

## Section 1 - Genomic Variant Analysis

### Exercise 1.1:

Identify the number of variants and samples in provided VCF file

(<http://www.1000genomes.org/wiki/Analysis/Variant%20Call%20Format/vcf-variant-call-format-version-40>)

```
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001
NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ
0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:,,
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50
0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ
1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
```

### Exercise 1.2:

Find the genotype of each sample for rs6054257 variant (from above provided VCF data).

### Exercise 1.3:

Analyze the following variants using Variant Effect Predictor (VEP)

(<http://www.ensembl.org/common/Tools/VEP?db=core>):

```
1 182712 . A C . . .
3 319780 . GA G . . .
19 110747 . G GT . . .
```

List high impact variants and affected genes.

### Exercise 1.5:

Are these CNVs present in controls? Check in DGV database

<http://dgv.tcag.ca/dgv/app/home?ref=NCBI36/hg18>

```
X:147729896-147754425
16:15387380-16270740
16:14960247-16307313
16:14808871-15935225
```

## Section 2-Gene Ontology and Pathway Analysis

### Exercise 2.1:

Draw a Protein Protein Interaction (PPI) network using [String](#) with provided gene list. Which type of evidence supports the interaction between **MBD5** and **DOCK8** proteins?

### Exercise 2.2:

Draw a Protein Protein Interaction (PPI) network using String with experimentally supported protein-protein interactions and exclude non-interacting proteins.

### Exercise 2.3:

What type of information is provided by STRING in “**Analysis Tab**”?

### Exercise 2.4:

For the same gene list find the most significant top 3 GO terms with adjusted p-value (Biological Processes and Molecular Function) using Enrichr.

### Exercise 2.6:

Are KEGG pathways associated with ASD? If yes then why?

### Exercise 2.7:

Which of your genes are involved in the most significantly enriched pathway (only for Reactome database).

### Exercise 2.8:

Which Human Phenotype Ontologies are significantly enriched for your gene list (top 5)?

### Exercise 2.9:

What are the differences in between both tools (Enrichr or String) regarding the information and usage? In your opinion, which tool is better?

### Exercise 2.10:

Prepare a Tree map for all GO Biological processes terms (only significant:  $< 0.05$ ) from Enrichr using the ReviGO tool (<http://revigo.irb.hr/>).