***TP53* MHPA pipeline documentation**

1. Include in one folder:

Python files:

* Analyzer.py
* v12-q50.py
* merger.py
* zeroscreator.py

Text files:

* genomeADalts.txt
* genomeADfreqs.txt
* genomeADposns.txt
* genomewithinst.txt
* exonCoordData.txt
* nt\_coding.txt
* chrnum.txt

Matlab files:

* Matlab\_commands\_3\_16\_GenomADupdate.m
* matlab\_input\_genomeADcoord.mat

Other files:

* MHPA\_TP53.bed
* dataWithaa.xlsm
* indelmacro.xlsm
* all .bam/.bai files to be analyzed (aligned to hg19)

1. Analysis using macOS: Use ‘python Analyzer.py’ command in Terminal. In the Analyzer window (see screenshot below) choose the *TP53* gene location (chr17:7571720-7590868), and specify the minimum variant allele frequency for SNV calls (‘Minimum Variant Allele Frequency’ field), the minimum variant allele frequency for indel calls (‘Minimum indel frequency’ field) and the minimum read count for the variant allele (‘Minimum read count (total)’ field). Proceed with ‘Begin Analysis’.

Graphical user interface, text, application, chat or text message

Description automatically generated

1. When the python analysis is complete, mergez.txt and namelist.txt output files are generated. Remove all ‘chr’ prefixes from the mergez.txt file (i.e., change ‘chr17’ to ‘17’ for all coordinates) and run the ‘Matlab\_commands\_3\_16\_GenomADupdate.m’ script in Matlab.
2. Open the Matlab analysis output files ( ‘snvdata.csv’ and ‘formatfilterindels.csv’) in excel. Open ‘dataWithaa.xlsm’ and ‘indelmacro.xlsm’, and use macros (Developer -> Macros) for annotation and formatting of the output files.

*Note:* SNVs are not always annotated with the proper amino acid change. Therefore, we recommend use of variant effect predictor (VEP) for hg19 (online tool: <https://grch37.ensembl.org/Homo_sapiens/Tools/VEP> ) for SNV annotation. Use the following VEP parameters:

* Transcript database to use: ‘RefSeq transcripts’
* Restrict results: ‘Show one selected consequence per variant’