

## **P53: GEVACT : GENOMIC VARIANT CLASSIFIER TOOL**

Dorien Daneels<sup>1</sup>, Isel Grau<sup>2</sup>, Dipankar Sengupta<sup>1</sup>, Maryse Bonduelle<sup>1</sup>, Dewan M. Farid<sup>2</sup>, Didier Croes<sup>1</sup>, Ann Nowé<sup>2</sup> & Sonia Van Dooren<sup>1</sup>

<sup>1</sup> Centre for Medical Genetics, Reproduction and Genetics, Reproduction Genetics and Regenerative Medicine, Vrije Universiteit Brussel, UZ Brussel

<sup>2</sup> Como - Artificial Intelligence Lab, Vrije Universiteit Brussel

### **INTRODUCTION**

With the emergence of new screening techniques, targeted, whole exome and genome screening are becoming standard diagnostic norms in clinical settings to identify the variants causative of a genetic disease. However, development of bioinformatics solutions for pathogenic classification of the variants still remains a big challenge and henceforth, making the process ponderous for geneticists and clinicians. In this work, we describe GEVACT (Genomic Variant Classifier Tool), a tool for classification of genomic single nucleotide and short insertion/deletion variants. The aim of this study was to design and implement a variant classification algorithm, based on a literature review of cardiac arrhythmia syndromes and existing knowledge of clinical geneticists.

### **METHODS**

The algorithm we propose for GEVACT is based on a published variant classification schema for cardiac arrhythmia syndromes (Hofman et al., 2013). It proposes two varying approaches: one to classify missense variants and another to classify nonsense and frameshift variants. The algorithm is implemented in two phases: pre-processing and classification.

In the pre-processing phase, an annotated tab-delimited variant file (.vcf.ann) retrieved from Alamut-batch (Interactive BioSoftware) can be refined based on the gene list for the disease-of-interest, so as to reduce the number of variants for the analysis. Filters are applied to look for variants that have already been reported in the Human Genome Mutation Database (Stenson et al., 2003) and in ClinVar (Landrum et al., 2014), or that have previously been detected and classified in an internal patient population. And lastly, the variants are filtered based on their location in the genome and their coding effect, followed by the check for minor allele frequency of the variant in a control population (Sherry ST et al. 2001).

Thereafter, in the classification phase, the filtered variants are classified as missense or nonsense/frameshift variants. For missense variants the classification is based on the parameters: amino acid substitution and its impact on protein function (Adzhubei et al., 2010; Kumar et al., 2009), biochemical variation (Mathe et al., 2006), conservation (Pollard et al., 2010), frequency of variant alleles in a control population (ESP6500), effects on splicing (Desmet et al., 2009), family and phenotype information and functional analysis. Whereas, for the nonsense and frameshift variants, it is based on: effects on splicing, frequency of variant alleles in a control population, family and phenotype information and functional analysis. For each parameter, a score is given to the variant, which is subsequently cumulated. Conclusively, based on the cumulative score each variant is classified into one of the five categories: Class I - Non-Pathogenic; Class II - VUS1 (unlikely pathogenic); Class III - VUS2 (unclear); Class IV - VUS3 (likely pathogenic); Class V - Pathogenic (Sharon et al., 2008).

In this study, we report a Java based tool called GEVACT, developed for classification of genomic variants. Input for the tool is an annotated vcf file, while the output depicts the cumulative classification score along with the class label for a variant. The tool was tested on a dataset of 130 cardiac arrhythmia syndrome patients, available at UZ Brussel. The results of the variant classification made by the tool were validated by manual curation, performed by the clinical geneticist. Definitively, the study indicates the tool to be promising but needs to be further validated on datasets from other diseases. In addition to, we are working on the tool to be adaptable for file inputs from other annotation software.

#### REFERENCES

- Adzhubei IA et al. *Nat Methods* 7(4), 248-249 (2010).  
Desmet et al. *Nucleic Acids Res* 37 (9): e67 (2009).  
Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA  
Hofman N et al. *Circulation* 128(14),1513-21 (2013).  
Kumar P et al. *Nat Protoc* 4(7), 1073–1081 (2009).  
Landrum MJ et al. *Nucleic Acids Res* 42(1), D980-5 (2014).  
Mathe E et al. *Nucleic Acids Res* 34(5),1317-25 (2006).  
Pollard K et al. *Genome Res* 20, 110-121 (2010).  
Sharon EP et al. *Hum Mutat.* 29(11), 1282–1291 (2008).  
Sherry ST et al. *Nucleic Acids Res* 29(1),308-11 (2001).  
Schulze-Bahr E et al. *Z Kardiol* 89 Suppl 4:IV12-22 (2000).  
Stenson et al. *Hum Mutat.* 21:577-581 (2003).  
Wilde AA & Tan HL *Circ J* 71, Suppl A:A12-9 (2007).