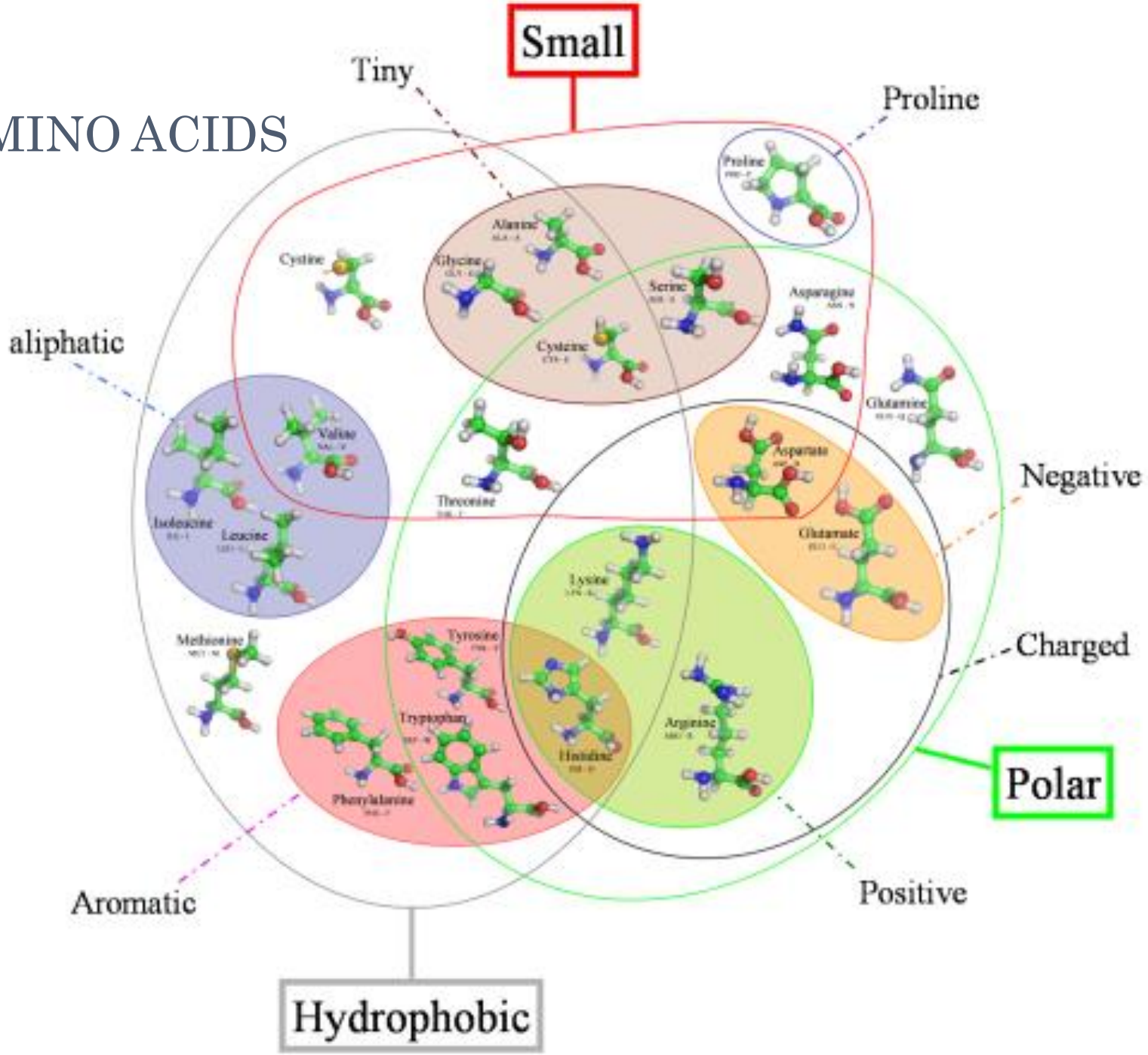
Sequence-based mutation analysis

Sonja Waldraff

25.6.2013

AMINO ACIDS


Small
Tiny
Proline
aliphatic
Negative
Charged
Polar
Positive
Aromatic
Hydrophobic
Proline
Alanine
Glycine
Serine
Cystine
Cysteine
Asparagine
Glutamine
Valine
Threonine
Isoleucine
Leucine
Aspartate
Glutamate
Lysine
Methionine
Tyrosine
Tryptophan
Histidine
Arginine
Phenylalanine

Prediction Tools

- Most common tools predict effect of single nucleotide polymorphisms (SNPs)

