Table summary

1. Files contains in each variant folder

|  |  |
| --- | --- |
| Files | Meaning |
| XXX.xtc | The trajectories of the protein. |
| index.ndx | The index file of the protein (Use “1 Protein” from the index file) |
| Topol.tpr | Topology for the protein |

1. GROMACS command to extract basic trajectories file

|  |  |
| --- | --- |
| Commands | Meaning |
| 1 1 | gmx rms -f XXX.xtc -n index.ndx -s topol.tpr | Extracting the RMSD based on the protein structure. |
| 1 | gmx trjconv -f XXX.xtc -n index.ndx -s topol.tpr -b 0 -e 0 -o conf.gro | Extracting the first frame of the protein and the file name as “conf.gro” |
| 1 | gmx gyrate -f D132H.xtc -n index.ndx -s topol.tpr | Extracting the gyration based on the protein structure. |
| 1 | gmx rmsf -f D132H.xtc -n index.ndx -s topol.tpr | Extracting the RMSF based on the protein structure. |
| 1| gmx sasa -f D132H.xtc -n index.ndx -s topol.tpr | Extracting the SASA based on the protein structure. |

1. Number of variants

|  |  |  |  |
| --- | --- | --- | --- |
| Protein | Benign variants | Pathogenic variants | Variants of Uncertain Significance |
| TP53 | 24 | 81 | 1143 |
| MLH1 | 8 | 44 | 447 |
| MSH2 | 12 | 38 | 1522 |
| MSH6 | 18 | 25 | 1275 |
| MUTYH | 55 | 63 | 175 |

1. Original data source

|  |  |
| --- | --- |
| Protein | Database |
| TP53 | ClinVar, COSMIC |
| MLH1 | ClinVar, UMD, LOVD |
| MSH2 | ClinVar |
| MSH6 | ClinVar |
| MUTYH | ClinVar, TCGA, COSMIC, LOVD |

1. Variant location

All the protein variants are located in the TP53\_variants.csv, MLH1\_variants.csv, MSH2\_variants.csv, MSH6\_variants.csv, and MUTYH\_variants.csv file. Each of the file contained the heading “Genome position (GRCh37)”, “Nucleotide”, “Amino Acid”, “Location”, and “File location”. The csv file can be read using the python pandas module, “read\_csv” function.