SomeNA

What is SomeNA?

- Diploma thesis by P. Hönigschmid (2012)
- is a neural network based classifier predicting whether a residue binds to D/RNA molecules
- is an overhauled version of DISIS

DISIS

- Idea: Similarly as in protein-to-protein interface prediction we would like to be able to predict DNA-binding sites in DNA-binding proteins using as little information as possible (i.e.: no information on the 3D structure)

- The first attempt to use evolutionary relationship and sequence alone to predict such features was made in 2004 by Ahmad et al.

- To date, some methods use tertiary structure, while others rely on sequence alone.

DISIS II

- The motivation behind DISIS is to deliver a prediction method that would not rely on 3D structure, as getting this information is complicated and costly.
- DISIS relies on the principle that DNA-binding residues have distinct biophysical characteristics, thus the method intends to demonstrate that these characteristics are so distinct that they enable accurate prediction of the residues that bind DNA directly from amino acid sequence
- DISIS has the advantage of being applicable to all proteins, as it doesn't require 3D structure
- DISIS uses machine learning, and in particular neural networks to predict DNA-binding sites
- DISIS has been later extended to DISIS2 which from sequence predicts: secondary structure, solvent accessibility, disorder, b-value, protein-protein interaction coiled coils, and evolutionary profiles, etc. The amount of predicted features is much larger than of DISIS (previous version). Finally, DISIS2 is able to predict DNA-binding residues from protein sequence of DNA-binding proteins.

Example PP output


```
cdallago@n03:~$ ls /mnt/project/ppcache20/18/1c/181c0652a018e803f00a5df690f7446e0cb55d0c/
query.blastPsiAli.gz  query.isis              query.prof1Rdb
query.blastPsiMat     query.loctreeAnimal     query.profAscii
query.blastPsiRdb     query.loctreeAnimalTxt  query.profb4snap
query.blastpSwissM8   query.loctreePlant      query.profbval
query.chk             query.loctreePlantTxt   query.profRdb
query.clustalngz      query.loctreeProka      query.proftmb
query.coils           query.loctreeProkaTxt   query.proftmbdat
query.coils_raw       query.mdisorder         query.prosite
query.disulfinder     query.nls               query.psic
query.fasta           query.nlsDat            query.segNorm
query.hmm2pfam        query.nlsSum            query.segNormGCG
query.hmm3pfam        query.nors              query.seqGCG
query.hmm3pfamDomTbl  query.norsnet           query.sumNors
query.hmm3pfamTbl     query.phdNotHtm         query.tmhmm
query.hsspPsiFil.gz   query.phdPred
query.in              query.phdRdb
```

