NCBI resources

dbSNP: the NCBI database of genetic variation

The Single Nucleotide Polymorphism database (dbSNP) is a public-domain archive for a broad collection of simple genetic polymorphisms.

Type of variation
Single nucleotide substitutions
insertion/deletion
Invariant regions of sequence
Microsatellite repeats

ClinVar

aggregates information about genomic variation and its relationship to human health.

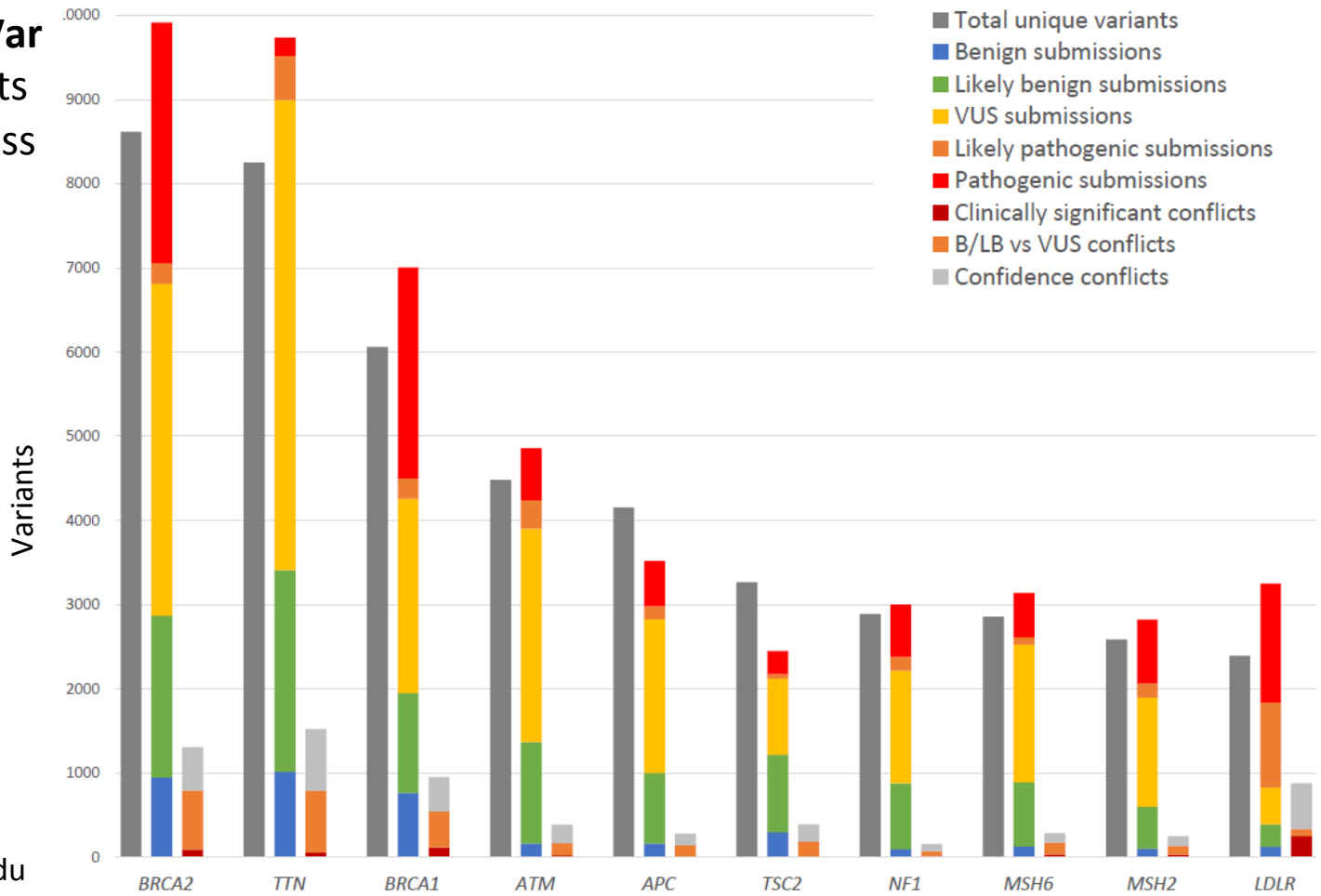
Data coming from many sources (clinical labs, researchers, databases, clinics, patient registries)

Jan 27th 2018:
580,861 submissions on 375,116 unique variants from 880 submitters from 63 countries

Accessions, of the format SCV000000000.0, are assigned to each submission.

