## Five Functional Prediction Algorithms for NGS Data

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Pub Year</th>
<th>Citations</th>
<th>Host Institution</th>
<th>Category</th>
<th>Distinguishing Characteristic</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>2003</td>
<td>&gt;1200</td>
<td>JCVI (UW)</td>
<td>Untrained</td>
<td>Popular, broadly applicable and intuitive method to identify functional mutations.</td>
<td>sift.jcvi.org</td>
</tr>
<tr>
<td>PolyPhen-2</td>
<td>2010</td>
<td>&gt;1000</td>
<td>Harvard/BWH</td>
<td>Trained</td>
<td>Provides two scores (HumDiv and HumVar) for applications to complex and Mendelian disease, respectively.</td>
<td>genetics.bwh.harvard.edu/pph2</td>
</tr>
<tr>
<td>MutationAssessor</td>
<td>2011</td>
<td>57</td>
<td>MSKCC</td>
<td>Untrained</td>
<td>Considers amino acid conservation in protein subfamilies to refine important functional regions. Interactive user interface.</td>
<td>mutationassessor.org</td>
</tr>
<tr>
<td>MutationTaster</td>
<td>2010</td>
<td>199</td>
<td>Charite-Berlin</td>
<td>Trained</td>
<td>Native support for DNA (rather than amino acid) variant analysis. Allows online submission of VCF files.</td>
<td>mutationtaster.org</td>
</tr>
<tr>
<td>FATHMM</td>
<td>2013</td>
<td>NA</td>
<td>University of Bristol</td>
<td>Trained (weighted)</td>
<td>Uses HMM method (rather than BLAST) to create MSA. Weighted extensions for human disease and cancer analysis.</td>
<td>fathmm.biocompute.org.uk</td>
</tr>
</tbody>
</table>

**Cautions on Usage**

- The common belief is that variants called damaging by multiple algorithms are most likely to have true disease causing potential. (Although this is not always true.)
- Published comparisons aren’t exhaustive, and usually focus on prediction performance for detecting a particular category of mutations.
- Each prediction tool has its own strengths and weaknesses and may carry certain biases based on the authors’ own research interests.
- All of the algorithms generally perform well for distinguishing between known damaging variants and known neutral variants.
- False positive rate can be high when the methods are applied to a broad range of variants of unknown significance.