Software for Graphical Notations in PathVisio and Lessons from its Usage

SBGN

4/29/2013

Augustin Luna

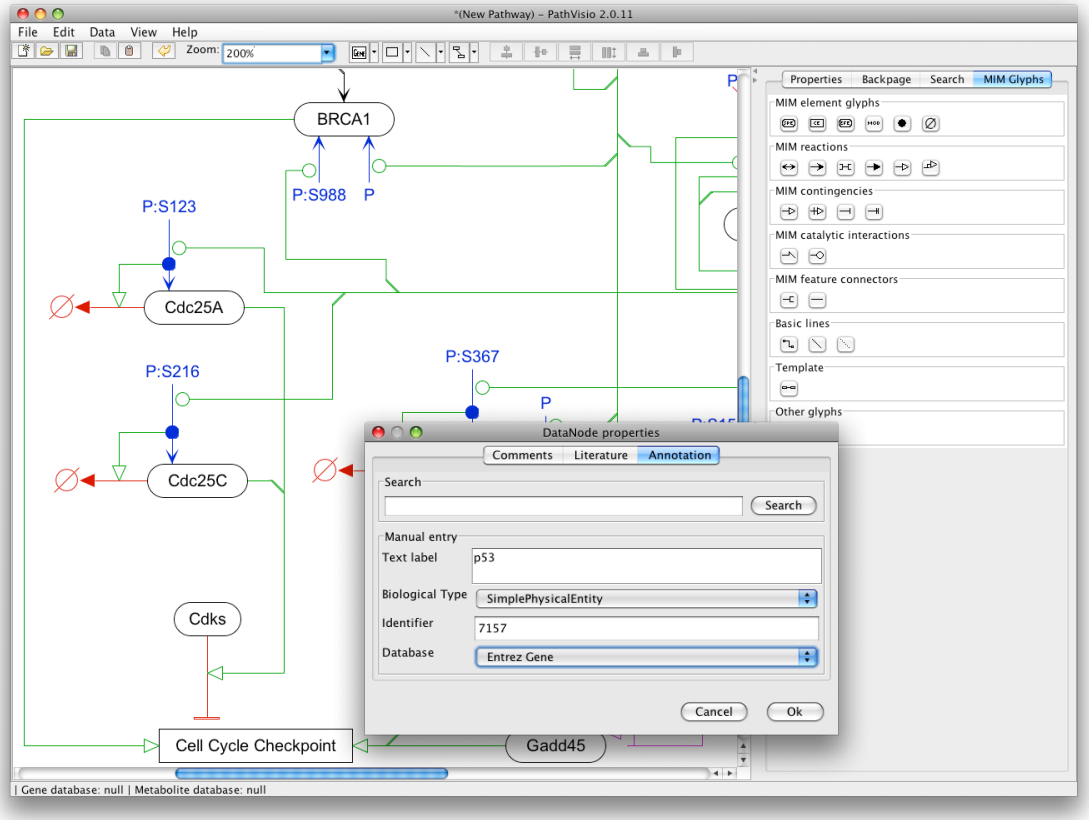
augustin@mail.nih.gov

Bioinformatics Program, Boston University, Boston, MA, USA

Laboratory of Molecular Pharmacology, National Cancer Institute, Bethesda, MD, USA

