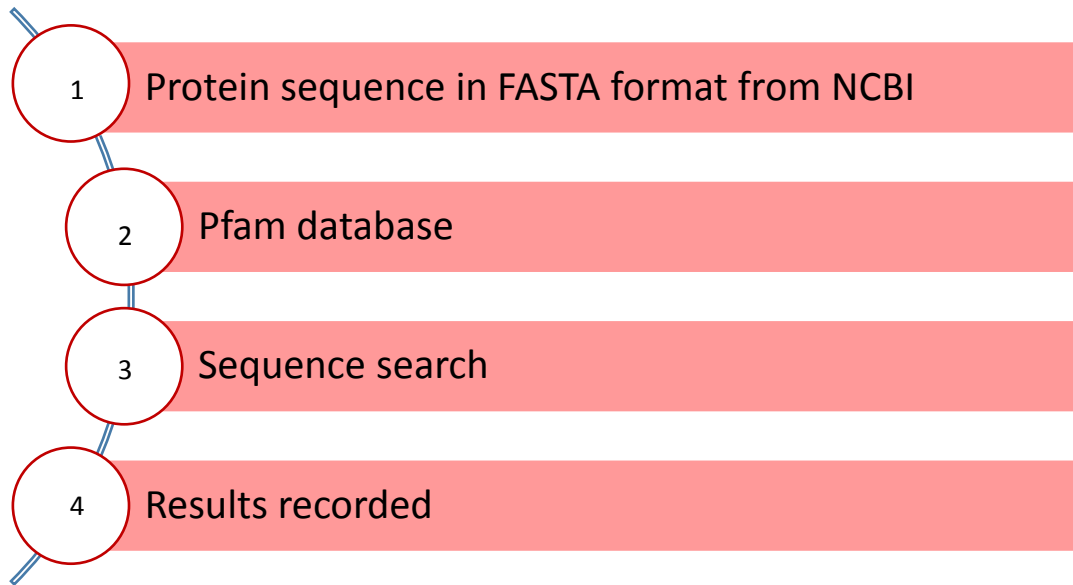


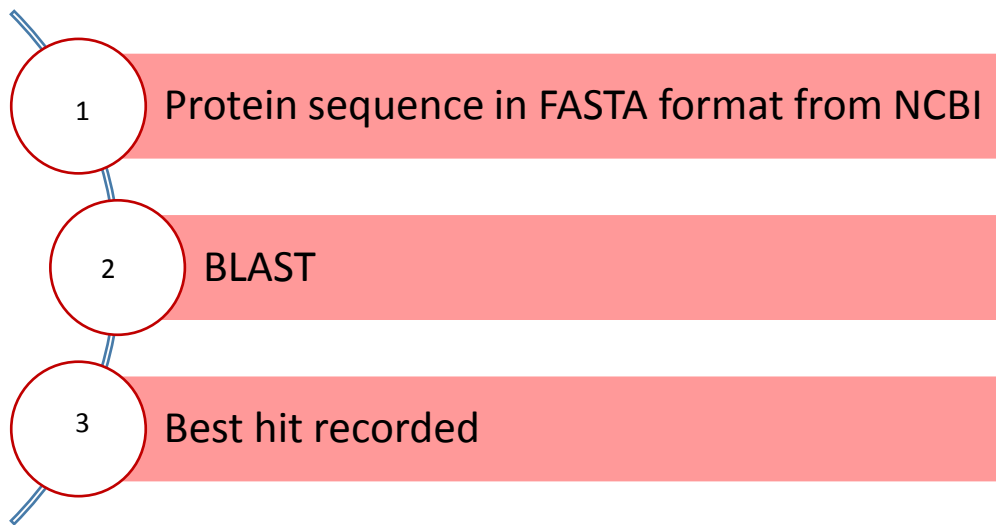
Putative gene: From Uniprot

Classifier 1: Pfam



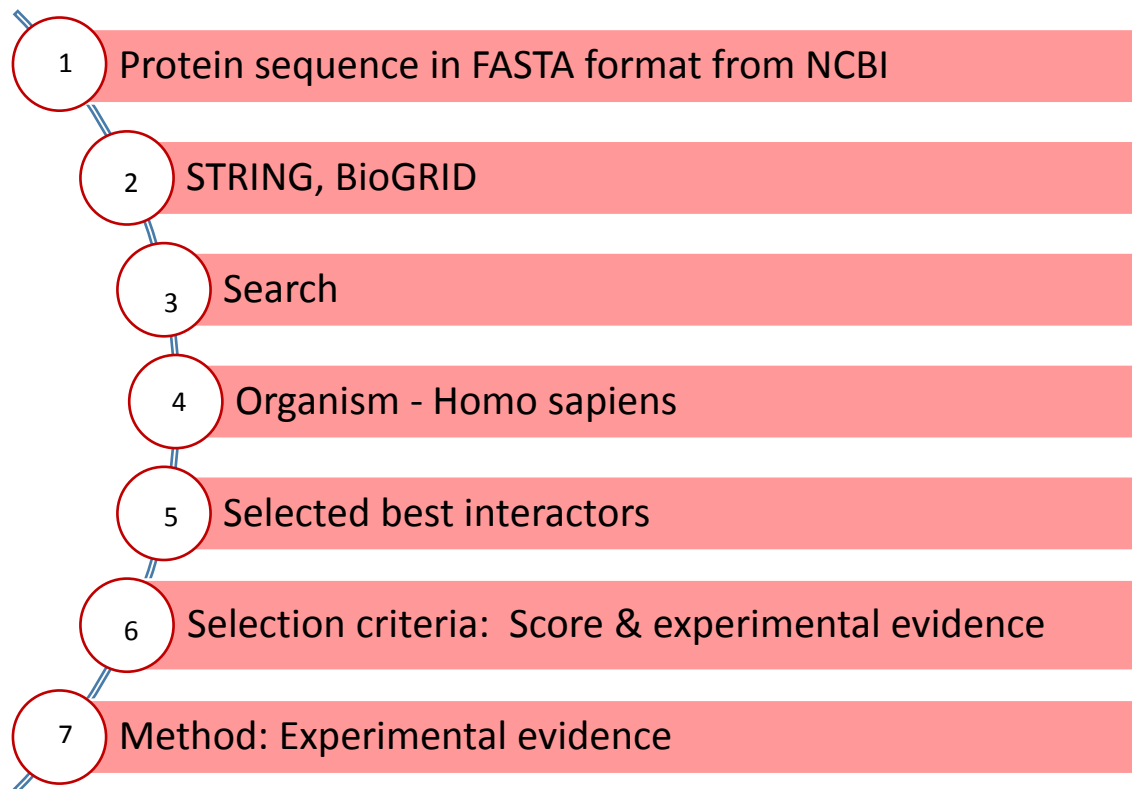
Scoring: Protein meet the classifier- 