VAN ANDEL RESEARCH INSTITUTE

We frequently encounter situations where we need a custom solution for a particular investigator's research project. Either the tool doesn't exist or an existing tool doesn't quite do what we need*. Thus, we'd like to see how you approach a coding problem.

Problem: An investigator has a VCF that they would like annotated in a specific way.

Coding challenge: We will provide you with the VCF file from the investigator, and we would like you to create a tool to output a table annotating each variant in the file. Every variant (if multiallelic, decompose into individual mutations) should be annotated with the following information derived from the VCF and querying the ExAC database via the API (documentation can be found at http://exac.hms.harvard.edu/):

1.Variant type (e.g. insertion, deletion, etc.). 2.Variant effect (e.g. missense, synonymous, etc.).

Note: If multiple variant types exist in the ExAC database, annotate with the most deleterious possibility.

3. Read depth at the site of variation. 4. Number of reads supporting the variant.
5.Percentage of reads supporting the variant versus those supporting reference reads.

6. Allele frequency of variant 7. (Optional) Any other information from ExAC that you feel might be relevant.

For this challenge please upload all relevant code (written in your preferred language) along with the annotated VCF file to a GitHub account and provide a link to ian.beddows@vai.org and marie.adams@vai.org. Work will be assessed based on quality of code and documentation to a greater degree than the variant annotation.

*Please note, we are aware that there are existing packages to do this type of work - please resist the urge to use one of them, and instead create your own code to accomplish the task - we want to see what you can do!