Multidisciplinary collaboration to facilitate hypotheses generation in Huntington’s Disease

Eleni Mina, Mark Thompson, Kristina M Hettne, Willeke van Roon-Mom, Rajaram Kaliyaperumal, Eelke van der Horst, Katherine Wolstencroft, Barend Mons, Marco Roos

Leiden University Medical Center
The Netherlands
eScience approach
our definition

• Collaboration
• Reusable components
• User interface
• Semantic web standards to expose methods and results
• Generate new insights for rare diseases
People involved

Willeke van Roon-Mom:
Huntington's Disease expert

Eleni Mina
Bioinformatician

Mark Thompson
Computer Scientist
Our eScience model

Hypothesis  Analysis  Interpretation  Outcome
Hypothesis
Progressive neurodegeneration
Caused by a single gene mutation (htt-QQQ)
Autosomal dominant inheritance
No cure or effective treatment yet

**Symptoms**
Chorea
Cognitive decline
Dystonia
Psychological & Behavioral Problems
Weight loss, muscle wasting
Metabolic dysfunction
Bradykinesia
Alterations in gene expression is a central mechanism in the development of HD
Mechanisms remain unknown
Epigenetics in HD

Changes in gene expression that are not due to alterations in DNA sequence

DNA coding defects are known

Epigenetic changes are not known
Epigenetic changes are associated with disease mechanisms

Link between gene expression and epigenetics in HD
General approach

Human brain data

Differentially expressed genes

map to chr location

Statistical analysis with R

data integration/analysis

CpG islands Chromatin marks

Hypothesis X

Hypothesis Y

Hypothesis Z

HD gene list

* wet lab validation

Nanopub.org

* microarray gene expression analysis of 3 different human brain regions, Hodges et al, 2006
eScience model

Analysis
Analysis:
Get chromosome location

Input:
file from microarray experiment

Output:
map genes to chromosome location
compute promoter region
Analysis:
Operate on genomic intervals

Inputs: gene file
epigenetic information

Transform files & Compute overlaps

Compute statistical significance
Analysis: output

Gene list affected by epigenetics

Statistical test measuring the effect
How is this gene list linked to Huntington's disease?

What can we tell more about this gene list?
eScience model

Interpretation
Concept profiling

Concept profile X:
All concepts associated with X in pubmed abstracts

Weight $w_{XA}$ calculated using the uncertainty coefficient
Interpretation: Link between gene list and HD

Concept profiling

Implicit associations
Inner product between vectors X & Y
Interpretation: concept profiling

Concept profiling score

Prioritize gene list
Prioritize genes related to *HD*
In blue: concepts not co-mentioned in literature before with HD (implicit associations)

<table>
<thead>
<tr>
<th>Concept</th>
<th>Sum</th>
<th>HTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCBP2</td>
<td>0.0165</td>
<td>0.0165</td>
</tr>
<tr>
<td>DLL1</td>
<td>0.0164</td>
<td>0.0164</td>
</tr>
<tr>
<td>ATXN10</td>
<td>0.0164</td>
<td>0.0164</td>
</tr>
<tr>
<td>AP2A2</td>
<td>0.0163</td>
<td>0.0163</td>
</tr>
<tr>
<td>GRIA3</td>
<td>0.0161</td>
<td>0.0161</td>
</tr>
<tr>
<td><strong>BAIAP2</strong></td>
<td><strong>0.0161</strong></td>
<td><strong>0.0161</strong></td>
</tr>
<tr>
<td>BCKDHA</td>
<td>0.0157</td>
<td>0.0157</td>
</tr>
<tr>
<td>BCL2L11</td>
<td>0.0154</td>
<td>0.0154</td>
</tr>
<tr>
<td>PRNP</td>
<td>0.0151</td>
<td>0.0151</td>
</tr>
<tr>
<td>CREB1</td>
<td>0.0151</td>
<td>0.0151</td>
</tr>
<tr>
<td>UBE2A</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>APLP2</td>
<td>0.0149</td>
<td>0.0149</td>
</tr>
<tr>
<td>APAF1</td>
<td>0.0149</td>
<td>0.0149</td>
</tr>
<tr>
<td>RAB3A</td>
<td>0.0149</td>
<td>0.0149</td>
</tr>
<tr>
<td>GABARAP</td>
<td>0.0149</td>
<td>0.0149</td>
</tr>
<tr>
<td>MAP2</td>
<td>0.0148</td>
<td>0.0148</td>
</tr>
<tr>
<td>DLX5</td>
<td>0.0148</td>
<td>0.0148</td>
</tr>
<tr>
<td>SYP</td>
<td>0.0147</td>
<td>0.0147</td>
</tr>
<tr>
<td>SCAMP5</td>
<td>0.0147</td>
<td>0.0147</td>
</tr>
<tr>
<td>NCOR2</td>
<td>0.0145</td>
<td>0.0145</td>
</tr>
</tbody>
</table>
The whole model with pictures of what goes where

Outcome
Outcome

Candidate genes implicated in HD

hypotheses

differential expression in the disease

affected by an epigenetic feature

associated with HD & epigenetics
Preservation of results and materials and methods to enable collaboration & reuse
Enabling collaboration & reuse

Three digital objects that enable short & long term collaboration
1. Workflows
2. Nanopublications
3. Research objects
Taverna Workflow management system:

- Digital web analogue of a lab protocol
- Helps discussion with collaborators
- Automate processes in multistep analysis
- Integrate web services
- Supports Java, R, import spreadsheets
- Modular
- Reusable
Expose any valuable statement in a machine readable format
Enable: computational analysis, data integration & knowledge discovery

Enabling collaboration & reuse: nanopublications

sub: <HDAC1>, <malaria>, <mutation X>
prd: <interacts with>, <is transmitted by>, <has frequency>
obj: <ParvB>, <mosquitos>, <0.25%>
Enabling collaboration & reuse: nanopublications

What is in a nanopublication?

http://nanopublication.url

nanopub is resolvable and citable

Assertion
minimal biological statement
- association between concepts
- minimal (biological) statement
- experimental result
- hypothetical inference

Provenance
how the assertion came to be
- experimental methods
- context
- conditions
- assumptions, etc.

Pub Info
who deserves attribution
- authors, institutions and funding
- organizations
- lab technicians
- creators, curators
- rightsholders (+licence info)
convert natural language statements to RDF:

“BAIAP2 is associated with Huntington's Disease and epigenetics”
There is a gene disease association that refers to gene X, Huntington’s Disease and epigenetics

*define:*

- entities (concepts) 1,2,3,4
- relations (predicates)
Enabling collaboration & reuse: research objects

Aggregate resources related to our experiment
Enabling collaboration & reuse: research objects

Research Object Portal (http://sandbox.wf4ever-project.org/portal/home)
(RO support in myExperiment.org under development)
• eScience model guides the collaboration
• Computational implementation:
  (i) repeat the analysis
  (ii) discuss the experiment
  (iii) provide our collaborators with an instant user interface
  (iv) reusable semantic data

• Digital objects allow for preservation of results & materials and methods
Biosemantics

Kees Burger
Kristina Hettne
Peter Bram 't Hoen
Rajaram Kaliyaperumal
Erik Schultes
Mark Thompson
Marco Roos

Huntington's Disease group

Melvin Evers
Barry Pepers
Menno Schut
Lodewijk Toonen
Willeke van Roon-Mom

Open PHACTS

RDoConnect

Huntington’s Disease Society of America

NBIC

Netherlands bioinformatics centre