FuncNet

a distributed platform for high-throughput protein function analysis

Andrew Clegg
University College London

funcnet.eu
Outline of talk

► Introduction and background
► Working with FuncNet
► APIs and extensions
► Further information
Aims of FuncNet

Given one set of proteins which are known to share a particular biological function...

… which of these other proteins also share that function?
Usage scenario

Erich Nigg – Max Planck Inst. – proteomics study.
150 known spindle proteins
> 600 unknown/novel proteins

How can we help identify true spindle proteins?

UCL, DTU and CNIO each contributed distinct prediction algorithms.
Experimentalists using this approach achieved **doubled hit rate** for wet-lab assays.
Motivations

Turning manual analysis into a pipeline...

- Accessibility (zero installation)
- Automation
- Throughput
- Robustness
Outline of talk

- Introduction and background
- Working with FuncNet
- APIs and extensions
- Further information
Reference proteins: known function

Query proteins: unknown function
Reference proteins: known function

Query proteins: unknown function

predictor_1  predictor_2  predictor_3
Outputs

Reference proteins: known function

Query proteins: unknown function

Fisher score = \(-2 \sum_{i=1}^{k} \log_e(p_i)\)
# Prediction services

<table>
<thead>
<tr>
<th>Service</th>
<th>Institution</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>CODA</td>
<td>UCL (London)</td>
<td>Genomic context – domain fusion</td>
</tr>
<tr>
<td>engineDB</td>
<td>ITB (Bari)</td>
<td>Similarity of GO annotations</td>
</tr>
<tr>
<td>GECO</td>
<td>UCL (London)</td>
<td>Gene expression correlation</td>
</tr>
<tr>
<td>hiPPI</td>
<td>UCL (London)</td>
<td>Protein interactions via domain homology</td>
</tr>
<tr>
<td>iHOP</td>
<td>CNIO (Madrid)</td>
<td>Text mining – MEDLINE co-occurrence</td>
</tr>
<tr>
<td>JACOP</td>
<td>SIB (Lausanne)</td>
<td>Subsequence homology</td>
</tr>
<tr>
<td>PIPS</td>
<td>Dundee University</td>
<td>Protein interactions via Bayesian classifier</td>
</tr>
<tr>
<td>SpindleP</td>
<td>CBS (Lyngby)</td>
<td>Neural network using sequence &amp; structure</td>
</tr>
</tbody>
</table>
Overview of workflow

Query proteins

Reference proteins

SpindleP adaptor

FuncNet front-end server

- CODA (CATH)
- CODA (Pfam)
- GECO
- hiPPI
- iHOP
- iHOP-Conn
- JACOP
- engineDB
- PIPS

Score aggregation

Pairwise predictions
FuncNet web interface

- Upload protein lists from your PC
- Monitor progress of job
- Download results as CSV or XML

http://funcnet.eu/gwt-client/
FuncNet session status report.

Large queries may take some time (minutes or hours) to complete on all predictors. This page refreshes every 60secs.

You can bookmark this page to return to this session later, or take a note of the job ID and the email address you used.

User: anonymous-web-user@funcnet.eu   Job ID: caffduba/-124554:1227e0ebeda-6849

- Cancel Job
- Refresh Now

CodaCathService: WORKING
CodaPfamService: WORKING
GecoService: WORKING
HippiService: COMPLETE
JacopService: WORKING
PIPsPredictorService: COMPLETE
iHOPFuncNetConnService: WORKING
iHOPFuncNetService: WORKING
FuncNet job results.

The following table shows the FuncNet scores for each pair of possibly interacting proteins, where a higher score indicates a better prediction. Also included are the prediction p-values for each predictor. The lower the p-value, the more likely that this is a genuine interaction, in the view of one particular predictor. Statistically significant predictions are highlighted in green, and predictors that failed due to server-side or network errors are show in red. Click a protein to see its UniProt page.

You can bookmark this page to return to for up to a week.