1, else 0. E-value: $<1 = 0$; >1 or near to 1 = 0

Classifier 2: Orthology



Scoring: E value: $<1 = 1$; $>1 = 0$

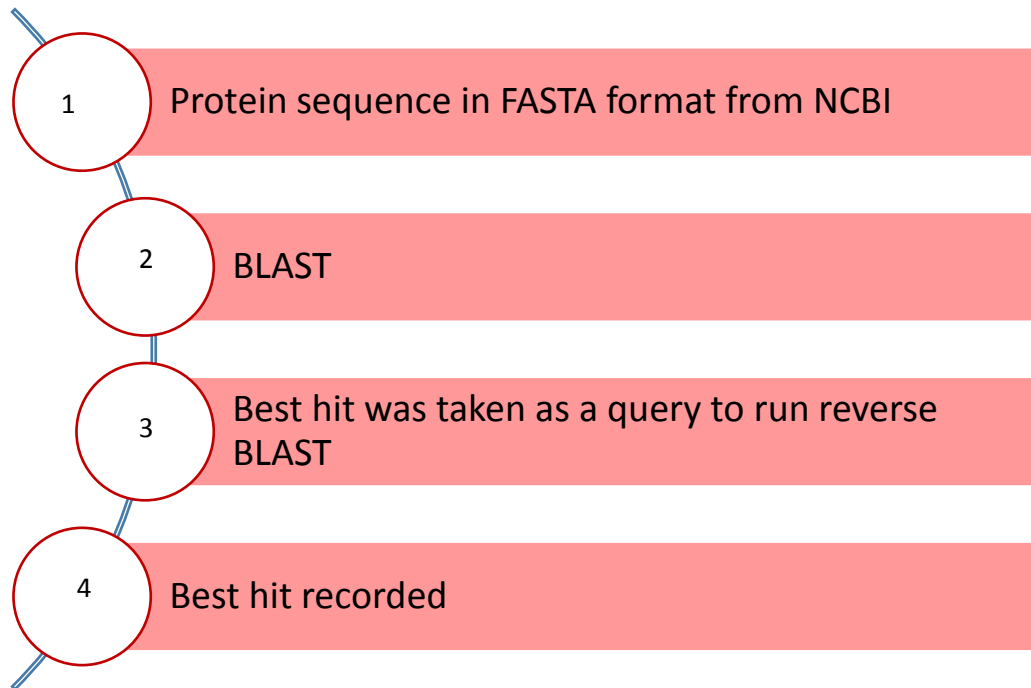
Classifier 3: Protein interactions/Association studies