Top 10 Genes in ClinVar

- Total unique variants
- Broken down by class
- Conflicts



Data retrieved Jan 2018
from ClinVar Miner
clinvarminer.genetics.utah.edu

NCBI resources videos

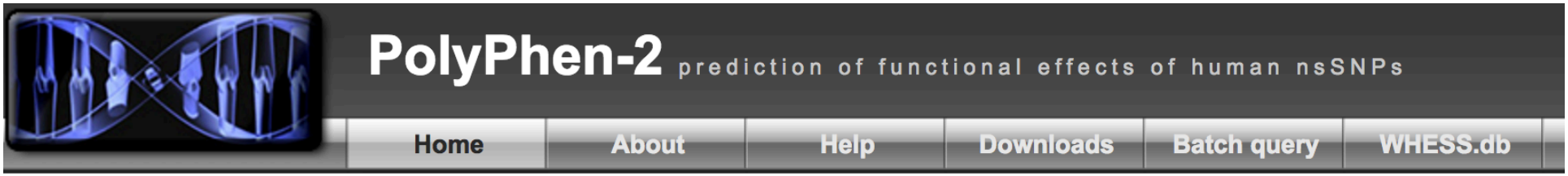
- Variation viewer:
  - https://www.youtube.com/watch?v=rnWZ9MFBwUM
- ClinVar
  - https://www.youtube.com/watch?v=A8G3ej83ZgU

Predicting functional effects

- What is the effect of a non-synonomous change on the protein function?
- the structure and function of the protein can give you information.
- If you don’t have structure data, there are tools that can make a prediction based on sequence.