PathVisio

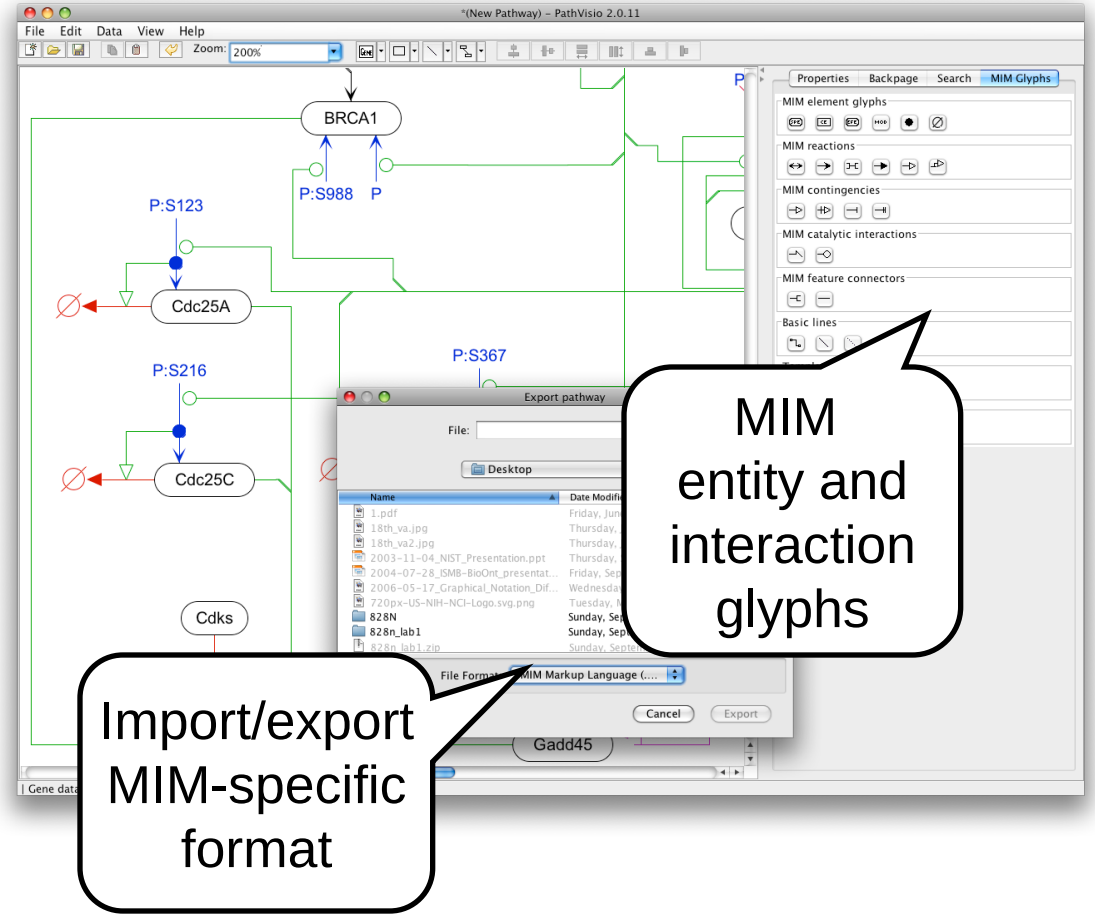
- Java-based
- Possesses a flexible plugin architecture
- Entities and interactions can be annotated with comments and literature references
- Associated pathway database, WikiPathways



van Iersel, MP et al. (BMC Bioinformatics) 2008

PathVisio-MIM Plugin

- Supports Molecular Interactions Map (MIM) notation created Kurt Kohn in 1999
  - Easy access to MIM glyphs
  - Import/export MIM-specific format meant for validation
- SBGN plugin has also been created for PathVisio