- SIFT
- PolyPhen2
- SNAP
- MutationTaster

SIFT - SORTING INTOLERANT FROM TOLERANT SUBSTITUTIONS

- Input:
  - FASTA sequence
  - SNP amino acid exchange, i.e. N61D (N is mutated to D at position 61)

- Method: "Multistep procedure"
  - Searches for similar sequences
  - Chooses closely related sequences that may share similar function →multiple alignment of these chosen sequences
  - Calculates normalized probabilities for all possible substitutions at each position from the alignment
  - Based on BLOSUM Matrix
  - Every substitution below a given cutoff is assumed to be deleterious

SIFT - SCORING

- **Score** ranges from 0 to 1:
  - Damaging with a score <= 0.05
  - Tolerated with a score > 0.05

- Score evaluation
  - **Median Info** ranges from 0 to 4.32 (=$\log_2$ 20)
    - Ideally between 2.75 and 3.25
    - To measure the diversity of the sequences used for prediction
    - Warning, if MI > 3.25 → indicates that the prediction was based on closely related sequences
  - **Seqs at Position**:
    - Number of sequences with an amino acid at the prediction-position
    - At beginning or end of a protein sometimes only few sequences are represented

SIFT - N61D

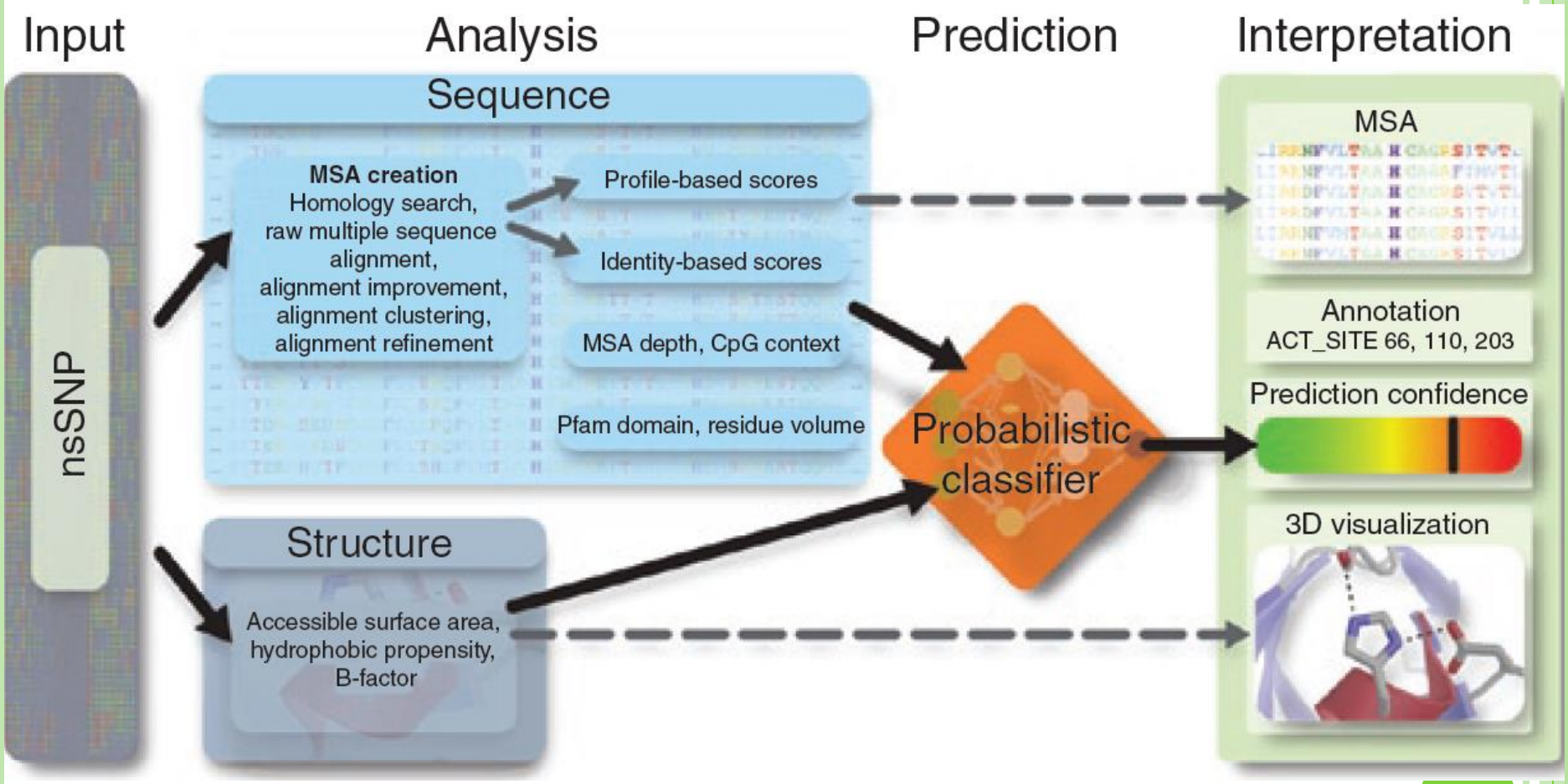
- Prediction
  - Substitution at pos 61 from N to D is predicted to AFFECT PROTEIN FUNCTION with a score of 0.03.
  - Median sequence conservation: 3.01
  - Sequences represented at this position:77

