Tempus Bioinformatics Technical Challenge

Thanks for providing this Technical Challenge. All the codes are in the folder of “Tempus”. The final annotated file is in the folder of “finalResult”.

Each variant must be annotated with the following pieces of information:

1. Type of variation (substitution, insertion, CNV, etc.) and their effect (missense, silent,

intergenic, etc.). If there are multiple effects, annotate with the most deleterious

possibility.

***RE:*** *Type of variation is in the column of “VariantType”. Type of SNV is in the column of “ExonicFunc.refGene”. The effect of variation is the column of “Func.refGene”. The amino acid change is the column of “AAChange.refGene”. The functional effects of the variants were predicted by different algorithms and provided from column of “CLINSIG” to column of “SiPhy\_29way\_logOdds\_rankscore”. In my group, we are frequently using “CADD” and “DANN” to prioritize the variants.*

2. Depth of sequence coverage at the site of variation.

***RE:*** *Depth of sequence coverage at the site of variation is in the column of “DP\_locus.normal” for normal sample, and “DP\_locus.vaf5” for vaf5 sample.*

3. Number of reads supporting the variant.

***RE:*** *Number of reads supporting the variant is in the column of “DP\_alt.normal” for normal sample, and “DP\_alt.vaf5” for vaf5 sample.*

4. Percentage of reads supporting the variant versus those supporting reference reads.

***RE:*** *VAF(variant allele fraction) is used to show the percentage of reads supporting the variant. It is in the column of “VAF.normal” for normal sample, and “VAF.vaf5” for vaf5 sample.*

5. Allele frequency of variant from ExAC API (API documentation is available here:

<http://exac.hms.harvard.edu/>).

***RE:*** *Population variant allele frequency from ExAC with different ethnicities is from the column of “ExAC\_ALL” to the column of “ExAC\_SAS”. gnomAD database has larger population than ExAC, and my group frequently use gnomAD. For your reference, I also annotate the population variant allele frequency with the results provided from the column of “AF” to column of “controls\_AF\_popmax”.*

6. Any additional annotations that you feel might be relevant.

***RE:*** *The functional effects of the variants should be important to prioritize the variants. As described in question 1, I have provided this information from column of “CLINSIG” to column of “SiPhy\_29way\_logOdds\_rankscore”. The information of the variants in COSMIC data base is also important, especially for cancer biology. I provided this information in the column of “cosmic70”. Please also note that “TempusAnnotated\_short.txt” in the folder of “finalResult” is the file with no variant detected by freeBays in either “normal” or “vaf5” sample (the file was generated by CheckGT.R).*