Luna A. et al. (Bioinformatics) 2011

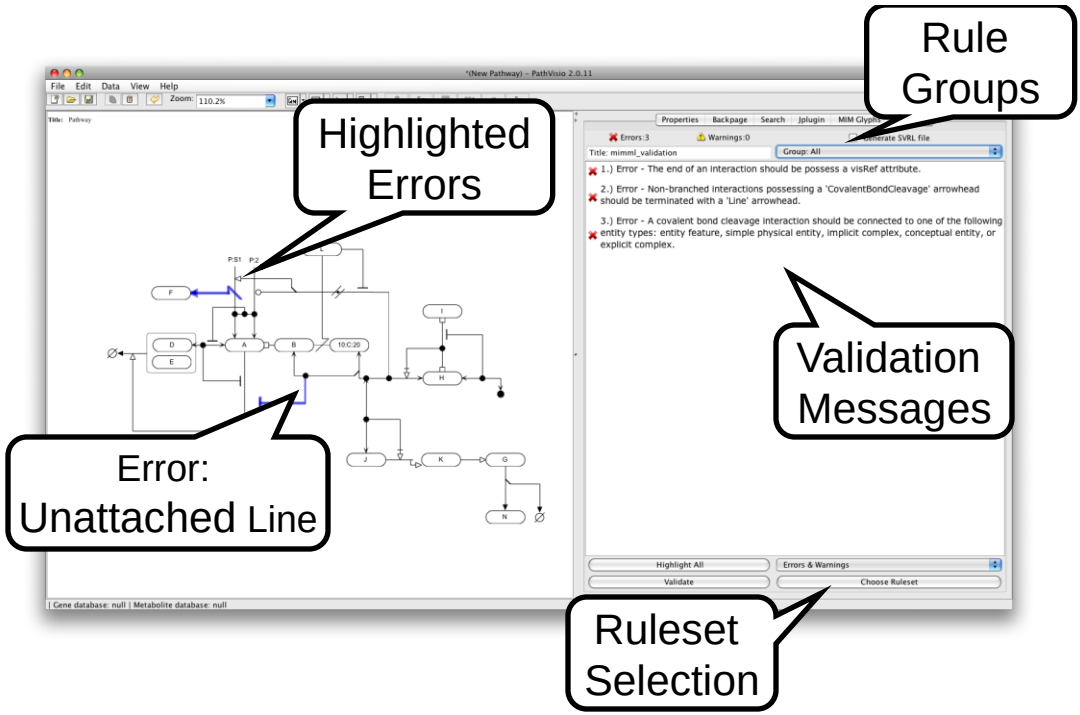
PathVisio-Validator Plugin

- Works with rulesets created using Schematron or Groovy
- Schematron: Generic XML validation framework
  - Language independent; stylesheets applied to XML datasets
- Groovy: Java-compatible scripting language
- MIM and SBGN rulesets created
- Other best practices rulesets created based on WikiPathways and E-MIMs curation rules