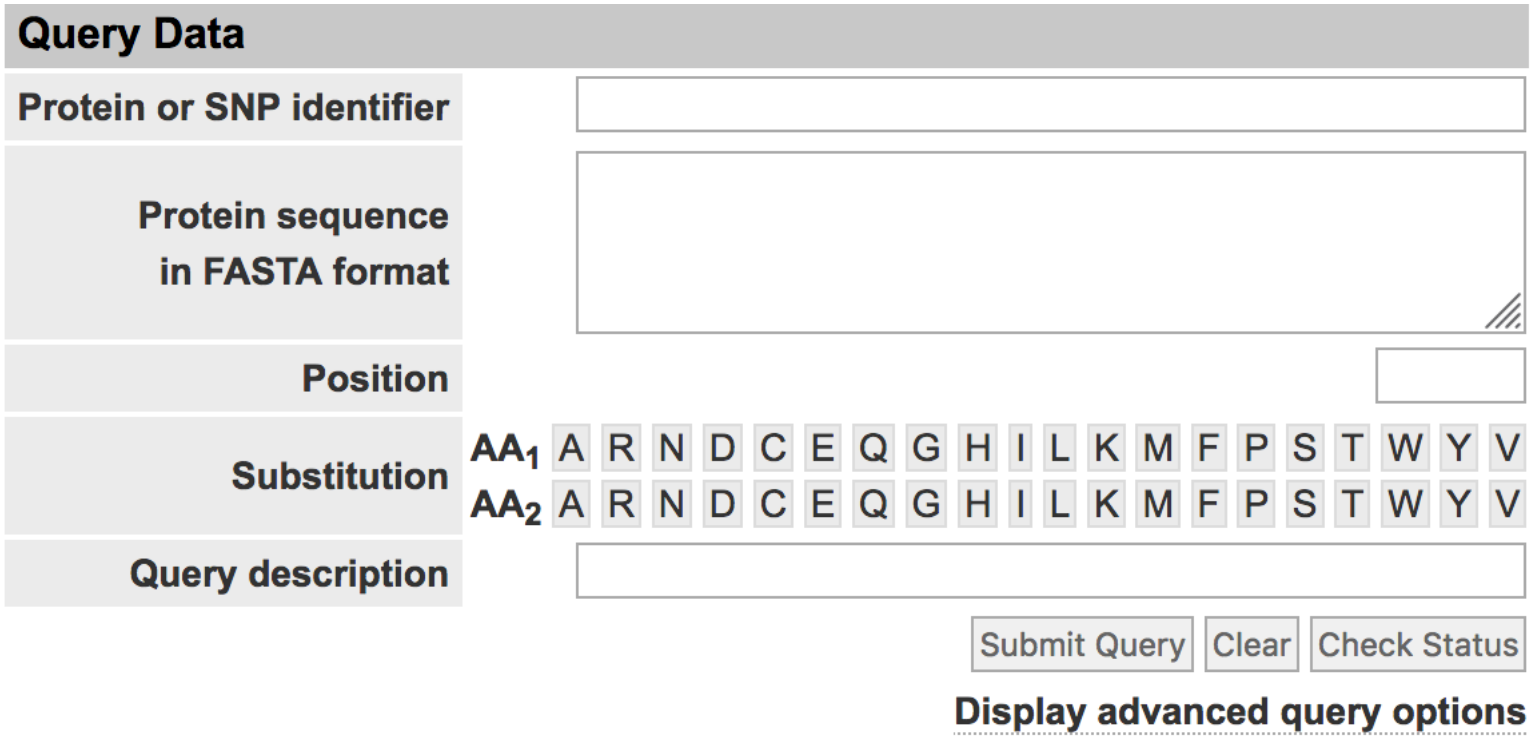
Two prediction tools

- PolyPhen 2.0

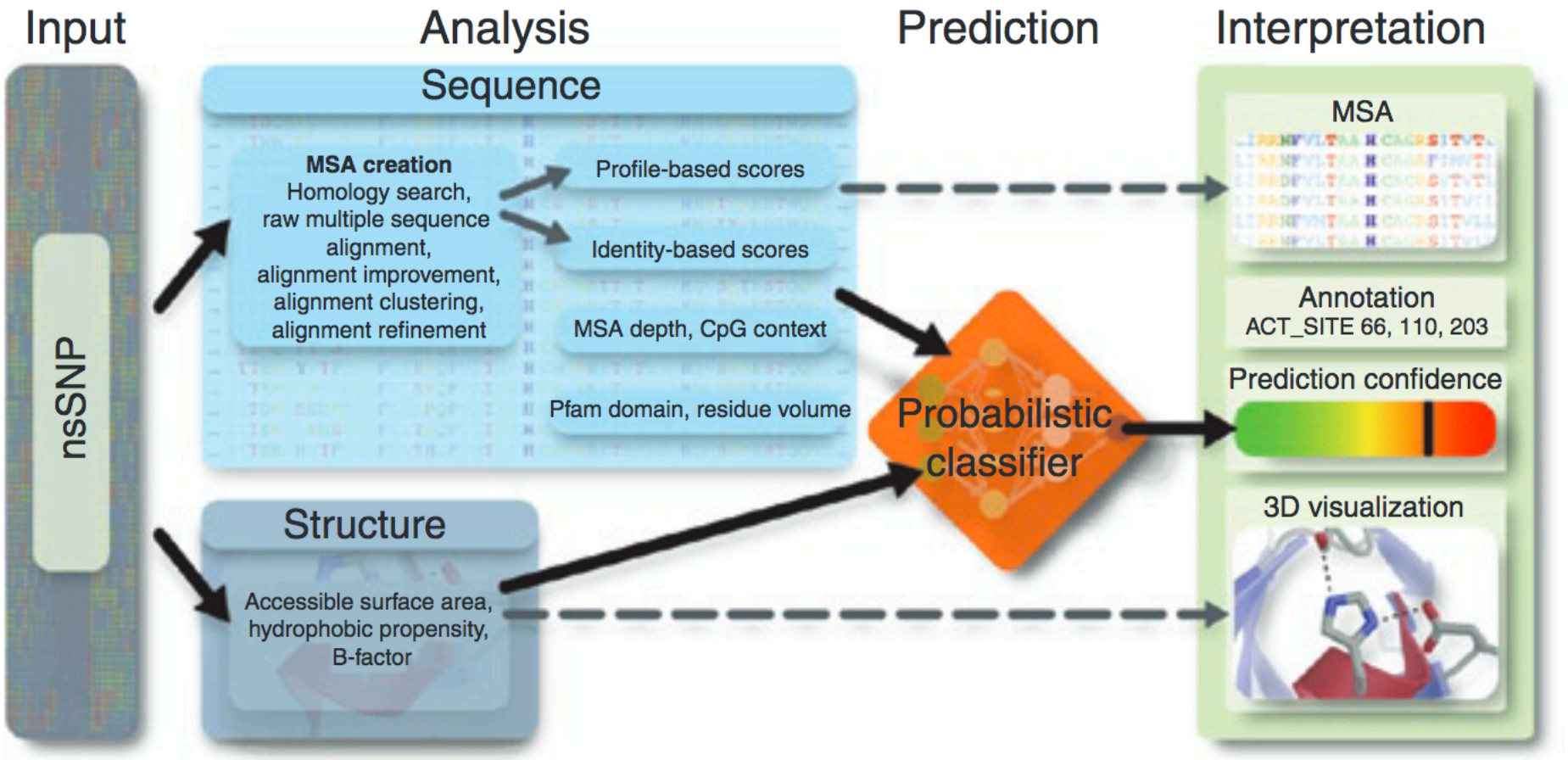


- SIFT