Scoring: If I Could find same interactors in atleast two databases, score = 1, else 0.

*BioGrid and STRING are given priority

Classifier 4: Bidirectional best blast hits (BBH)

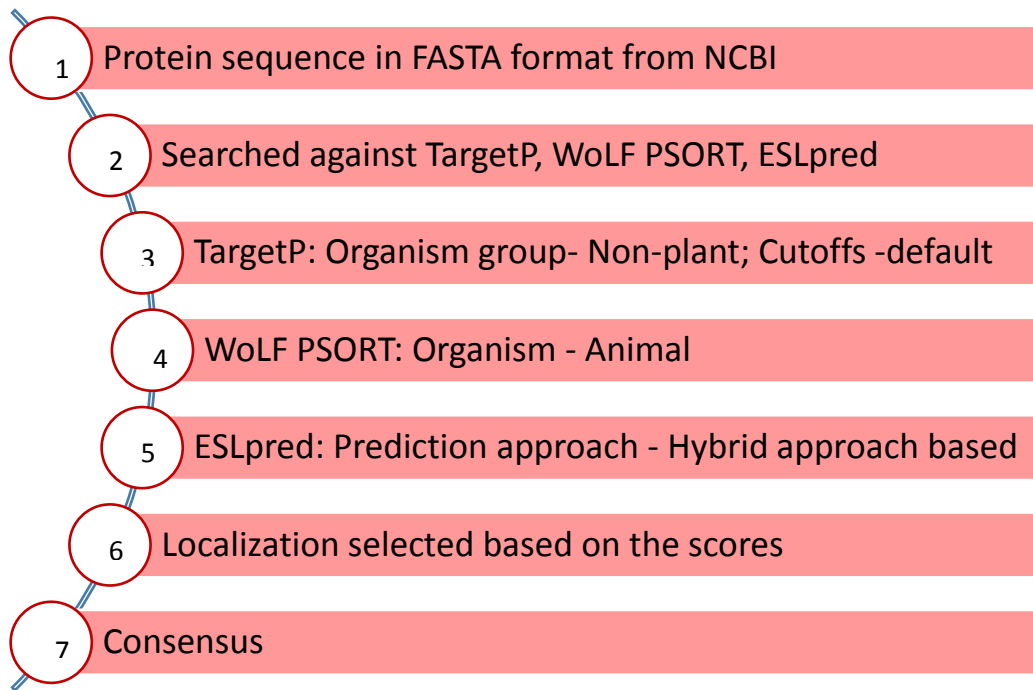


Scoring: If reverse blast of protein A resulted in protein A as a best hit (BBH), gave score as 1, else 0

