Provenance and Knowledge Abstraction for Reachability: A Framework for Knowledge Discovery

Delroy Cameron, Ph.D. Student
Ohio Center for Excellence in Knowledge Enabled Computing (Kno.e.sis)
Wright State University

Mentor: Olivier Bodenreider
Outline

• Background
• Reachability
  ▪ Provenance Metadata
  ▪ Knowledge Abstraction
• Approach
• Preliminary Results
• Knowledge Discovery
• Future Work
• Conclusion
Background

Crystal structures of beta-neurexin 1 and beta-neurexin 2 ectodomains and dynamics of splice insertion sequence 4.
Pre-synaptic nephrins (NRXs) bind to postsynaptic neurons (NLSs) to form Ca(2+)-dependent complexes that bridge neural synapses. beta-NRXs present crystal structures of the Delta isoforms of the LNS domains from beta-NRX1 and beta-NRX2, crystallized in the presence of Ca(2+) ions, in splice insertion site, with one coordinating ligand and donated by a glutamic acid from an adjacent beta-NRX1 molecule. NMR studies of beta-NRX1 and beta-NRX2 insertion sequence 4 may function independent of neuroligin binding.

Contribution of the BOP1 gene located on 3p24 to colorectal tumorigenesis
The most common form of genomic instability observed in colorectal cancer is chromosomal instability (CIN), whose molecular bases remain elusive. FES1, ORC6L, and RPL3, involved in ribosome biogenesis, altered chromosome segregation. To determine the contribution to colorectal tumour short fluorescent fragments, a screen to identify the BOP1 gene was conducted, revealing that in colorectal tumours the gene was found to be significantly mutated. The present study aims to elucidate the role of BOP1 in colorectal tumorigenesis and to identify potential therapeutic targets.

Physical and functional interaction between BOP1 and ribosomal proteins
Molecular mechanisms of mammalian cells. We showed that FES1 mutants selected by their ability to prevent or inhibit the efficient incorporation of ribosomal subunits into ribosomes. FES1 mutants have been shown to inhibit the association of ribosomal subunits with polysomes and to reduce the efficiency of translation initiation. This interaction is likely to be mediated by the FES1 protein, which has been shown to interact with ribosomal proteins. The interaction between FES1 and ribosomal proteins is likely to play a role in the regulation of translation initiation and in the assembly of ribosomes.
Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis.


INSERM U614, Faculté de Médecine et de Pharmacie, 22 Boulevard Gambetta, 76180 Rouen, France.

Abstract

The most common form of chromosomal instability observed in colorectal cancer is chromosomal instability (CIN), whose molecular bases remain to be determined. We have previously demonstrated that inactivation in human cells of several components of the Pest-Bop1 complex (BOP1, GRWD1, PEST1, ORC6L, and RPL3) involved in ribosome biogenesis, alterd chromosome segregation. To determine the contribution to colorectal tumorigenesis of somatic alterations of genes involved in ribosome biogenesis, we screened 58 primary colorectal cancers, using quantitative multiplex PCR of short fluorescent fragments, a sensitive method for the detection of gene dosage alterations. We found that dosage increase of the BOP1 gene was a frequent event, being detected in 39% of the tumors, and that we show an association with increased BOP1 mRNA. Scanning of 8q24, on which BOP1 is located, revealed that in colorectal cancers, gene dosage increase of BOP1 can be independent from that of MYC and was more frequent than that affecting MYC. Finally, transient overexpression of BOP1 in human cells increased the percentage of multipolar spindles. Together with our previous results, the present study strongly suggests that deregulation of the BOP1 pathway contributes to colorectal tumorigenesis.

PMID: 16804918 (PubMed - indexed for MEDLINE)

Publication Types, MeSH Terms, Substan...