Schematron-Based Ruleset

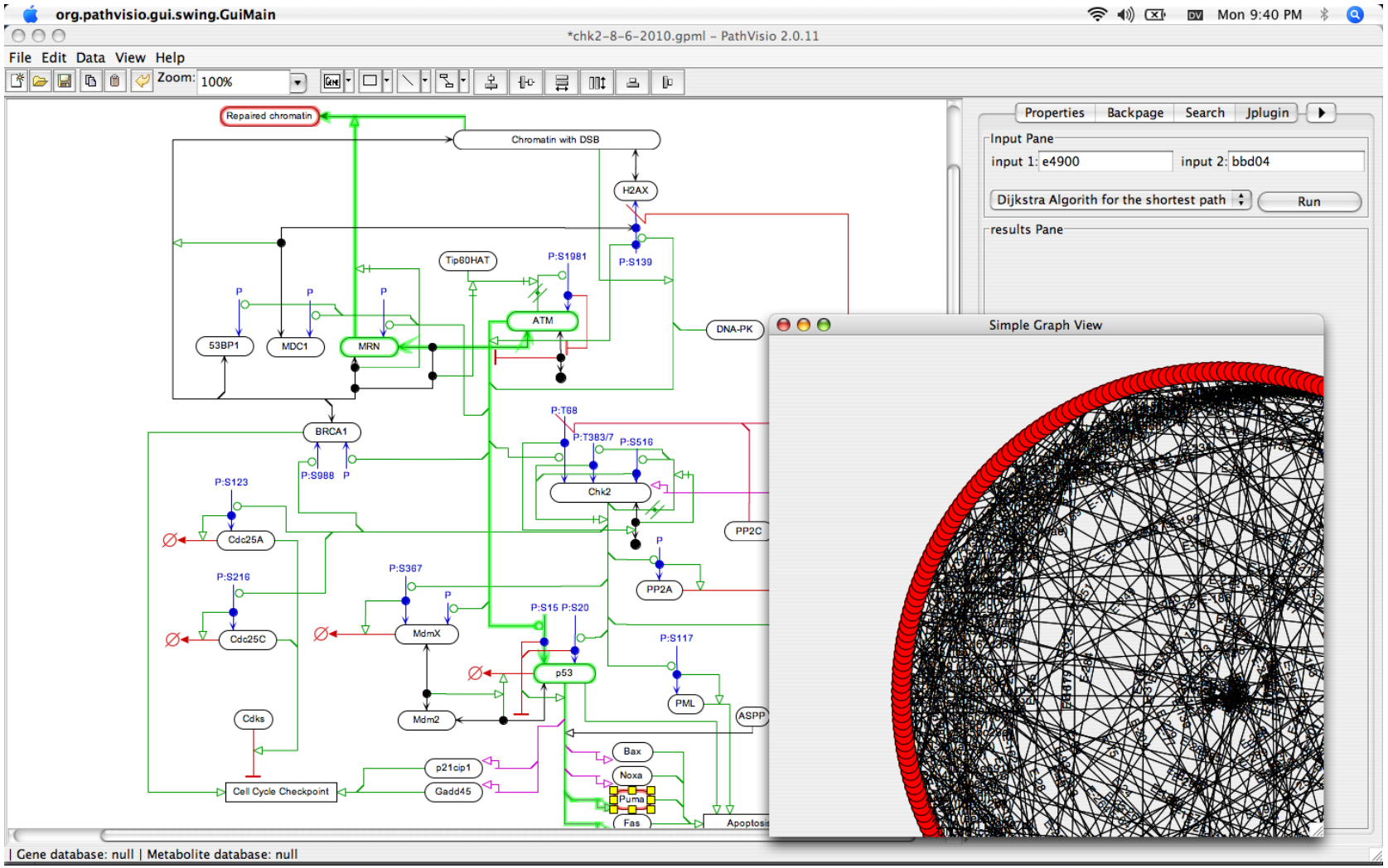
```
<!-- Validate Branching Interaction Arrowhead Prohibited Sets -->
<iso:pattern name="validate-branching-interaction-prohibited-sets" id="validate-br
    <iso:rule context="mimVis:InteractionGlyph/mimVis:Point[starts-with(@arrowHead

        <iso:let name="vis-id" value="../@visId"/>
        <iso:let name="inter" value="//mimVis:InteractionGlyph[@visId=$vis-id]"/>
```



Luna A. et al. (BMC Bioinformatics) 2011; Chandan K. et al., (Bioinformatics) 2012; van Iersel, MP et al. (Bioinformatics) 2012

