Roche CV mapping project
Automating alternative views of the GO

David Osumi-Sutherland, EBI
ROCHE CV

- 364 terms
- No definitions
- Each mapped to set of GO terms
- Term names mostly refer to
  - cell type
  - chemical
  - protein or protein family
  - biological process
  - molecular function
ROCHE CV

• What is it used for:
  • Term enrichment
    • Reduced redundancy compared to working with whole GO
    • Limited list of terms with appropriate specificity & relevance
    • Enrichment to orthogonal concepts useful to their research
      • cell type specific process
      • tissue specific processes
    • Example: mapping to cartilage processes useful when studying ectopic cartilage formation in fibrosis.
ROCHE CV
Approach

- Map each term to an OWL class expression:
  - Use this class expression to generate a list of mapped terms

- Store data as tsv on GitHub

- Generate tsv reports to
  - Compare manual mapping to automated mapping
  - Capture feedback from Roche
Approach - OWL

• Make union relations
  
  • e.g. `part_of OR_regualtes OR_capable_of_part_of`

  • Simple class expression + ELK (VERY FAST)

  • e.g. `part_of OR_regualtes OR_capable_of_part_of`

  • `some DNA repair`
Design patterns for terms

*is_a_OR_part_of_OR_regulates_OR_capable_of*

**label:** is a, part, regulates or functions in \{ key_process \}

**def:** \{ key_process \} OR a part of \{ key_process \} OR a process that regulates \{ key_process \} OR a cell component that functions in \{ key_process \}.

**equivalentTo:** 'process super query' some \{ key_process \}

*is_a_OR_part_of_OR_regulates*

**label:** is a, part of or regulates \{ key_process \}

**def:** \{ key_process \} OR a part of \{ key_process \} OR a process that regulates \{ key_process \}

**equivalentTo:** 'part_of_or_regulates' some \{ key_process \}
GitHub interaction

• Markdown
  • README.md in each directory for instructions – displayed by default
  • Autogenerated summary reports

---

**How to use this folder**

To get an overview, navigate to the `results_summary`.

• Tables are named for the RocheCVterm mapped, except that underscores are used in place of spaces.
• Tables in this folder may be edited to record whether a mapping has been checked, or blacklist.
  ◦ Use 0 for false and 1 for true.
  ◦ Edits to other cells will be overwritten during building.
  ◦ When editing, please bear in mind that:
    ▪ Each row must retain the correct number of tabs.
    ▪ The header rows must never be edited.
## BDNF pathway RCV_000002

- **Key class:** brain-derived neurotrophic factor receptor signaling pathway
- **Pattern:** is_a/OR_part_of/OR_regulates
- **map summary:** Roche_cvit: RCV_000002; class_expression 1C127FE1-B049-4E09-8DCC-8B323644160F some GO_0031547; manual_list_count 7, generated_list_count 8
- **Results**

<table>
<thead>
<tr>
<th>name</th>
<th>ID</th>
<th>manual</th>
<th>auto</th>
<th>checked</th>
<th>blacklisted</th>
</tr>
</thead>
<tbody>
<tr>
<td>brain-derived neurotrophic factor receptor signaling pathway</td>
<td>GO_0031547</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>regulation of brain-derived neurotrophic factor receptor signaling pathway</td>
<td>GO_0031548</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>negative regulation of brain-derived neurotrophic factor-activated receptor activity</td>
<td>GO_0031552</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>positive regulation of brain-derived neurotrophic factor-activated receptor activity</td>
<td>GO_0031553</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>negative regulation of brain-derived neurotrophic factor receptor signaling pathway</td>
<td>GO_0031549</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>positive regulation of brain-derived neurotrophic factor receptor signaling pathway</td>
<td>GO_0031550</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>regulation of brain-derived neurotrophic factor-activated receptor activity</td>
<td>GO_0031551</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>brain-derived neurotrophic factor-activated receptor activity</td>
<td>GO_0060175</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
GitHub interaction

• Version control
  • Git $\Rightarrow$ very solid versioning system
  • GitHub allows easy user interaction via web
Issue tracker

- Simple API
- Simple, flexible system of tags and milestone

This ticket is intended for discussion of the definition and mapping of BDNF pathway RCV_000002

- Summary of mapping
- Results table for mapping

@badilla: Mapping looks good. If you agree. Please sign off by closing this ticket with the comment 'resolved'. Thanks! David
Workflow

Map to pattern → Check results → Refine pattern/edit ontology → Roche approves pattern-based definition?

- No → Roche edits and approves mapping?
  - No → Release to Roche
  - Yes → Finished
- Yes → Roche approves pattern-based definition?
  - No → Release to Roche
  - Yes → Finished
Results

- 308/364 RCV terms:
  - Mapped to OWL queries
  - Results reviewed and accepted by Roche
  - New mapping now in daily usage at Roche
- Many gaps and errors found in GO and corrected
Manual mappings still required

- Most terms require only small numbers of additional, manual mappings.
Automation identified many missing RCV mappings

![Bar chart showing number of RCV and auto-only mappings](chart.png)
Paper in progress

• Emphasis:
  • Automating generation of alternative GO views with relevance to pharmacological research.

• Plan to submit to SWAT4LS 2015 with aim of getting a J. Bio-Med Semantics paper.
Work in progress

• Automated generation of RCV mappings (Jenkins)
• System for adding new RCV terms
Acknowledgements

- Roche:
  - Laura Badi
  - Enrico Ponta

- GO:
  - Jane Lomax
  - Melanie Courtot
  - Helen Parkinson