Polyphen2 – Polymorphism phenotyping version 2

- Input:
  - Protein ID, SNP ID or FASTA sequence
  - Needs position and substitution (AA_1 & AA_2)

- Better than SIFT and SNAP

POLYPHEN2 -COMBINATION OF SEQUENCE-BASED AND STRUCTURE-BASED FEATURES



POLYPHEN2

- Two trained datasets:
  - **HumDif**:
    - to evaluate rare alleles at loci potentially involved in complex phenotypes, dense mapping of regions found with GWAS and analysis of natural selection
    - "probably damaging": [0.957,1]
    - "possibly damaging“: [0.453,0.956]
    - "benign“: [0,0.452]
  - **HumVar**:
    - to distinguish mutations with drastic effect from other human variation, including mildly deleterious alleles (diagnostic of Mendelian diseases)
    - "probably damaging“: [0.909,1]
    - "possibly damaging“: [0.447,0.908]
    - "benign“: [0,0.446]

POLYPHEN2-EXAMPLE (N61D)

- HumDif:
  - This mutation is predicted to be benign with a score of **0.029** (sensitivity: **0.95**; specificity: **0.82**)
- HumVar:
  - This mutation is predicted to be benign with a score of **0.087** (sensitivity: **0.91**; specificity: **0.67**)
- Multiple sequence alignment and 3D visualizations are available

SNAP - Screening for non-acceptable polymorphisms

- neural network to predict SNP effects
- Uses different features like
  - psi-blast frequency profile
  - PSIC scores (position specific independent counts)
  - Structure prediction:
    - relative solvent accessibility prediction (PROFace)
    - secondary structure prediction (PROFsec)
  - predicted residue flexibility (PROFbval)
  - Pfam information, SwissProt annotations …
- Looks at the environment of the SNP (window of five amino acids)
- Better than SIFT and PolyPhen, especially on tough cases

SNAP

- Input: Fasta-Sequence
- Amino acid substitution and position like Sift: (AA_1PosAA_2)
- Output:
  - Divides between neutral and non-neutral
  - Reliability Index 0(low) - 9(high):
    - reflects level of confidence of prediction
    - higher RI indicates better accuracy, but lower coverage
  -

nsSNP	Prediction	Reliability Index	Expected Accuracy
N61D	Non-neutral	1	63%

MutationTaster

- Input:
  - HGNC gene symbol, NCBI Gene ID, Ensembl gene ID
  - Possible to analyse SNPs, insertions and deletions
  - Nucleotide substitutions and its position
- Analyses, if an amino acid is changed or not
- Method:
  - Integrates information from different biomedical databases and uses analysis tools
  - three different prediction models:
    - 'silent' synonymous or intronic alterations (without_aae)
    - single amino acid(simple_aae)
    - complex changes in the amino acid sequence (complex_aae)

N61D – A181G

- Prediction
  - disease causing
  - Model: simple_aae
  - prob: 0.9999999901823

- MutationTaster performed best in terms of accuracy and speed (according to their paper)

Thank your for your attention!

Any questions?

References

- A method and server for predicting damaging and missense mutations
- *SNAP: predict effect of non-synonymous polymorphisms on function.* Yana Bromberg and Burkhard Rost
  Nucleic Acids Research, 2007, Vol. 35, No. 11 3823-3835 (PubMed)
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. Nat Methods**7**(4):248-249 (2010)
- Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, Nature Protocols 4, - 1073 - 1081 (2009)
- Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods. 2010 Aug;7(8):575-6.*
- Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. Genome Res 11: 863–874.
- http://swift.cmbi.ru.nl/teach/Wageningen/IMAGE/aa_venn_diagram.png , 15.6.13