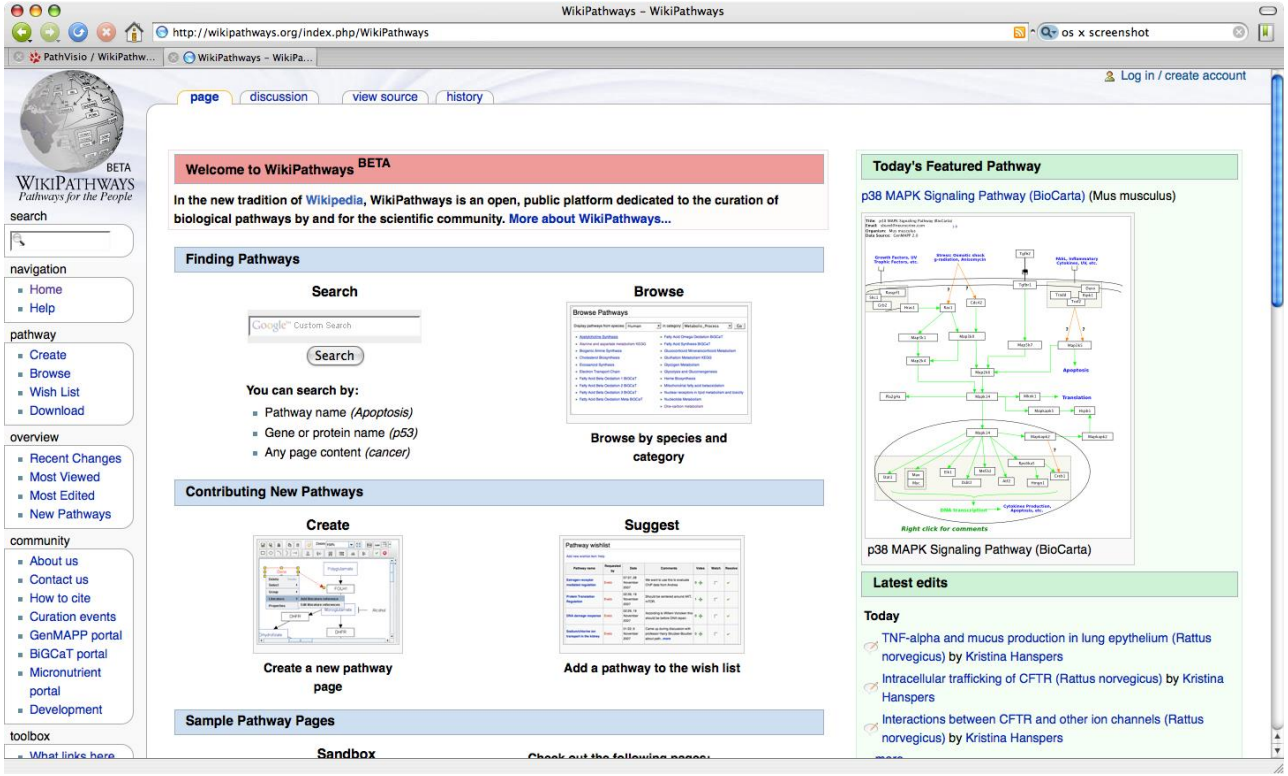
Network Analysis: Are Entities in a MIM Connected?