Scooner Logo

Welcome, kno.eas

My Projects

New Project

New Project

Dashboard

Search: bop1

Workbench

bop1 (HPCO)

bop1 (UNLS)

Crystal structures of beta-neurexin 1 and beta-neurexin 2 ectodomains and dynamics of splice insertion sequence 4. Pre-synaptic neurexins (NRXs) bind to postsynaptic neuregins (NRGs) to form Ca(2+)-dependent complexes that bridge neural synapses. beta-NRX1 present crystal structures of the Delta domains of the N-terminal domain from beta-NRX1 and beta-NRX2, crystallized in the presence of Ca(2+) ions, splice insertion site, with one coordinating ligand donated by a glutamic acid from an adjacent beta-NRX1 molecule. NMR studies of beta-NRX1 beta-NRX1 (1-q24) insertion sequence 4 may function in roles independent of neural ligand binding.


Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis.

The most common form of chromosomal instability observed in colorectal cancer is chromosomal instability (CIN), whose molecular bases remain to be determined. Pest1, ORC6L, and RPL3, involved in ribosome biogenesis, alterd chromosome segregation. To determine the contribution to colorectal tumorigenesis of somatic alterations of genes involved in ribosome biogenesis, we screened 58 primary colorectal cancers, using quantitative multiplex PCR of short fluorescent fragments, a sensitive method for the detection of gene dosage alterations. We found that dosage increase of the BOP1 gene was a frequent event, being detected in 39% of the tumors, and that we show an association with increased BOP1 mRNA. Scanning of 8q24, on which BOP1 is located, revealed that in colorectal cancers, gene dosage increase of BOP1 can be independent from that of MYC and was more frequent than that affecting MYC. Finally, transient overexpression of BOP1 in human cells increased the percentage of multipolar spindles. Together with our previous results, the present study strongly suggests that deregulation of the BOP1 pathway contributes to colorectal tumorigenesis.


Physical and functional interaction between Pest1 and PEST1.

Molecular mechanisms of mammalian ribosome biogenesis. Pest1 mutants selected by their ability to reversibly arrest the essential for the efficient incorporation of Pest1 into nucleolar ribosomal subunits.

Outline

• Background
• **Reachability**
  ▪ Provenance Metadata
  ▪ Knowledge Abstraction
• Approach
• Preliminary Results
• Knowledge Discovery
• Future Work
• Conclusion
Reachability (Vertex)

“the notion of being able to get from one vertex in a directed graph to some other vertex”
Reachability (Documents)

“is the notion of being able to cover the documents in a document set, using the vertices in a directed graph from one vertex to some other vertex”
Outline

• Background
• Reachability
  ▪ Provenance Metadata
  ▪ Knowledge Abstraction
• Approach
• Preliminary Results
• Knowledge Discovery
• Future Work
• Conclusion
Provenance

• Meaning: origin, source of

• Provenance of Predications
  ▪ Biomedical Knowledge Repository (BKR)
    – UMLS Metathesaurus
      ○ 8 million relations
    – Biomedical Literature using SemRep
      ○ 13.4 million predications
  ▪ ProvenAnce Context Entity (PACE)¹

¹S.S. Sahoo, O. Bodenreider, P. Hitzler, A. Sheth, K., Thirunarayan, “Provenance Context Entity (PaCE): Scalable provenance tracking for scientific RDF data.”, in the 22nd International Conference on Scientific and Statistical Database Management (SSDBM) 2010
Outline

• Background
• Reachability
  ▪ Provenance Metadata
  ▪ Knowledge Abstraction
• Approach
• Preliminary Results
• Knowledge Discovery
• Future Work
• Conclusion
Knowledge Abstraction

- **Scenarios**
  - No Reachable Docs in DSP
  - No Successors in DSPG + Reachable Docs in DSP
  - No Successors in DSPG + No Reachable Docs in DSP

Stopping Condition (vertex)
- no successors
- no reachable docs
- no pseudo-sibling in a doc
Reachability Framework

- External Knowledge Base Graph Plane
- Document Set Predications Graph (DSPG) Plane
- Knowledge Abstraction
- Provenance
- Document Set Plane (DSP)
Dataset

