G>A | p.R574Q |

**Part B – *In silico* Variant Analysis**

This work requires you to use skills learned from bioinformatics exercise 2.

*In silico* analysis tools:  
<https://gnomad.broadinstitute.org/>  
<https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html>  
<http://genetics.bwh.harvard.edu/pph2/>  
<https://www.ncbi.nlm.nih.gov/>  
See Bioinformatics exercise 2 for details.

For each sequence sample in your table, work through an *in silico* analysis and classify each piece of evidence as whether it suggests the variant is pathogenic, benign, or uncertain. Colour evidence accordingly, with red suggesting pathogenic, green suggesting benign, and black uncertain.

Using the information present in your table, for each variant write a preliminary conclusion of benign, possibly benign, uncertain, possibly pathogenic, or pathogenic.

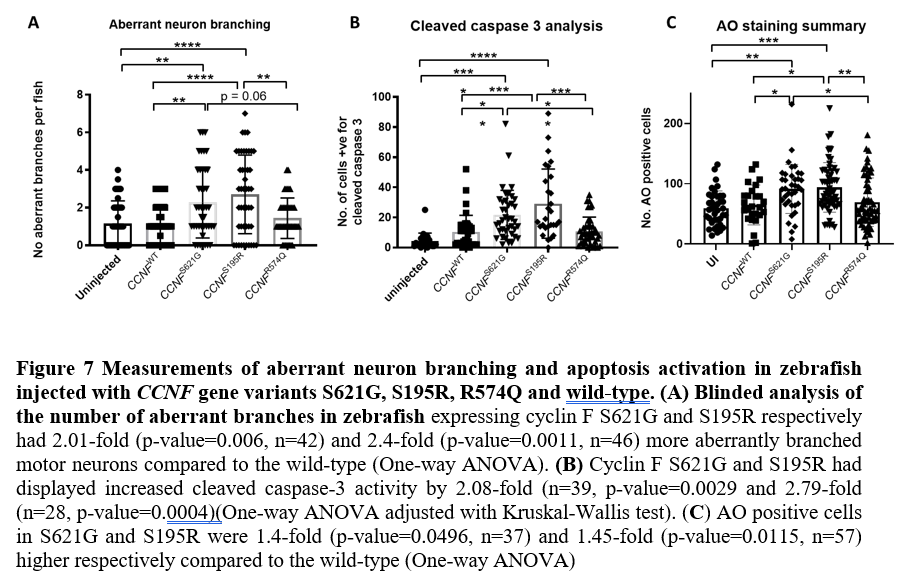
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***CCNF sample sequence*** | **DNA variant** | **Effect on protein** | **Gnomad allele frequency** | **Sift protein affect prediction** | **Polyphen protein affect prediction** | **Conclusion from evidence** | **Clinvar pathogenicity rating** |
| Sample A | None | None | N/A | N/A | N/A | Not pathogenic | N/A |
| Sample B | c.1861A>G | p.S621G | 4.29e-5 (Pathogenic) | AFFECT PROTEIN FUNCTION (0.01) | 0.227 benign | Unsure | Pathogenic |
| Sample C | c.583A>C | p.S195R | 7.07e-6 (Pathogenic) | AFFECT PROTEIN FUNCTION (0.00) | 0.684 possibly damaging | Unsure | Pathogenic |
| Sample D | c.1721G>A | p.R574Q | 6.90e-5 (missense) | AFFECT PROTEIN FUNCTION (0.00) | 1.00 probably damaging | Unsure | Not in dataset |

**Part C – *In vivo* Variant Analysis**

This work requires you to interpret and demonstrate an understanding of the primary scientific literature.

The Macquarie University Centre for MND ran these variants through a workflow to screen pathogenic gene mutations (Chen, De Luca et al., 2021, <https://doi.org/10.3389/fnmol.2021.627740>). Below is a sample of some the results from the zebrafish studies. Please note that this work was performed in compliance with the Animal Ethics Committee and Biosafety Committee, Macquarie University (AEC Reference No. 2015/034-29; NLRD 5974). As always, *in vitro* and *ex vivo* work was performed on cell culture material before moving into animal work.

It has been previously reported that expression of MND linked mutations, but not wild-type or non-pathogenic SNPs cause aberrant motor neuron branching and activation of apoptosis pathways in embryonic zebrafish models of neurodegenerative disease (Bosco et al., 2010; Kabashi et al., 2010; Laird et al., 2010). We therefore performed these assays with the sample sequences that you have been provided. For details of the experiments, please see Hogan et al., 2017 (<https://academic.oup.com/hmg/article/26/14/2616/3746877>)



How does this functional *in vivo* analysis compare to your *in silico* analysis?

|  |  |  |  |
| --- | --- | --- | --- |
| ***CCNF sample sequence*** | **DNA variant** | **Effect on protein** | **Phenotype in zebrafish?** |
| Sample A | None | None |  |
| Sample B | c.1861A>G | p.S621G | 2.01 fold abberant branches |
| Sample C | c.582A>G (this prac) c.585T>G (Williams, 2016) | p.S195R | 2.4 fold abberant branches |
| Sample D | c.1721G>A | p.R574Q |  |

What could sample D be?