Protein A - - - -BLAST- - - - Best hit - - - REVERSE BLAST - - - protein A - 1

Protein A - - - -BLAST- - - - Best hit - - - REVERSE BLAST - - - protein B - 0

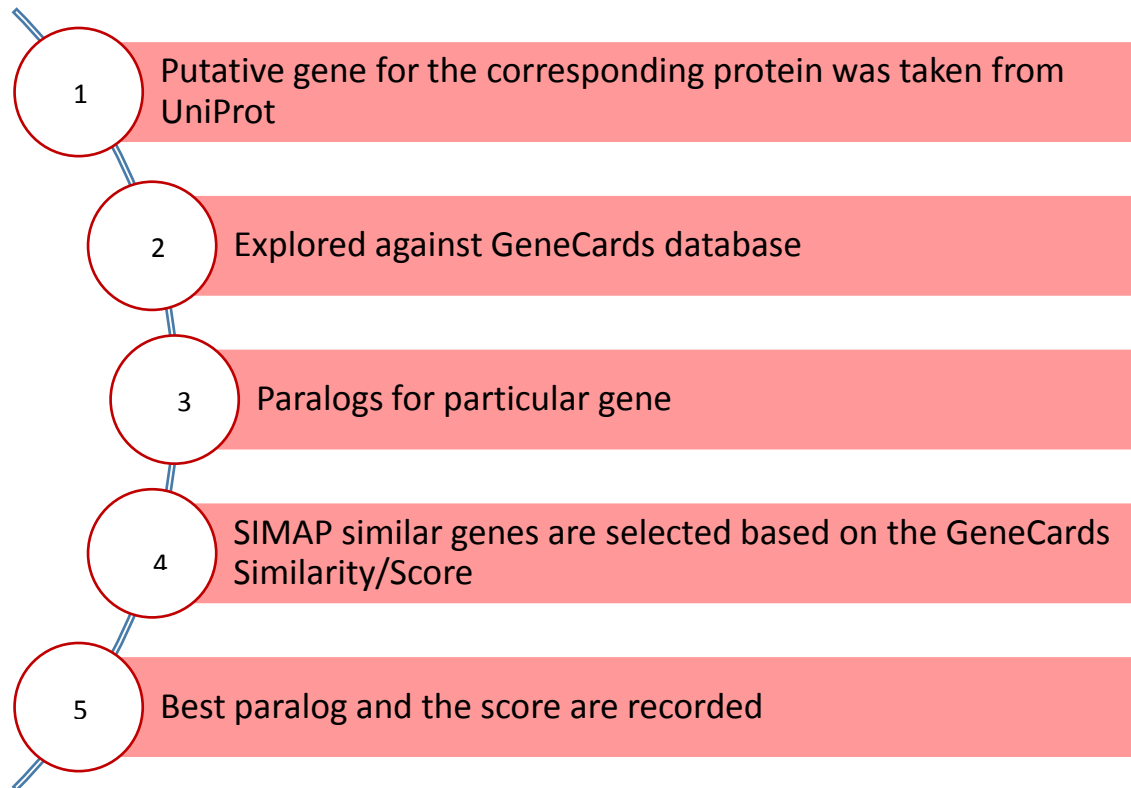
Classifier 5: Sorting signals



Scoring: Could come to consensus from three databases, score =1, else 0

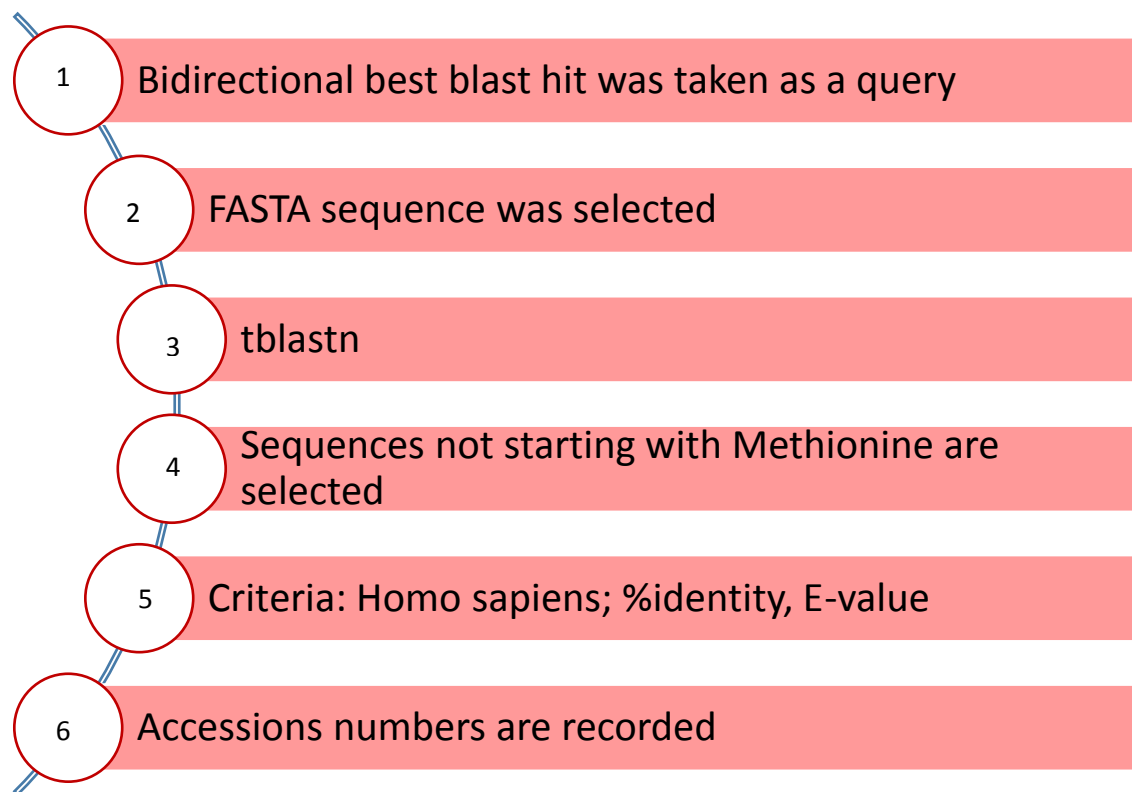
