*Aniket Kuhikar, UWL*

*rearranging the pattern of intron and exon elements that are joined by splicing to alter the mRNA*

splicing.

This practical will show you how to assess sequence variants and their predicted effect on

isoforms could be related to the change in protein function that you are trying to assess.

your samples have the sample protein isoform. There may also be a potential that differing

assess the cause for change in protein function, it is a good step to initially ensure that all of

When assessing proteins as a bioinformatician, especially for studies where you want to

*coding sequence.*

*protein variants (isoforms) that may have different cellular functions or properties. It occurs by*

*Alternative splicing is a process that enables a messenger RNA (mRNA) to direct synthesis of different*

protein. It is defined in Nature as follows:

Alternative splicing is a process that results in the production of different isoforms of a

Introduction

*Introduction to Genetics and Genomics*

*MSc Bioinformatics*

*Predicted Effect on Splicing using SpliceAI*

*Annotation of Genetic Variants with their*

for the whole genome. It consists of 2 sections, the VCF header, which contains

Protocol:

1.

Due to the large scale production of genotyping and sequencing data from major

projects such as the 1000 Genomes Project, files in Variant Call Format (VCF) which

are tab-delimited text files used in bioinformatics have been developed and

commonly used for storing gene sequence variations without having to store data

•

metadata explaining the contents of the body of the file; the body follows the

header. The body consists of a minimum of 8 mandatory columns (more optional

columns are often present).

2.

The Download the provided VCF files for Sample 1 and Sample 2. These files contain

only the VCF data for one gene, AOC1, therefore it can be opened in EXCEL. VCF files

can be extremely large and therefore will often need to be analysed

programmatically.

Two VCF files (AOC1\_Sample1.vcf and AOC1\_Sample2.vcf)

•

SpliceAI web interface (https://spliceailookup.broadinstitute.org/)

•

For future reference – a command line version of the tool is also available

•

(https://github.com/Illumina/SpliceAI)

Resources and Materials:

1000 Genomes Project - About variant identifiers:

•

https://www.internationalgenome.org/category/vcf/

What is a splice donor and splice acceptor site?

•

http://www.imgt.org/IMGTeducation/Aide-memoire/\_UK/splicing/

SpliceAI publication for further information about this application.

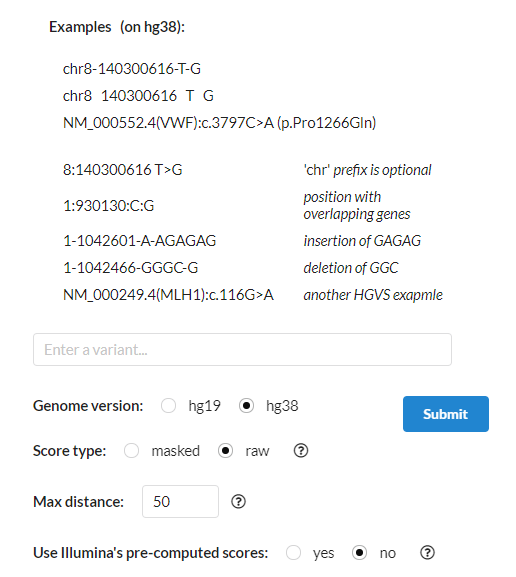


Figure 1: SpliceAI web interface. Enter the variant data in the input box using the format shown in

the examples.

box where it says “Enter a variant”.

3.

Examine the contents of the VCF header and the body of the file. Are there any

differences between the contents of the two files?

4.

Visit the SpliceAI webpage, the ID region of the VCF files contain information of the

genetic variants present in the samples. Enter this information into the SpliceAI input

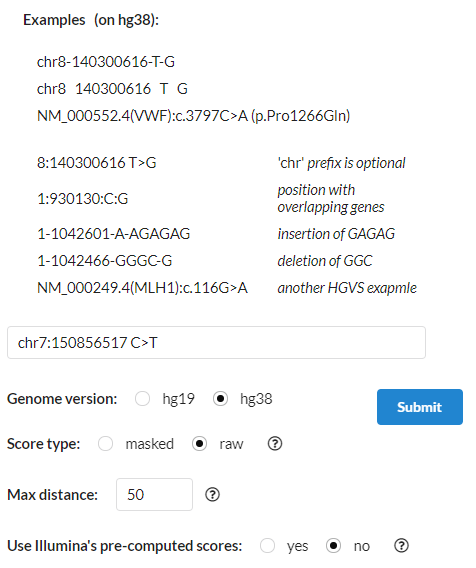


Figure 2: The input box containing the variant information. Hover the mouse over the question

marks for further information regarding the options.

5.

Read the additional information in each section of the options by hovering the

mouse over the question mark (?). Click submit once the variant information is

entered.

the Github).

independently (requires an Ubuntu terminal, other dependencies are explained on

included in the resources if you wish to explore the command-line version

use command-line tools for many tasks. The Github page for SpliceAI has been

datasets rather than just two samples, hence the preference of Bioinformaticians to

very repetitive and time consuming. Usually as Bioinformaticians, we work with large

You may have observed, that analysing each variant for each sample, one at a time is

8.

samples? What conclusions would you draw from this analysis?

Can you observe any difference in the predicted splicing scores between the two

Read the provided resources to obtain more information to interpret your results?

7.

Use the resources linked at the top of the page to help you interpret your results.

variant into an EXCEL file. Do this for both samples and then compare the results.

The results will appear at the bottom of the page. Extract the results from each

6.