Dear Candidate,

Thank you very much for your application to the Bioinformatics position in the Tertiary Analysis Operations team. We do consider your application and would appreciate seeing your bioinformatics skills put to work.

The exercise is as following:

Incorporate in an SQL database (or other data structure you are familiar with) the dutch Allele Frequency Database GoNL (http://www.nlgenome.nl/).

- Get the SNV and indels data in VCF (it is sufficient to limit the analysis to one chromosome, such as chromosome 22 for example).
- Conceive an SQL database schema, which will permit you to store the variant data, such as genomic position, nucleotide changes, alternative allele count, total number of alleles, variant frequency, a unique variant ID (identifier) and the dbSNP reference (rsid).
- The data scheme should be optimised for queries based on genomic intervals (e.g. find all of the variants present in chromosome 22 between genomic positions n and m).
- Write a program in Python (or another programming language you are more comfortable with) to populate this database from the VCF file from GoNL (SNVs and indels). Exclude variants "Inaccessible".

Please return a structured report along with scripts and materials needed to reproduce your analysis.

If possible can you return the results to us in one week.

Please do not hesitate to contact us if you have and further queries or require more information.