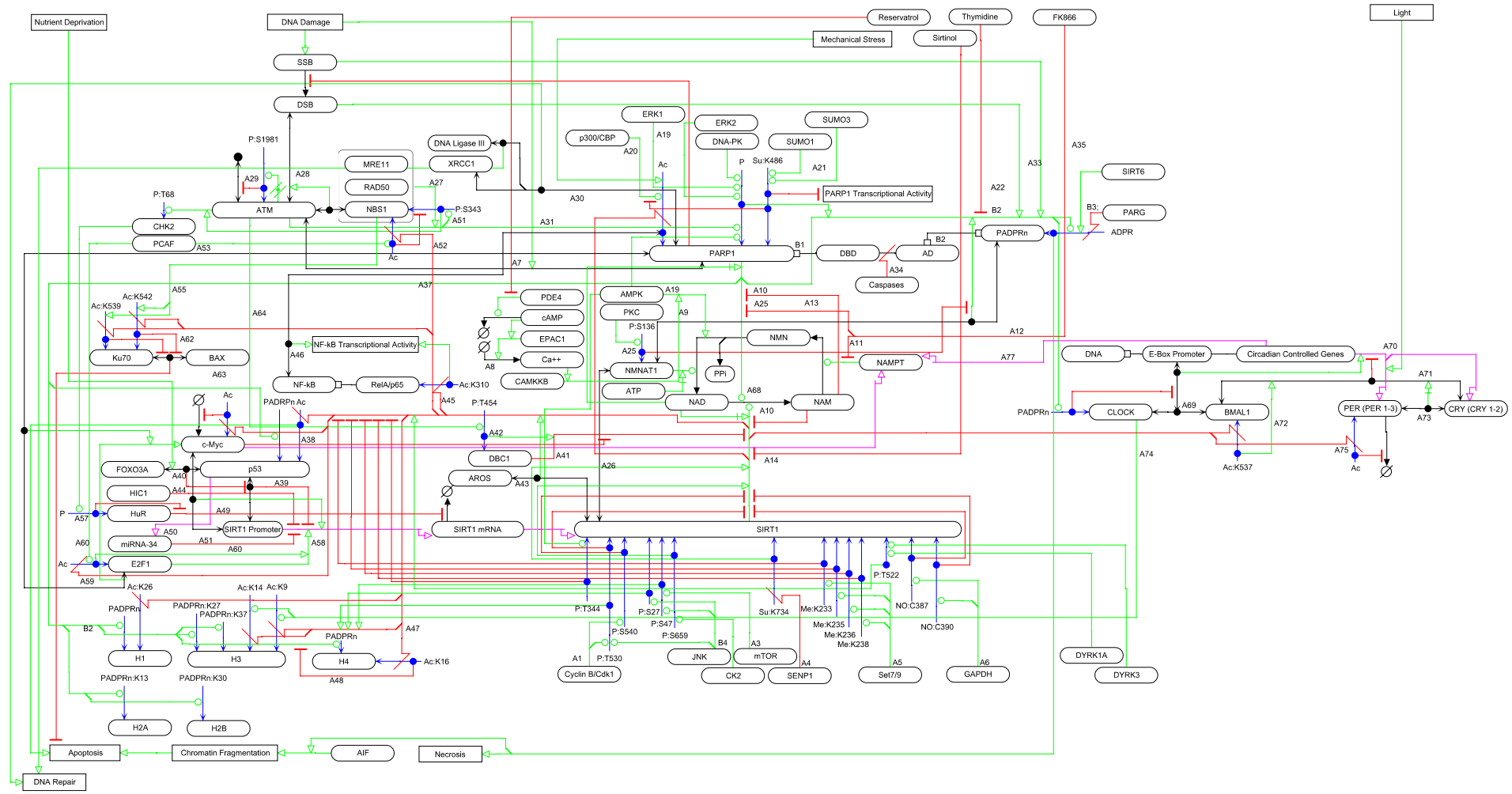
Chandan, K. and Luna, A.

WikiPathways Database

- Search for pathways, genes, proteins, etc.
- Need to extend support for SBGN and MIM



