Secretion pathway prediction
ProteinCenter (Proxeon Bioinformatics, Odense, Denmark; http://www.cbs.dtu.dk/services)

**SignalP (version 4.1)**
(http://www.cbs.dtu.dk/services/SignalP)

Classical secretion pathway 예상

hidden Markov model algorithms 이용

SignalP uses amino acid sequences to predict the existence and location of signal peptide cleavage sites. The hidden Markov model algorithms calculate the probability of whether the submitted sequence contains a signal peptide or not. A protein is considered classically secreted if it receives a default cutoff values Y (0.5 > D-cutoff for TM networks, 0.45 > D-cutoff for no-TM model networks).
Parameter

1. Fasta 형식 또는 파일 준비
   - 한 번의 submissive에 2,000 entries, 총 200,000 amino acids가능
   - 하나의 entry는 6,000 amino acids을 넘지 않아야 함 (넘을 경우 N-terminal에서부터 6000번째까지만 사용)
2. Organism group 선택
3. Output format: data 정리 및 분석 작업을 위해서는 “short” 선택

4. D-cutoff value: default 권장 (0.45이상 SignalP-noTM network, 0.5이상 SignalP-TM network)
   *TM: transmembrane state (sequence의 hydrophobicity에 따라 integral transmembrane sequence 예상, 4개 이상 존재하면 TM으로 인지)
5. Method: default 권장 (input sequences may include TM regions)
6. Graphics output: 다수의 proteome 분석 시 default 권장 (no graphic)
Fig 4. Graphics output “PNG (inline)” 선택 시, output

■ 결과 해석

SignalP 4.1 Server - prediction results
Technical University of Denmark

Fig 5. SignalP의 prediction results

D (D-cutoff)
? (Y: D-cutoff 0.45 ≥ SignalP-noTM, 0.5 ≥ SignalP-TM, N: 0.45 < SignalP-noTM, 0.5 < SignalP-TM)
SecretomeP (version 2.0)

- (http://www.cbs.dtu.dk/services/SecretomeP)
- nonclassical secretion pathway 예상
- neural network 이용

SecretomeP uses a neural network that combines six protein characteristics to determine if a protein is nonclassically secreted. These characteristics include the number of atoms, number of positively charged residues, presence of transmembrane helices, presence of low-complexity regions, presence of pro-peptides and subcellular localization. A protein is considered nonclassically secreted if it receives an NN-score $\geq 0.5$.\textsuperscript{iii}

---

Fig 6. SecretomeP homepage
■ Parameter

1. Fasta 형식 또는 파일 준비
   - 한 번의 submission에 100 entries, 총 200,000 amino acids 가능
   - 하나의 entry는 40~4,000 amino acids여야 함. (넘을 경우 N-terminal에서부터 4,000번째 까지만 사용)

2. Organism group 선택

■ 결과 해석

![Secretome 2.0 Server - prediction results](image)