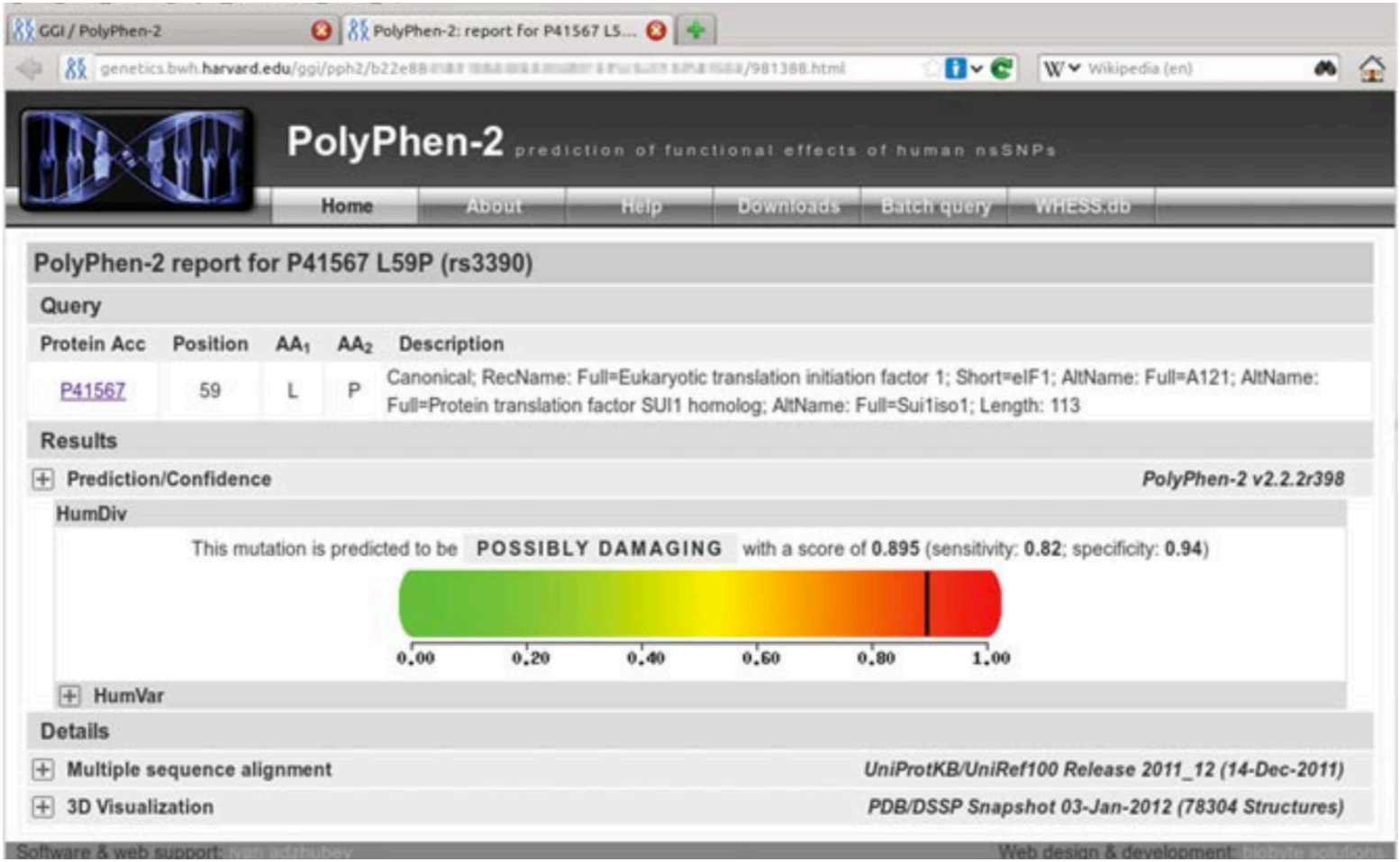
Polyphen 2


Query Data
Protein or SNP identifier
Protein sequence in FASTA format
Position
Substitution
AA1 A R N D C E Q G H I L K M F P S T W Y V
AA2 A R N D C E Q G H I L K M F P S T W Y V
Query description
Submit Query
Clear
Check Status
Display advanced query options