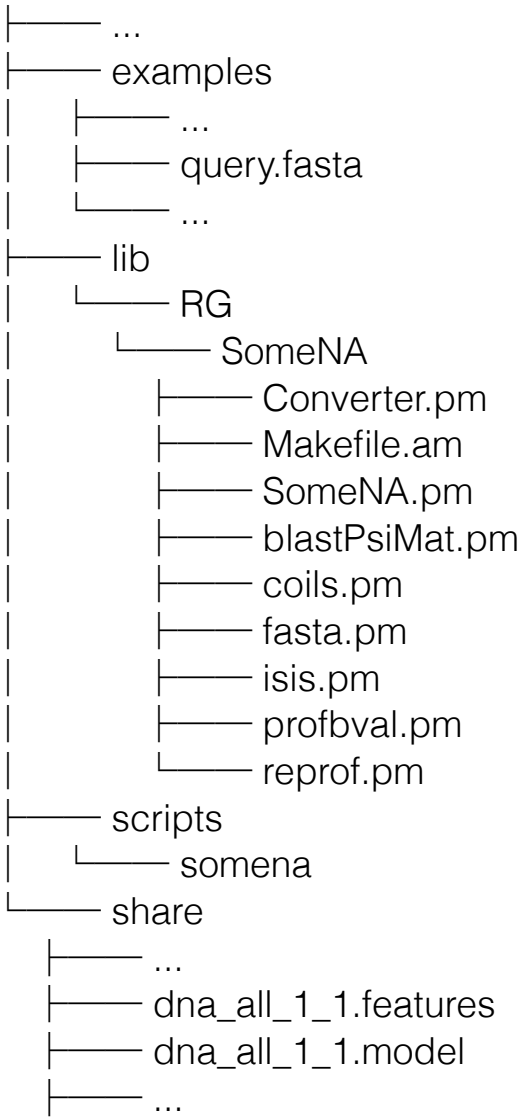
Example SomeNA output I


```
############################################################
# SomeNA - Prediction of DNA- and RNA-Binding Proteins
############################################################
# Row Format
#  The first column is the TYPE of the row
#  #                : Comment
#
#  DNA_JURY         : DNA-binding single prediction
#  RNA_JURY         : RNA-binding single prediction
#  XNA_JURY         : XNA-binding single prediction
#  DNA_COMB_JURY    : Combined DNA-binding prediction
#                     (has XNA-binding as prequisite and excludes RNA-binding; very precise)
#  RNA_COMB_JURY    : Combined RNA-binding prediction
#                     (has XNA-binding as prequisite and excludes DNA-binding; very precise)
#
#  HEADER           : The header line for the PRP rows
#
#  PRP              : Per residue predictions
############################################################
# PRP Column Format
#  NO               : Amino acid position
#  RES              : Amino acid one-letter code
#  DNA_AVG          : Average of direct scores of the DNA models
#  DNA_JURY         : Fraction of models that predicted DNA-binding
#  RNA_AVG          : Average of direct scores of the RNA models
#  RNA_JURY         : Fraction of models that predicted RNA-binding
#  XNA_AVG          : Average of direct scores of the XNA models
#  XNA_JURY         : Fraction of models that predicted XNA-binding
```