User: anonymous-web-user@funcnet.eu

Results for job caffduyaf/-124554:1227e0ebda:-683d

<table>
<thead>
<tr>
<th>Query protein</th>
<th>Reference protein</th>
<th>FuncNet score</th>
<th>CodaCathService p-value</th>
<th>CodaPfamService p-value</th>
<th>GecoService p-value</th>
<th>HippiService p-value</th>
<th>iHOPFuncNetConnService p-value</th>
<th>iHOPFuncNetSeq p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q99728</td>
<td>P38398</td>
<td>62.48168</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.00582</td>
<td>0.00006</td>
</tr>
<tr>
<td>Q99728</td>
<td>P51587</td>
<td>25.65159</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.11650</td>
<td>0.30791</td>
</tr>
<tr>
<td>Q9P0W2</td>
<td>P51587</td>
<td>22.92018</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.42149</td>
<td>0.000003</td>
</tr>
<tr>
<td>Q9Y6A5</td>
<td>O75410</td>
<td>20.97225</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.12490</td>
<td>0.21760</td>
</tr>
<tr>
<td>Q9UM11</td>
<td>P38398</td>
<td>19.85311</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.42149</td>
<td>0.74789</td>
</tr>
<tr>
<td>Q9Y6A5</td>
<td>P43487</td>
<td>17.19031</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>Q9Y6A5</td>
<td>O95359</td>
<td>16.62138</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>0.54428</td>
<td>0.36526</td>
</tr>
</tbody>
</table>
Outline of talk

► Introduction and background
► Working with FuncNet
► APIs and extensions
► Further information
FuncNet SOAP API

- Workflow tools like Taverna, KNIME, Bioclipse, Pipeline Pilot, etc. etc.
- Open-source clients (Java and Perl so far)
- Your own scripts or programs
- Each predictor can be queried separately
FuncNet in EnVision2

EnVision II - Microsoft Internet Explorer

Address: [http://www.ebi.ac.uk/enfin-srv/ envision2/pages/homePage.jsf](http://www.ebi.ac.uk/enfin-srv/envision2/pages/homePage.jsf)

Start EnVision
Please note that Envision2 uses sessions. This means that after a certain period of inactivity (120 minutes) your session and all its associated information (datasets, results) will be deleted.
You can monitor the status of all your datasets from the Dataset Manager Page.

If you need help just click on this icon or go to the User Manual Page for help.

To perform an Envision 2 query, optionally provide a name for your dataset here

- e.g. 'myDataset'

and provide input for your dataset

- Protein Accessions
- Protein Sequences
- Enfin XML Document

Please select how you want to provide the input for this dataset

- Direct Input
- File Upload
- URL Load

Please input or paste various Protein Accession Numbers below
Spindle protein prediction

Query proteins → FuncNet front-end server

Reference proteins → SpindleP adaptor

SpindleP → Score aggregation

CODA (CATH) → CODA (Pfam) → GECO → hiPPI → iHOP → iHOP-Conn → JACOP → engineDB → PIPS

Pairwise predictions

Integrating with SpindleP
Query set

Reference set = SpindleP training set

SpindleP neural network

FuncNet front-end service

Score integration
Outline of talk

► Introduction and background
► Working with FuncNet
► APIs and extensions
► Further information
Work in progress

- Evaluation – precision & recall
- Load testing
- Other species – currently human-only
- Score integration – Bayesian classifier
- Annotation of results
- Visualization
- Wet-lab partners?
The FuncNet consortium

University College London
- Christine Orengo
- Corin Yeats
- Jonathan Lees
- Ian Sillitoe
- Adam Reid

University of Malaga
- Juan Ranea
- Ian Morilla

SIB, Lausanne
- Marco Pagni
- Heinz Stockinger

EBI, Hinxton
- Henning Hermjakob
- Florian Reisinger

CNR-ITB, Bari
- Andreas Gisel

CNIO, Madrid
- Alfonso Valencia
- José María Fernández González
- David González Pisano
- Ana Rojas

CBS, Lyngby
- Søren Brunak
- Kristoffer Rapacki

University of Dundee
- Geoff Barton
- Mark McDowall
- Michelle S. Scott
- Tom P. Walsh

... and myself, Andrew Clegg
# FuncNet

a distributed protein function analysis network

## Current status

<table>
<thead>
<tr>
<th>Service Identifier</th>
<th>Status</th>
<th>Response Time</th>
<th>Checked On</th>
</tr>
</thead>
<tbody>
<tr>
<td>FrontEndService</td>
<td>OK</td>
<td>72.30s</td>
<td>April 30, 2009 10:30:05 AM GMT+01:00</td>
</tr>
<tr>
<td>GecoService</td>
<td>OK</td>
<td>2.48s</td>
<td>April 30, 2009 10:31:18 AM GMT+01:00</td>
</tr>
<tr>
<td>HippiService</td>
<td>OK</td>
<td>0.17s</td>
<td>April 30, 2009 10:31:20 AM GMT+01:00</td>
</tr>
<tr>
<td>CodaCathService</td>
<td>OK</td>
<td>2.55s</td>
<td>April 30, 2009 10:31:21 AM GMT+01:00</td>
</tr>
</tbody>
</table>