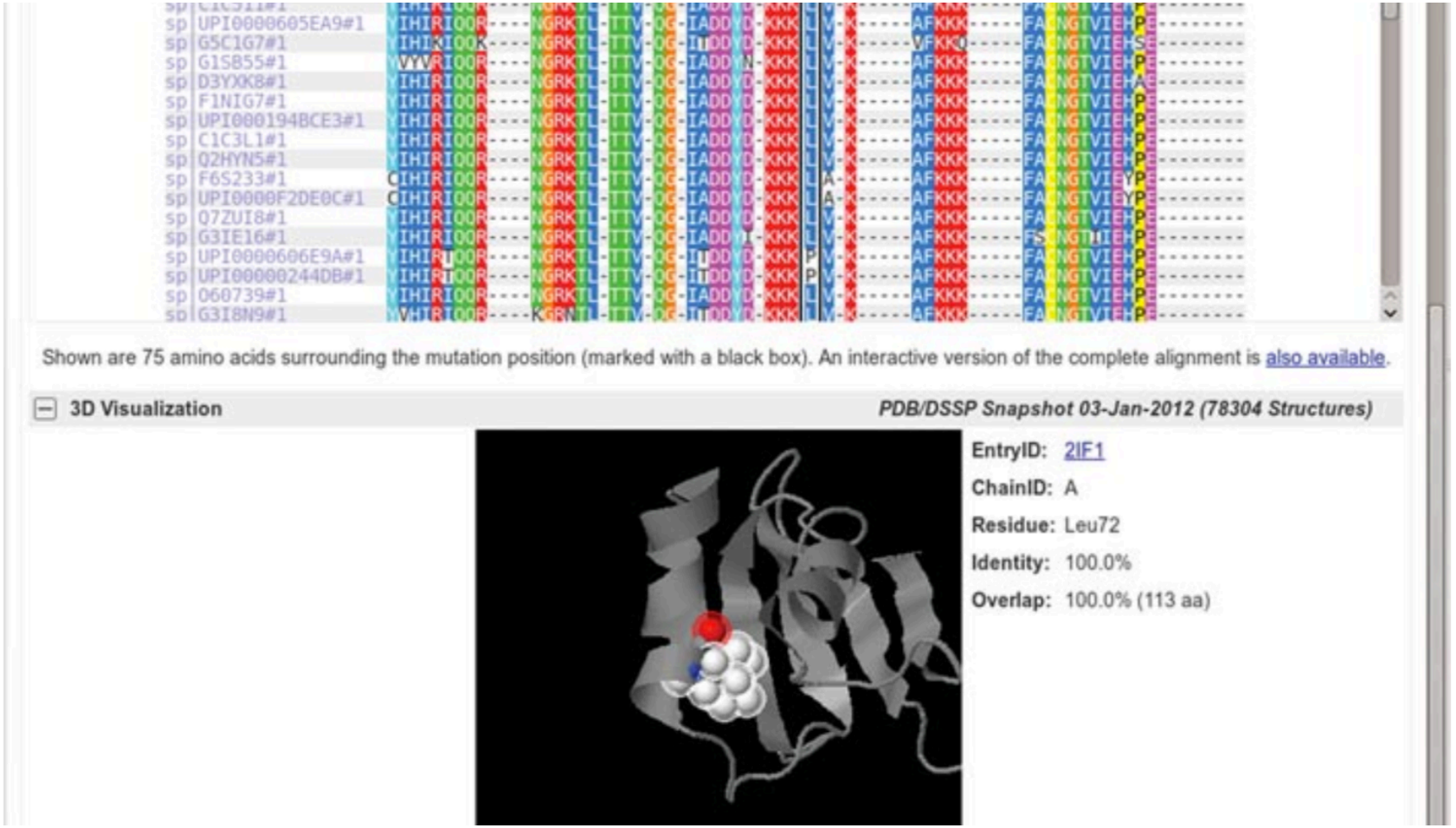
Polyphen 2


Input
Analysis
Prediction
Interpretation
nsSNP
Sequence
MSA creation
Homology search,
raw multiple sequence
alignment,
alignment improvement,
alignment clustering,
alignment refinement
Profile-based scores
Identity-based scores
MSA depth, CpG context
Pfam domain, residue volume
Structure
Accessible surface area,
hydrophobic propensity,
B-factor
Probabilistic
classifier
MSA
Annotation
ACT_SITE 66, 110, 203
Prediction confidence
3D visualization

Polyphen 2


GGI / PolyPhen-2
PolyPhen-2: report for P41567 L5...
genetics.bwh.harvard.edu/ggi/pph2/b22e88
/981388.html
Wikipedia (en)
PolyPhen-2 prediction of functional effects of human nsSNPs
Home
About
Help
Downloads
Batch query
WHESS.db
PolyPhen-2 report for P41567 L59P (rs3390)
Query
Protein Acc
Position
AA1
AA2
Description
P41567
59
L
P
Canonical; RecName: Full=Eukaryotic translation initiation factor 1; Short=eIF1; AltName: Full=A121; AltName: Full=Protein translation factor SUI1 homolog; AltName: Full=Sui1iso1; Length: 113
Results
Prediction/Confidence
PolyPhen-2 v2.2.2r398
HumDiv
This mutation is predicted to be POSSIBLY DAMAGING with a score of 0.895 (sensitivity: 0.82; specificity: 0.94)
0.00
0.20
0.40
0.60
0.80
1.00
HumVar
Details
Multiple sequence alignment
UniProtKB/UniRef100 Release 2011_12 (14-Dec-2011)
3D Visualization
PDB/DSSP Snapshot 03-Jan-2012 (78304 Structures)
Software & web support:
Web design & development:

