

# Overview of Single Nucleotide Variants

Louis Catherine\*

Department of Molecular Medicine, Central Research Institute, USA

## Description

Massive volumes of genomic variant information are produced by whole-exome sequencing (WES) and genome-wide association studies (GWAS) and figuring out which changes cause disease or influence phenotypic features is a significant issue. Since exonic non-synonymous single nucleotide variations (nsSNVs) account for the majority of known disease-causing mutations, most research concentrate on how these nsSNVs alter protein function. The impact of nsSNVs on protein function is demonstrated by computational studies to reflect sequence homology and structural information, and to be predicted by statistical methods, machine learning techniques, or models of protein evolution. One of the biggest problems in genetics is figuring out how genotype and phenotype are related. Between two unrelated humans, there are more than four million DNA variations. Approximately 80% of these variances are single nucleotide variations because additions and deletions often have a greater impact and are selected against more frequently (SNVs). An estimated 81 to 93 percent of human genes have at least one SNV across the whole population. Non-synonymous single nucleotide variations (nsSNVs), which make up a relatively tiny percentage of variants, account for about 85% of known illness relationships and are detected in about 10,000 pairs of unrelated individuals. The high-yield category of non-synonymous coding SNVs has historically received the most attention in techniques for forecasting the effects of SNVs. Interest in other types of mutations has increased due to the presence of disease-associated synonymous mutations and non-coding variations that affect lincRNA, miRNA, and promoters. However, different tools will be needed to analyze these types of variations, and these tools are still relatively new and untested. Not every nsSNV, though, affects how proteins function. Some variations might not alter the protein at all, in which case the mutation might not be harmful.

Some of the earliest techniques for anticipating the effects of nsSNVs relied only on structure. They attempted to determine the free energy change of folding as a result of a mutation as they hypothesized that harmful nsSNVs disturb protein folding. This idea is supported by the fact that roughly 75 percent of amino acid mutations that cause Mendelian disorders also influence protein stability. In general, homology-based approaches assume that the underrepresented changes in a protein family are harmful and the overrepresented ones have a neutral effect on the function of the protein. This implies two theories: that all homologs have the same function and that each substitution has an independent impact on protein function (no epistasis) (the fitness landscape is constant). Most approaches try to minimize this issue by optimizing the sequence selection to primarily orthologous proteins, hence avoiding changes in the fitness landscape. However, the prediction accuracy is greatly impacted by the failure of these hypotheses. Although non-native alignments might occasionally increase a method's accuracy, rationally tailoring the sequence alignment requires a considerable deal of skill and experience. In more recent homology implementations, substitution matrices and homology information were coupled. The alignment-based score that Provean employs

gauges the difference between the query sequence's sequence similarity to each of its homologs both before and after the introduction of the mutation. The BLOSUM62 matrix is used to measure similarity, and it can forecast the effects of multiple insertions, deletions, and substitutions of amino acids [1-3].

As an alternative, the Evolutionary Action method uses an equation that states that the impact of a mutation is a function of the functional significance of the altered residue and of the amino acid similarity of the substitution to simulate the genotype-to-phenotype link. The Evolutionary Trace technique approximates the functional importance, and substitution matrices that depend on the functional significance of the residues and, optionally, on their structural characteristics, approximate the similarity of the amino acids. Overall, the prevalence of such techniques demonstrates homology's independence from other variables and precision in predicting the impact of nsSNVs. Regardless of whether it is used alone or in conjunction with structural information, homology has consistently been a key factor in nsSNV effect prediction, although its predictive power has several drawbacks. Lower prediction accuracy may be caused, in example, by the absence of available homologous sequences. For instance, the Provean technique typically employs 100–200 homologous sequences, but accuracy suffers when this number falls below 50. About 20,000 proteins that could be used as therapeutic targets were initially revealed by the Human Genome Project (HGP), which has deciphered the complete sequence of nucleotide base pairs that make up the human genome. Unexpectedly, hundreds of thousands of non-coding RNAs that were previously thought of as "junk DNA" have been found thanks to later large-scale annotation projects like the Encyclopedia of DNA Elements (ENCODE) project. Long non-coding RNAs (lncRNAs), which are among them, are extensively transcribed in mammalian genomes. LncRNAs can be categorized into five main groups according to their locations and properties: intergenic, antisense, sense, intronic, and overlapping. About half of these lncRNAs are transcribed from locations within 10 kb of protein-coding loci, making them more likely to play a role in the cis-regulatory control of the expression levels of nearby genes. Transcripts that are located further away from nearby genes appear to have less chance of cis-regulatory control [4,5].

## Conflict of Interest

None.

## References

- Gerstung, Moritz, Christian Beisel and Niko Beerenwinkel. "Reliable detection of subclonal single-nucleotide variants in tumour cell populations." *Nature communi* 3 (2012): 1-8.
- Goya, Rodrigo, Mark G.F. Sun and Janine Senz. "SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors." *Bioinfo* 26 (2010): 730-736.
- Capriotti, Emidio, Russ B. Altman and Yana Bromberg. "Collective judgment predicts disease-associated single nucleotide variants." *BMC Genomics* (2013): 1-9.
- Gehring, Julian S., Bernd Fischer, Michael Lawrence and Wolfgang Huber. "Somatic Signatures: inferring mutational signatures from single-nucleotide variants." *Bioinfo* 31 (2015): 3673-3675.
- Bühler, Kora-Mareen, Elena Giné and Jose Antonio López-Moreno. "Common single nucleotide variants underlying drug addiction: More than a decade of research." *Addict Biol* 20 (2015): 845-871.

\*Address for Correspondence: Louis Catherine, Department of Molecular Medicine, Central Research Institute, USA, E-mail: cathlouis981@gmail.com

Copyright: © 2022 Catherine L. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Date of Submission: 05-May-2022, Manuscript No: jmgm-22-69481; Editor assigned: 07-May-2022, PreQC No. P-69481; Reviewed: 12-May-2022, QC No. Q-69481; Revised: 18-May-2022, Manuscript No. R-69481; Published: 24-May-2022, DOI: 10.37421/1747-0862.2022.16.556

How to cite this article: Catherine, Louis. "Overview of Single Nucleotide Variants." *J Mol Genet Med* 16 (2022): 556.