http://wikipathways.org

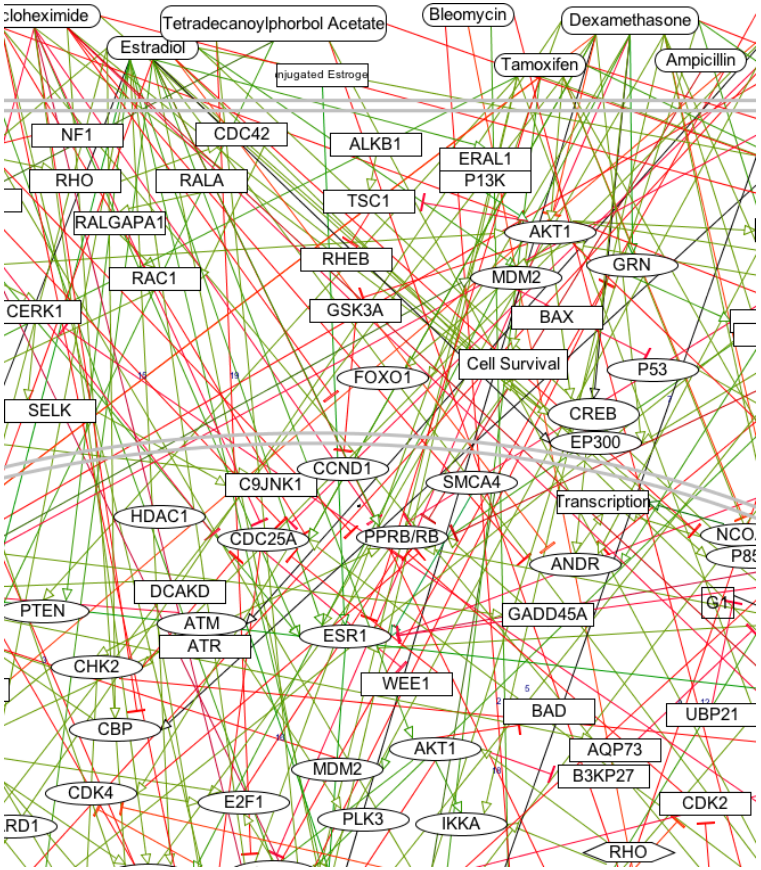


Review of SIRT1 and PARP1 Interactions: ~50 proteins, ~200 interactions; ~100 papers

Luna, A., (in preparation)

Layout, layout, layout…

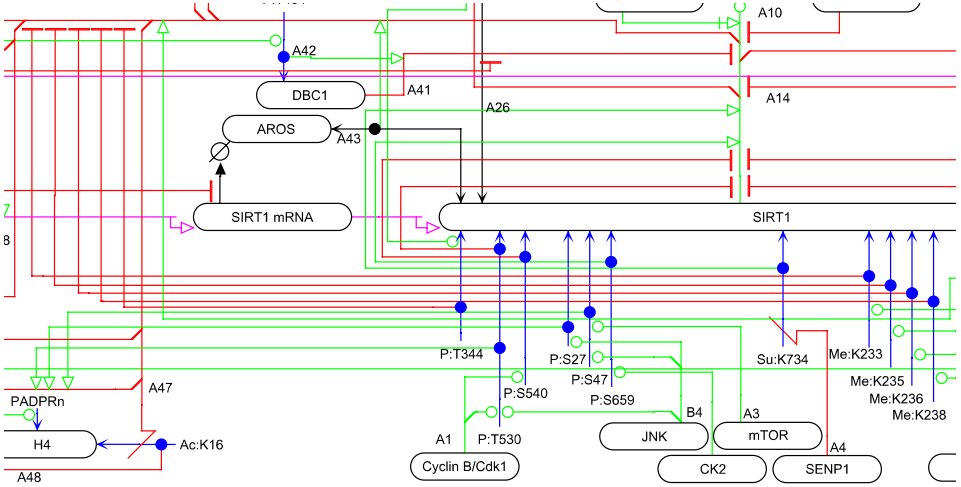
- Layout is the biggest challenge in large diagrams
- Orthogonal layouts help keep diagrams readable
  - PathVisio has limited interaction routing capabilities; can keep lines orthogonal
- Difficult to extract smaller diagrams with entity selection or expand diagrams with new entities
- PathVisio-MIM allows more connection points on entities



Breast Cancer Pathway
WikiPathways

Repetitive Interactions

- Repetitive interaction types
  - Add to the visual complexity
  - Make large diagrams difficult manage and layout