• Documents (TREC 2006 Corpus)
  ▪ 26 Questions
  ▪ 1381 Gold Standard Documents
  ▪ 3461 Passages

• Document Predications Graph
  ▪ 2105 Vertices, 16942 Edges
  ▪ 1141 documents
  ▪ 240 no predications, 3 not processed

• External Knowledge Base Graph
  ▪ 13 million from UMLS Metathesaurus
  ▪ 8 million from Literature using SemRep
# TREC Genomics Track 2006 (Questions)

<table>
<thead>
<tr>
<th>Topic ID</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>What is the role of PrnP in mad cow disease?</td>
</tr>
<tr>
<td>161</td>
<td>What is the role of IDE in Alzheimer's disease</td>
</tr>
<tr>
<td>163</td>
<td>What is the role of APC (adenomatous polyposis coli) in colon cancer?</td>
</tr>
<tr>
<td>165</td>
<td>How do Cathepsin D (CTSD) and apolipoprotein E (ApoE) interactions contribute to Alzheimer's disease?</td>
</tr>
<tr>
<td>167</td>
<td>How does nucleoside diphosphate kinase (NM23) contribute to tumor progression?</td>
</tr>
<tr>
<td>168</td>
<td>How does BARD1 regulate BRCA1 activity?</td>
</tr>
<tr>
<td>169</td>
<td>How does APC (adenomatous polyposis coli) protein affect actin assembly</td>
</tr>
<tr>
<td>170</td>
<td>How does COP2 contribute to CFTR export from the endoplasmic reticulum?</td>
</tr>
<tr>
<td>172</td>
<td>How does p53 affect apoptosis?</td>
</tr>
<tr>
<td>173</td>
<td>How do alpha7 nicotinic receptor subunits affect ethanol metabolism?</td>
</tr>
<tr>
<td>174</td>
<td>How does BRCA1 ubiquitinating activity contribute to cancer?</td>
</tr>
<tr>
<td>176</td>
<td>How does Sec61-mediated CFTR degradation contribute to cystic fibrosis?</td>
</tr>
<tr>
<td>177</td>
<td>How do Bop-Pes interactions affect cell growth?</td>
</tr>
<tr>
<td>178</td>
<td>How do interactions between insulin-like GFs and the insulin receptor affect skin biology?</td>
</tr>
<tr>
<td>179</td>
<td>How do interactions between HNF4 and COUP-TF1 suppress liver function?</td>
</tr>
<tr>
<td>180</td>
<td>How do Ret-GDNF interactions affect liver development?</td>
</tr>
<tr>
<td>184</td>
<td>How do mutations in the Pes gene affect cell growth?</td>
</tr>
<tr>
<td>185</td>
<td>How do mutations in the hypocretin receptor 2 gene affect narcolepsy?</td>
</tr>
<tr>
<td>186</td>
<td>How do mutations in the Presenilin-1 gene affect Alzheimer's disease?</td>
</tr>
</tbody>
</table>
Graph Entry Point (GEP)

Q160: What is the role of PrnP in mad cow disease?

EG_5621: PRNP
C1418941: PRNP gene

Q161: What is the role of IDE in Alzheimer's disease

C2258633: insulin-degrading enzyme activity
C2258634: insulin-degrading neutral proteinase activity
Outline

• Background
• Reachability
  ▪ Provenance Metadata
  ▪ Knowledge Abstraction
• **Approach**
• Preliminary Results
• Knowledge Discovery
• Future Work
• Conclusion
Reachability Workflow

Graph Entry Point

Entity Identification Parser

Question

MDFS Algorithm

Paths

External Knowledge Base Graph

Document Set Predications Graph

Document Set
Reachability Ratio ($r_D$)

$$r_D = \frac{\text{Actual Number of Documents Reached}}{\text{Total Number of Documents to be Reached}}$$

Actual: 9
Total: 10
$$r_D = \frac{9}{10} = 0.9$$
Reachability Path ($\rho$)

