

An introduction to the predictors of pathogenic point mutations

Author : Muniba Faiza

Categories : [Algorithms](#), [Softwares](#), [Tools](#)

Date : September 7, 2017



Single nucleotide variation is a change in a single nucleotide in a sequence irrespective of the frequency of the variation. Single nucleotide variants (SNVs) play a very important role in causing several diseases such as the tumor, cancer, etc. Many efforts have been made to identify the SNVs which were initially based on identifying non-synonymous mutations in coding regions of the genomes.

Nowadays, the classifiers developed to identify SNVs are more focused on making genome-wide predictions [1,2]. Combined Annotation–Dependent Depletion (CADD) developed by Kircher et al., 2014, is one of the methods which predict the pathogenic point mutations, which integrates many diverse annotations into a single C-score for each variant. The C-score correlate with the annotations of functionality, allelic diversity, disease severity, pathogenicity, complex trait associations, and experimentally measured regulatory effects [1]. However, the performance of CADD was challenged later [3]. DANN is a deep learning approach to annotate pathogenicity of genetic variants [4]. It is based on the deep neural networks, which overcome the issue of the methods based on support vector machines

which are unable to capture nonlinear relationships among the features, thereby, limiting the performance [4].

Recently, a new method has been introduced by Vander Velde et al., (2017) [5], which adjusts the C-score in a gene-specific manner, known as Gavin (Gene-Aware Variant INterpretation) [5]. It outperforms the CADD method of pathogenic point mutation identification and assigns the Benign and Pathogenic labels to simplify the interpretation. Another method has also been introduced recently, known as FATHMM-XF (Functional Annotation Through Hidden Markov Model- eXtended Feature) [5]. It uses Platt scaling [7] to assign a confidence score, i.e., p-score to each prediction and focus the analysis on a subset of high-confidence predictions. The SNVs in FATHMM-XF have been characterized into feature groups of 27 data sets from ENCODE (The ENCODE Project Consortium, 2012) and NIH Roadmap Epigenomics [8], which were significantly applicable in some other domains such as Genome tolerance browser (GTB) is an online browser to visualize the predicted tolerance of a genome to mutation [2,9]. Four other feature groups have been developed from the conservation score, annotated gene models, the Variant Effect Predictor [10], and the sequence itself.

References

1. Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*, 46(3), 310-315.
2. Shihab, H. A., Rogers, M. F., Gough, J., Mort, M., Cooper, D. N., Day, I. N., ... & Campbell, C. (2015). An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics*, 31(10), 1536-1543.
3. Liu, X. et al. (2016). The performance of deleteriousness prediction scores for rare non-protein-changing single nucleotide variants in human genes. *J. Medical Genetics*.
4. Quang, D., Chen, Y., & Xie, X. (2014). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, 31(5), 761-763.
5. van der Velde, K. J., de Boer, E. N., van Diemen, C. C., Sikkema-Raddatz, B., Abbott, K. M., Knopperts, A., ... & Sinke, R. J. (2017). GAVIN: Gene-Aware Variant INterpretation for medical sequencing. *Genome biology*, 18(1), 6.
6. Mark F. Rogers, Hashem A. Shihab, Matthew Mort, David N. Cooper, Tom R. Gaunt, Colin Campbell; FATHMM-XF: accurate prediction of pathogenic point mutations via extended features, *Bioinformatics*, btx536
7. Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3), 61-74.
8. Bernstein, B. E. et al. (2010). The NIH roadmap epigenomics mapping consortium. *Nat.*

Biotechnology, 28(10), 1045–1048.

9. Shihab, H. A., Rogers, M. F., Ferlino, M., Campbell, C., & Gaunt, T. R. (2017). GTB—an online genome tolerance browser. *BMC bioinformatics*, 18(1), 20.

10. McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. (2016). The ensembl variant effect predictor. *Genome biology*, 17(1), 122.

Sharing is caring. Spread the love!

- [Print](#)
- [Email](#)
- [LinkedIn](#)
- [Twitter](#)
- [Facebook](#)
- [Google](#)
-