Polyphen 2

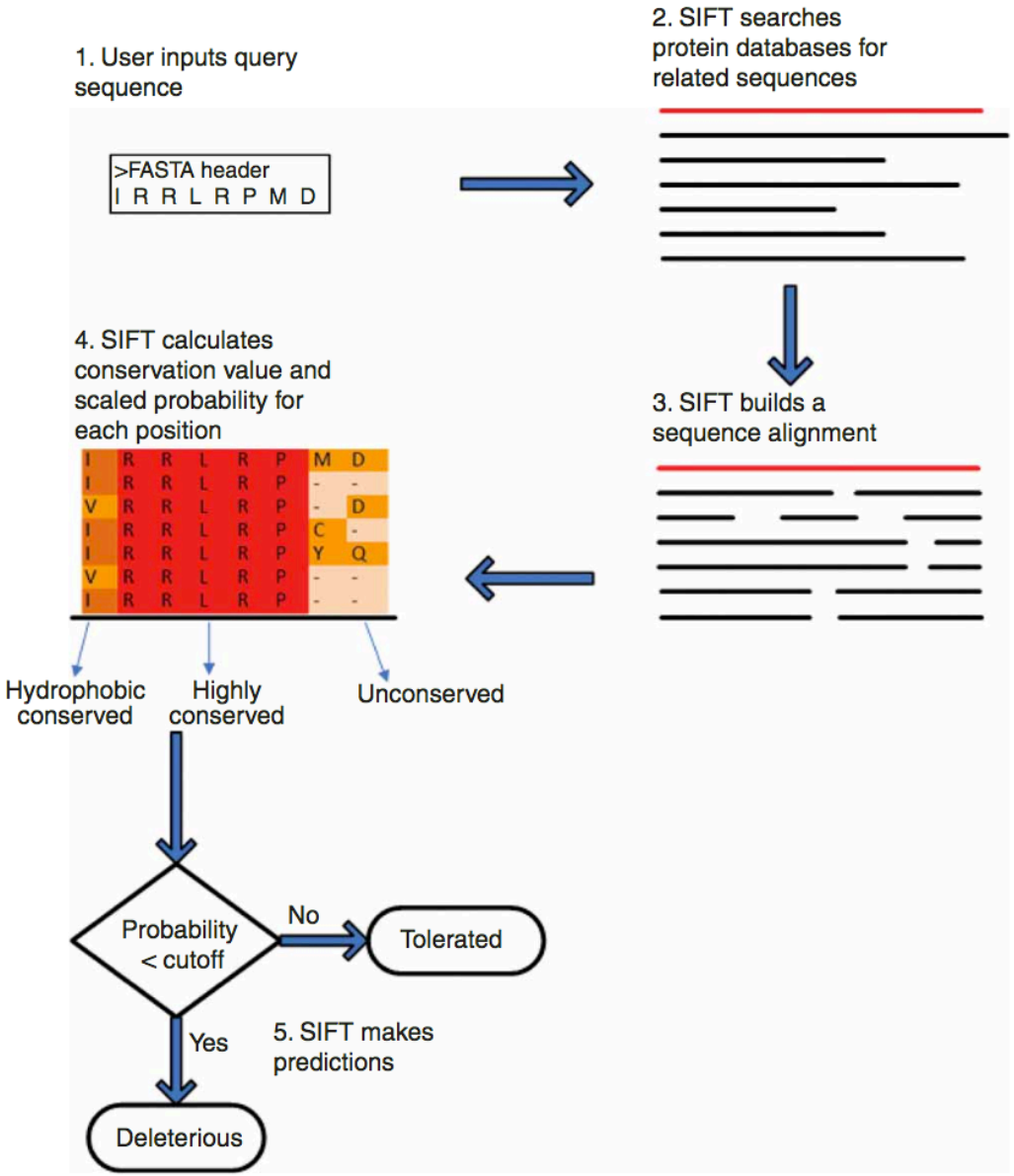

sp|UPI0000605EA9#1
sp|G5C1G7#1
sp|G1SB55#1
sp|D3YXK8#1
sp|F1NIG7#1
sp|UPI000194BCE3#1
sp|C1C3L1#1
sp|Q2HYN5#1
sp|F6S233#1
sp|UPI0000F2DE0C#1
sp|Q7ZUI8#1
sp|G3IE16#1
sp|UPI0000606E9A#1
sp|UPI00000244DB#1
sp|O60739#1
sp|G3I8N9#1
Shown are 75 amino acids surrounding the mutation position (marked with a black box). An interactive version of the complete alignment is also available.
3D Visualization
PDB/DSSP Snapshot 03-Jan-2012 (78304 Structures)
EntryID: 2IF1
ChainID: A
Residue: Leu72
Identity: 100.0%
Overlap: 100.0% (113 aa)

Polyphen 2

- Polyphen 2 can also work as a standalone software.
- Perl is required
- Can use the tool to annotate genomic SNP, to annotate protein variant or to classify probabilistic variant

SIFT

- Sorting intolerant from tolerant
- Uses sequence similarity + physicochemical properties of aa



UCSC

- UCSC (https://genome.ucsc.edu/) also has resources to annotate and interpret variants. This tool is called Variant Annotation Integrator (https://genome.ucsc.edu/cgi-bin/hgVai). The final output is similar to the Variant Effect Predictor from Ensembl.