For ESLpred, order of priority: Nuclear > Cytoplasm > Mitochondria > Extracellular

Classification 6: Functional linkages from known databases



Scoring: If paralog was found, score = 1, if not 0

Classification 7: HPs linked to Pseudogenes



Scoring: 1. Predicted and synthetic sequences are ignored.

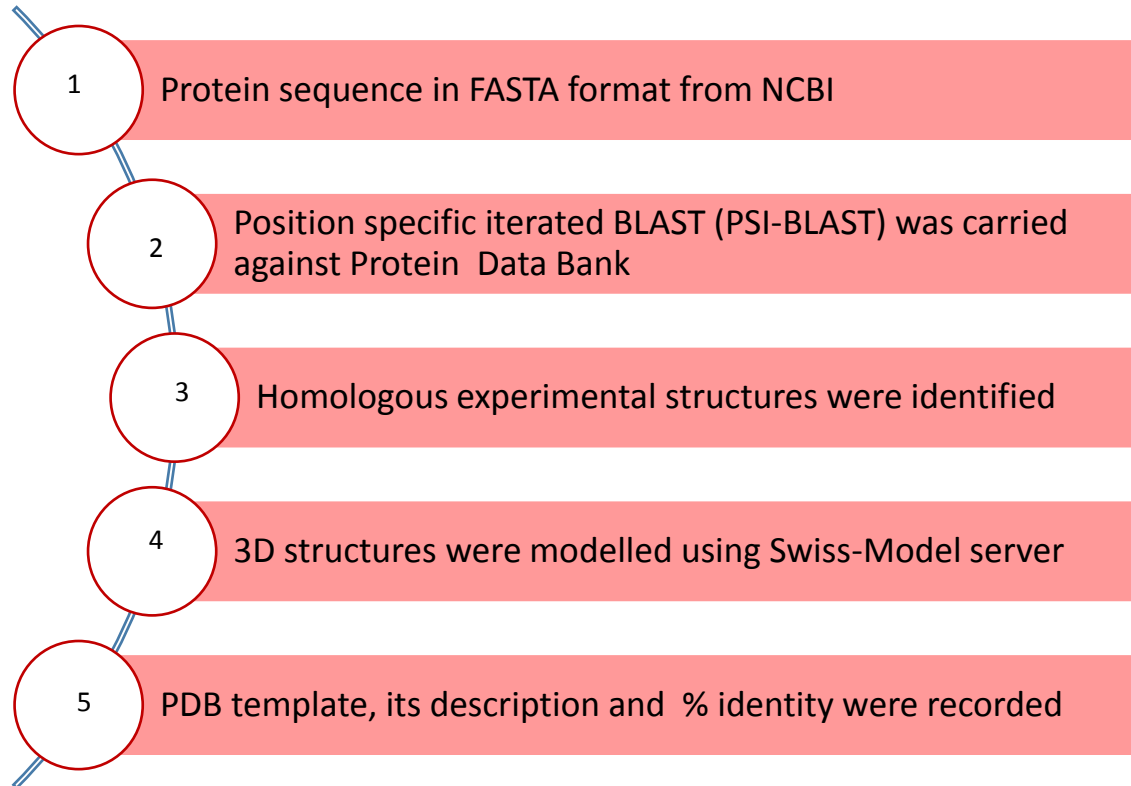
2. Sequences from Homo sapiens are considered.