An ordered list of ternary relations

root

External KB + Document Set Predications

Document Set Predications + Document Set

Document Set Plane
Path Analysis

- Inception Depth
- Minimal Maximum Reachability Path (MMRP)
- Most Relevant Subtree (MRS)
- Unconstrained Navigation
Inception Depth + MMRP

Inception Depth - Maximum number of document levels that must be explored to cover all documents

MMRP – Paths of Minimum Length among those Paths with Maximum coverage

BOP1 → PES1 -- 16043514
PES1 → rRNA processing -- 16043514
rRNA processing → Cell physiology ← WDR12 -- 16043514
WDR12 → S Phase -- 16043514
S Phase → Aneuploidy ← TP53 -- 12690111

Q177: How do Bop-Pes interactions affect cell growth?
Path Length = 5, Inception Level = 2, #GS = 2, #reached = 2

Collapsed Path: BOP1 inhibits PES1 affects WDR12 affects TP53.
Q174: How does BRCA1 ubiquitinating activity contribute to cancer?

Path Length = 59
Inception Depth=2
#Gold Std Docs = 19
#reached = 17
Q174: Path Length = 59
Inception Depth = 2
#GS = 19
#reached = 17
Outline

• Background
• Reachability
  ▪ Provenance Metadata
  ▪ Knowledge Abstraction
• Approach
• Preliminary Results
• Knowledge Discovery
• Future Work
• Conclusion
<table>
<thead>
<tr>
<th>Topic ID</th>
<th>Graph Entry Point (GEP)</th>
<th>#reached/#GS docs</th>
<th>Reachability Ratio ($r_D$)</th>
<th>Inception Depth ($\delta$)</th>
<th>MMRP ($\rho_{mm}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>165</td>
<td>C0007427/CATHEPSIN D</td>
<td>9/10</td>
<td>0.90</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>168</td>
<td>C1332381/BARD1 gene</td>
<td>30/43</td>
<td>0.69</td>
<td>9</td>
<td>65</td>
</tr>
<tr>
<td>170</td>
<td>EG_1080/CFTR (try C0538130/COP23)</td>
<td>3/4</td>
<td>0.75</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>171</td>
<td>EG_23246/BOP1</td>
<td>2/2</td>
<td>1.0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>172</td>
<td>C0021665/Insulin-Like Growth Factor I</td>
<td>3/3</td>
<td>1.0</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>173</td>
<td>EG_3172/HNF4A</td>
<td>12/12</td>
<td>1.0</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>174</td>
<td>EG_23481/PES1</td>
<td>2/2</td>
<td>1.0</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>175</td>
<td>C1113688/orexins</td>
<td>12/13</td>
<td>0.92</td>
<td>3</td>
<td>13</td>
</tr>
</tbody>
</table>
Outline

- Background
- Reachability
  - Provenance Metadata
  - Knowledge Abstraction
- Approach
- Preliminary Results
- Knowledge Discovery
- Future Work
- Conclusion
Most Relevant Subtree

Reachability Relation

Level-0
Level-1
Level-2
Level-3
Level-4
Q177: How do Bop-Pes interactions affect cell growth?

MRS: [BOP1, PES1, rRNA processing, __Cell physiology__, WDR12, S Phase]
Future Work

• Complete Reverse Engineering

• Pattern Learning
  ▪ Statistical Analysis
  ▪ Machine Learning

• Extensive Knowledge Abstraction
  ▪ Associative
  ▪ Hierarchical
  ▪ Hybrid

• Target Venue
  ▪ WWW 2011, Hyderabad, India.
  ▪ Deadline October 29, 2010
Conclusion

• Reachability
  • Provenance
  • Knowledge Abstraction
  • Path Analysis

• Knowledge Discovery
  • Reverse Engineering
  • Pattern Filtering
Acknowledgements

• Olivier Bodenreider
• Thomas Rindflesch
• Dongwook Shin
• John Nguyen
• Mohammed Cyclegar
• May Cheh
• Dr. Clement McDonald
• Lister Hill Center, NLM