Many inhibitions stemming from entity modifications

Entity Identification Problems

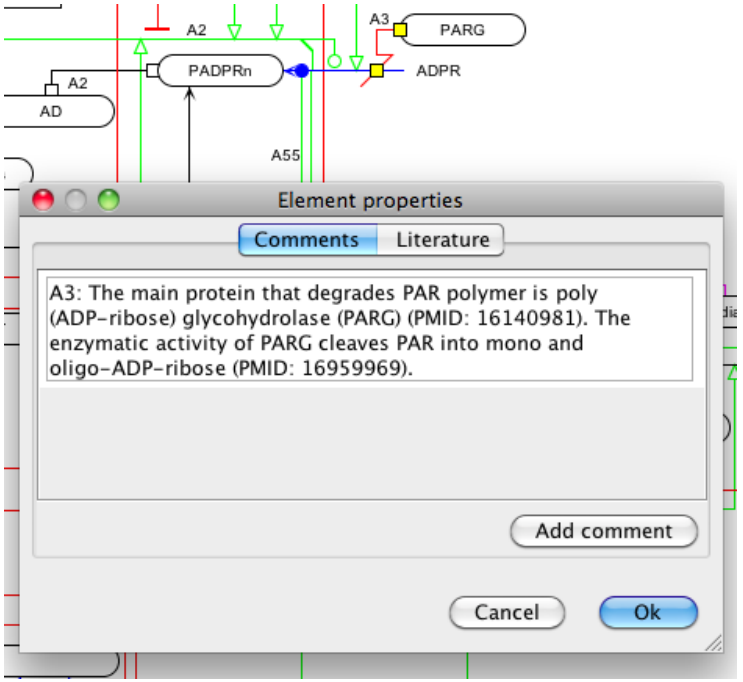
- Lack of information from publications for referencing some entities
  - Histones
  - Protein Kinase C (PKC); rarely are specific isomers referred to in publications

HIST1H1A	histone cluster 1, H1a
HIST1H1B	histone cluster 1, H1b
HIST1H1C	histone cluster 1, H1c
HIST1H1D	histone cluster 1, H1d
HIST1H1E	histone cluster 1, H1e
HIST1H1PS1	histone cluster 1, H1, pseudogene 1
HIST1H1PS2	histone cluster 1, H1, pseudogene 2
HIST1H1T	histone cluster 1, H1t
HIST1H2AA	histone cluster 1, H2aa
HIST1H2AB	histone cluster 1, H2ab
HIST1H2AC	histone cluster 1, H2ac
HIST1H2AD	histone cluster 1, H2ad
HIST1H2AE	histone cluster 1, H2ae
HIST1H2AG	histone cluster 1, H2ag
HIST1H2AH	histone cluster 1, H2ah
HIST1H2AI	histone cluster 1, H2ai
HIST1H2AJ	histone cluster 1, H2aj
HIST1H2AK	histone cluster 1, H2ak
HIST1H2AL	histone cluster 1, H2al

HGNC Identifiers for Histones

Linking Diagrams, Annotations, and Publication Text

- All interactions possess annotation labels
- Best practices MIM ruleset checks for properly annotated MIMs and annotation format
- Annotations eventually extracted to table
- Manuscript/diagram syncing tedious
  - PathVisio issue no easy way to renumber annotations



Annotation with multiple references

Desnoyers et al. 1999; Hassa, Haenni et al. 2006; Messner, Altmeyer et al. 2010). PARP1 is a transcriptional co-activator where PAR acts as a signal helping to regulate transcription (Luo and Kraus 2012). PAR is quickly cleaved by poly (ADP-ribose) glycohydrolase (PARG) [[A3]] (Hassa, Haenni et al. 2006). PARP1 becomes highly activated by DNA strand breaks;

Publication Manuscript

Notation/Tool Instructions

- Instruction overload for many users
  - How do you use the notation?
  - How do I use the tool?
  - How do I use the analysis plugins?

Contents

Faceted-Search Plugin Help

Acknowledgements

- NCI LMP
  - Mirit Aladjem
  - Kurt Kohn
  - Margot Sunshine
  - Sohyoung Kim
  - Lucas Chang (Cornell)
  - Noah Zweben (Columbia)
  - Katya Malison
- Bogazici University
  - Turkan Haliloglu
  - Can Ozturan
  - Itir Karac
  - Mine Edes
- NCI-Frederick
  - Ruth Nussinov
- Pathvisio Community
  - Martjin van Iersel (EBI)
  - Martina Kutmon (U of Maastricht)
- Google Summer of Code
  - Alexander Pico (UCSF)
  - Kumar Chandan (RIT)
  - Jake Fried (UMD)
- BioPAX Community
- SBGN Community