3. E- value less than zero is considered

4. No end to end alignment

Sequences not starting with Methionine and meeting all the above criteria, given score 1, else 0.

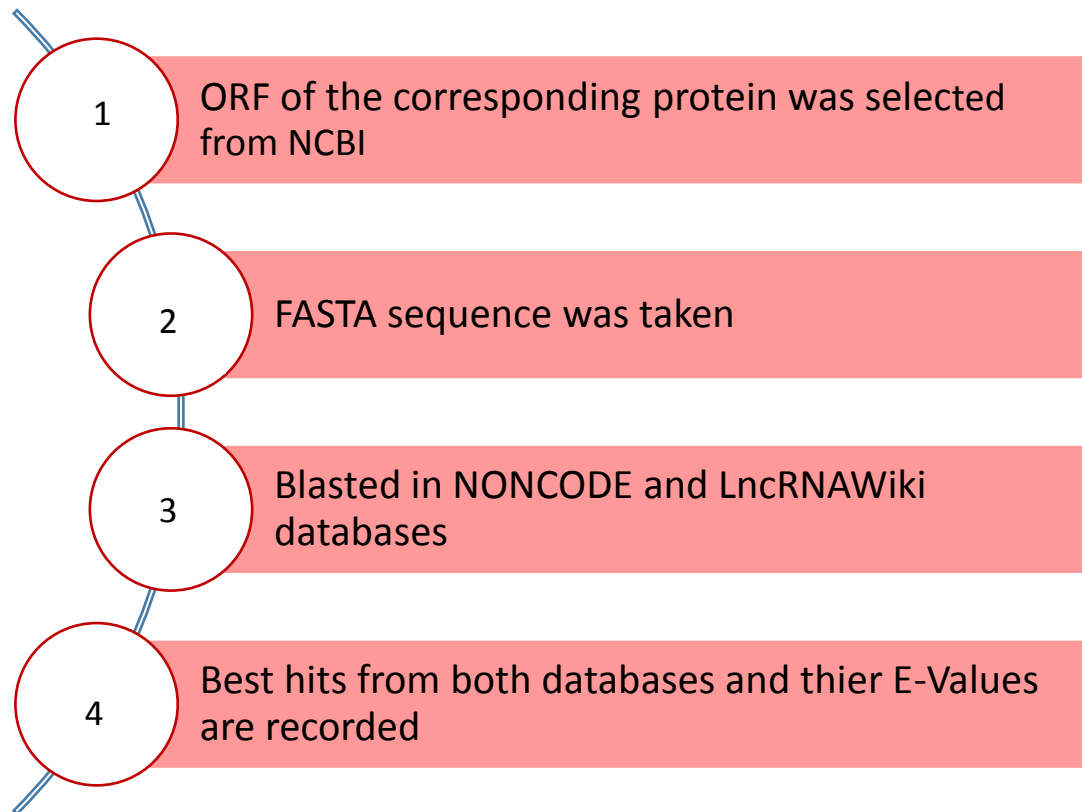
Classification 8: Homology modelling



Scoring: Based on % Identity between query and PDB template

If there is more than 30% similarity, score =1, else 0.

Classification 9: HPs linked to Non-coding RNAs



Scoring: Only sequences from Homo sapiens are considered (NONCODE)

Criteria for scoring

H. sapiens sequence in top three hits

Top five hits are considered when there is no considerable difference between scores of first five hits.

E value less than zero

If the above criteria are met, score 1, else 0.