Non-classically secreted proteins should obtain an NN-score exceeding the normal

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>NN-score</th>
<th>Odds</th>
<th>Weighted</th>
<th>Warning</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>ADAM17_HUMAN</td>
<td>0.190</td>
<td>0.372</td>
<td>0.001</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>02</td>
<td>ARVR_HUMAN</td>
<td>0.532</td>
<td>1.103</td>
<td>0.002</td>
<td>-</td>
</tr>
<tr>
<td>03</td>
<td>ASNS_HUMAN</td>
<td>0.427</td>
<td>0.730</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td>04</td>
<td>ATRN_HUMAN</td>
<td>0.230</td>
<td>0.560</td>
<td>0.001</td>
<td>-</td>
</tr>
<tr>
<td>05</td>
<td>CATE_HUMAN</td>
<td>0.625</td>
<td>1.717</td>
<td>0.003</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>06</td>
<td>CBPE_HUMAN</td>
<td>0.464</td>
<td>0.933</td>
<td>0.002</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>07</td>
<td>CLCA_HUMAN</td>
<td>0.464</td>
<td>4.778</td>
<td>0.010</td>
<td>-</td>
</tr>
<tr>
<td>08</td>
<td>CLH2_HUMAN</td>
<td>0.443</td>
<td>0.868</td>
<td>0.002</td>
<td>-</td>
</tr>
<tr>
<td>09</td>
<td>C5A31_HUMAN</td>
<td>0.053</td>
<td>0.129</td>
<td>0.000</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>10</td>
<td>CD78_HUMAN</td>
<td>0.003</td>
<td>1.909</td>
<td>0.004</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>11</td>
<td>CRL1_HUMAN</td>
<td>0.521</td>
<td>1.140</td>
<td>0.002</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>12</td>
<td>ERFL7_HUMAN</td>
<td>0.500</td>
<td>1.465</td>
<td>0.003</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>13</td>
<td>FSFRK_HUMAN</td>
<td>0.611</td>
<td>1.514</td>
<td>0.003</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>14</td>
<td>FVNC_HUMAN</td>
<td>0.371</td>
<td>0.686</td>
<td>0.001</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>15</td>
<td>FPRP_HUMAN</td>
<td>0.711</td>
<td>2.350</td>
<td>0.005</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>16</td>
<td>GFP5_HUMAN</td>
<td>0.692</td>
<td>5.521</td>
<td>0.011</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>17</td>
<td>GNR_HUMAN</td>
<td>0.590</td>
<td>1.342</td>
<td>0.003</td>
<td>signal peptide predicted by SignalP</td>
</tr>
<tr>
<td>18</td>
<td>H3K27_HUMAN</td>
<td>0.953</td>
<td>1.959</td>
<td>0.004</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>IEP59_HUMAN</td>
<td>0.819</td>
<td>3.877</td>
<td>0.008</td>
<td>signal peptide predicted by SignalP</td>
</tr>
</tbody>
</table>

Fig 7. SecretomeP의 prediction results

SignalP 에 의해 signal peptide가 예상된 단백질을 제외하고 NN-score가 0.5이상인 경우, nonclassical secretion pathway를 통해 secretion 되었다고 예상
TMHMM (version 2.0)

- (http://www.cbs.dtu.dk/services/TMHMM)
- transmembrane helices in integral membrane proteins 예상
- hidden Markov model 이용

The TMHMM program is based on a hidden Markov model approach. The hidden Markov model can incorporate hydrophobicity, charge bias, helix lengths and grammatical constraints into one model for which algorithms for parameter estimate and prediction already exist.

Fig 8. TMHMM homepage
Parameter

1. Fasta 형식 또는 파일 준비
   · 한 번의 submission에 4,000 entries 가능

2. Output format 선택: one line per protein 추천

결과해석

Fig 9. TMHMM의 prediction results (Extensive, with graphics)
**TMHMM result**

*HELP with output formats*

**Fig 10. TMHMM prediction results (one line per protein)**

- **Length ("len="):** the length of the protein sequence.
- **Number of predicted TMHs:** The number of predicted transmembrane helices.
- **Exp number of AAs in TMHs ("ExpAA="):** The expected number of amino acids in transmembrane helices. If this number is larger than 18 it is very likely to be a transmembrane protein (OR have a signal peptide).
- **Exp number, first 60 AAs ("First60="):** The expected number of amino acids in transmembrane helices in the first 60 amino acids of the protein. If this number is more than a few, you should be warned that a predicted transmembrane helix in the N-term could be a signal peptide.
- **Total prob of N-in:** The total probability that the N-term is on the cytoplasmic side of the membrane.
- **POSSIBLE N-term signal sequence:** a warning that is produced when "Exp number, first 60 AAs" is larger than 10.
- **"PredHel=":** The number of predicted transmembrane helices by N-best.
- **"Topology=":** The topology predicted by N-best. ("i" if the loop is on the inside or "o" if it is on the outside)

