IMHOTEP 

Integrating **M**olecular **H**euristics and **O**ther **T**ools for **E**ffect **P**rediction

Combined Effect Pathogenicity Prediction for Human Non-synonymous Variants

Aim and method

This tool combines the predictions of nine third party prediction tools in an integrative approach. The individual prediction tools utilized are:

PhyloP (1), Grantham Score (2), PolyPhen-2 (HumVar model) (3), SNPs&GO (4), MutPred (5), SIFT (6), MutationTaster2 (7), Mutation Assessor (8) and FATHMM (inherited disease model) (9).

Number of prediction tools utilized: smart selection

Our pipeline offers the user the opportunity to choose between predictions based upon all nine prediction tools or only on the five most relevant tools. These latter tools are termed the 'smart selection' and comprise

PolyPhen-2, SNPs&GO(4)(4), MutPred(5)(5), MutationTaster2 and FATHMM.

Our study (10) showed that predictions made with the five tools of the smart selection are nearly identical to those with the application of all nine individual tools.

Input

The input of this integration method can either be manually entered or uploaded from a file. It consists of the continuous scores of the nine prediction tools mentioned above. For MutationTaster2, in addition to the continuous score, the binary decision (effect yes or no) is required. Since these individual prediction tools are the property of their respective authors and institutions, they cannot be provided on this web page. The respective scores have to be calculated by the users themselves.

Calculation of the scores:

- PhyloP: e.g. by the UCSC Genome Browser, dbNSFP, ANNOVAR
- Grantham Score: The Grantham Score can be calculated directly on the basis of Grantham's original data (http://www.genome.jp/dbget-bin/www_bget?aax2:GRAR740104) or e.g. by using ANNOVAR.
- PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/ (the required score corresponds to pph2_prob)
- SNPs&GO: http://snps-and-go.biocomp.unibo.it/snps-and-go or http://snps.biofold.org/snps-and-go/index.html
- MutPred: http://mutpred.mutdb.org/
- SIFT: http://sift.bii.a-star.edu.sg
- MutationTaster2: http://www.mutationtaster.org/ Here, in addition to the continuous score the binary decision is required (effect yes or no).
- Mutation Assessor: http://mutationassessor.org/r3/
- FATHMM: http://fathmm.biocompute.org.uk/inherited.html

Manual entry

For manual entry, just enter the nine continuous scores into the corresponding fields. For MutationTaster2, the binary decision of the tools (effect yes or no) is also required.

Upload a file

The input file should have either 11 (all nine tools) or 7 (smart selection of tools) tab separated columns. These columns are

For all 9 tools

```
1) Id
2) Score of PhyloP
3) Score of GranthamScore
4) Score of PolyPhen-2
5) Score of SNPs&GO
6) Score of MutPred
7) Score of SIFT
8) Score of MutationTaster2
9) Binary decision of MutationTaster2:
   1 – effect (consequential); 0 – no effect (inconsequential)
10)  Score of Mutation Assessor
11)  Score of FATHMM
```

For the 5 tools of the smart selection

```
1) Id
2) Score of PolyPhen-2
3) Score of SNPs&GO
4) Score of MutPred
5) Score of MutationTaster2
6) Binary decision of MutationTaster2:
   1 – effect (consequential); 0 – no effect (inconsequential)
7) Score of FATHMM
```

Lines beginning with "#" will be ignored (e.g. a header). By the number of columns, the prediction will automatically be performed for all 9 tools or only the 5 tools of the smart selection.

An example input file can be downloaded (http://www.uni-kiel.de/medinfo/cgi-bin/predictor/example_input.txt).

Output

All binary decisions are given as inconsequential (no effect, non-functional, neutral, not pathogenic) or consequential (effect, functional, damaging, pathogenic)

- Scores of the individual tools as given in the input
  The binary decision of the individual tools and the normalized score is also given.
- Binary decision of our integration methods, **random forest, decision tree, logistic regression and binary summation**
  The best performing integration method of our study was random forest, closely followed by decision tree and logistic regression. Binary summation is given here because of its simplicity and easy interpretation but it performed worse than the other three integration methods.
- Scores and thresholds for random forest, logistic regression and binary summation
  For decision tree, no score is available since it yields only a binary classification.
- If values are missing or inappropriate a comment is given.

An example output file can be downloaded (http://www.uni-kiel.de/medinfo/cgi-bin/predictor/example_output.txt).

Publication

For details on the statistical models and the development of the integration tools see our publication (10).

References

1. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110-121.
2. Grantham, R. (1974) Amino-acid difference formula to help explain protein evolution. *Science*, **185**, 862-864.
3. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248-249.
4. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P.L. and Casadio, R. (2009) Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.*, **30**, 1237-1244.
5. Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D. and Radivojac, P. (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, **25**, 2744-2750.
6. Ng, P.C. and Henikoff, S. (2001) Predicting deleterious amino acid substitutions. *Genome Res.*, **11**, 863-874.
7. Schwarz, J.M., Cooper, D.N., Schuelke, M. and Seelow, D. (2014) MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods*, **11**, 361-362.
8. Reva, B., Antipin, Y. and Sander, C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.*, **39**, e118.
9. Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N. and Gaunt, T.R. (2013) Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.*, **34**, 57-65.
10. Knecht, C., Mort, M., Junge, O., Cooper, D.N., Krawczak, M. and Caliebe, A. (2016) A composite score integrating popular tools for predicting the functional consequences of non-synonymous sequence variants. *submitted to Nucleic Acids Research*.