Example SomeNA output II

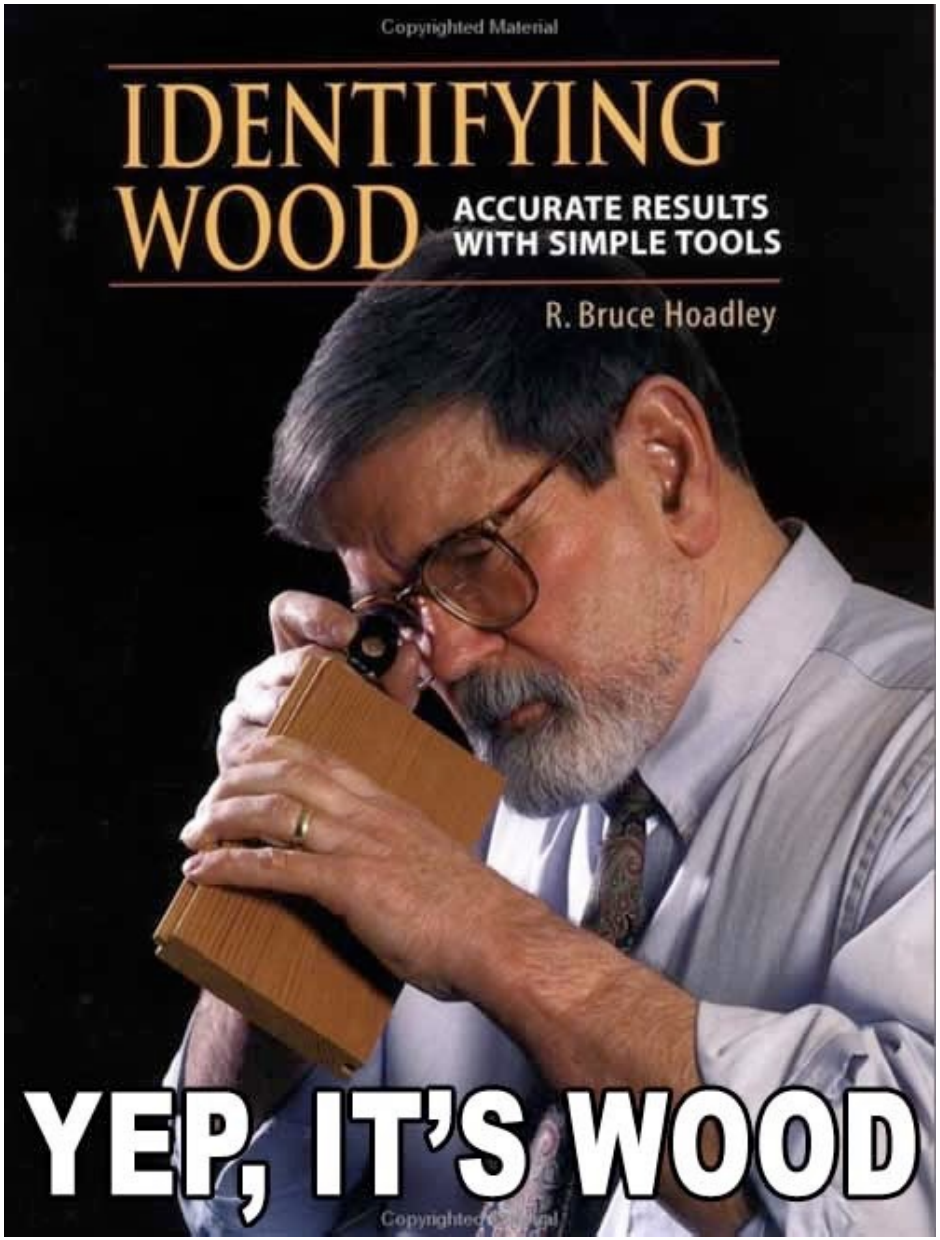

```
############################################################
# Notes
#  The _JURY suffixed rows and columns can be interpreted as
#  positive/yes prediction if they show a value above 0.5,
#  meaning that the majority of network models predicted
#  the positive class
############################################################
DNA_JURY 1.00
RNA_JURY 1.00
XNA_JURY 1.00
DNA_COMB_JURY   0.20
RNA_COMB_JURY   0.60
HEADERNO RESDNA_AVG   DNA_JURY  RNA_AVG  RNA_JURY    XNA_AVG XNA_JURY
PRP1  A  -0.16 0.40  0.33  0.80  0.06  0.80
PRP2  R  0.66  1.00  0.78  1.00  0.79  1.00
PRP3  T  0.44  1.00  0.42  1.00  0.48  1.00
PRP4  K  0.62  1.00  0.74  1.00  0.78  1.00
PRP5  Q  0.31  1.00  0.63  1.00  0.51  1.00
PRP6  T  0.20  0.80  0.47  1.00  0.38  1.00
PRP7  A  0.31  1.00  0.53  1.00  0.38  1.00
PRP8  R  0.43  1.00  0.68  1.00  0.57  1.00
PRP9  K  0.34  1.00  0.39  1.00  0.53  1.00
PRP10 S  0.37  1.00  0.20  0.80  0.32  1.00
PRP11 T  0.44  1.00  0.49  1.00  0.54  1.00
PRP12 G  0.26  1.00  0.43  1.00  0.31  1.00
PRP13 G  0.19  0.80  0.20  0.80  0.30  1.00
PRP14 K  0.45  1.00  0.70  1.00  0.63  1.00
PRP15 A  0.10  0.80  0.47  1.00  0.21  0.80
PRP16 P  0.18  0.80  0.56  1.00  0.47  1.00
.....
```

Folder Structure



About documentation

Virtually none, only `scripts/somena` file contains information about input and output


Copyrighted Material
IDENTIFYING WOOD
ACCURATE RESULTS WITH SIMPLE TOOLS
R. Bruce Hoadley
YEP, IT'S WOOD

- scripts/somena is the main executable
- `SomeNA.pm` contains prediction algorithm
- `blastPsiMat.pm`, `coils.pm`, `fasta.pm`, `isis.pm`, `profbval.pm`, and `reprof.pm` are used to parse corresponding PP output files

- Docker allows to build, ship and run applications
- Github for executables with all their dependencies
- Download a Docker image and run it without worrying about libraries and operating system

How to

- Download docker from https://www.docker.com/
- Type

```
docker run -it -d -P --name somena -v /local/folder/with/PP/output:/shared rost/
somena bash
```

- Then you will have a running instance of a SomeNA Virtual Machine on your computer that has a shared folder with your host machine
- You can then use

```
docker attach somena
somena -i /shared
```

Next steps

- Contacted P. Hönigschmid
- Obtain his thesis
- Refine Docker image
- Create documentation for SomeNA