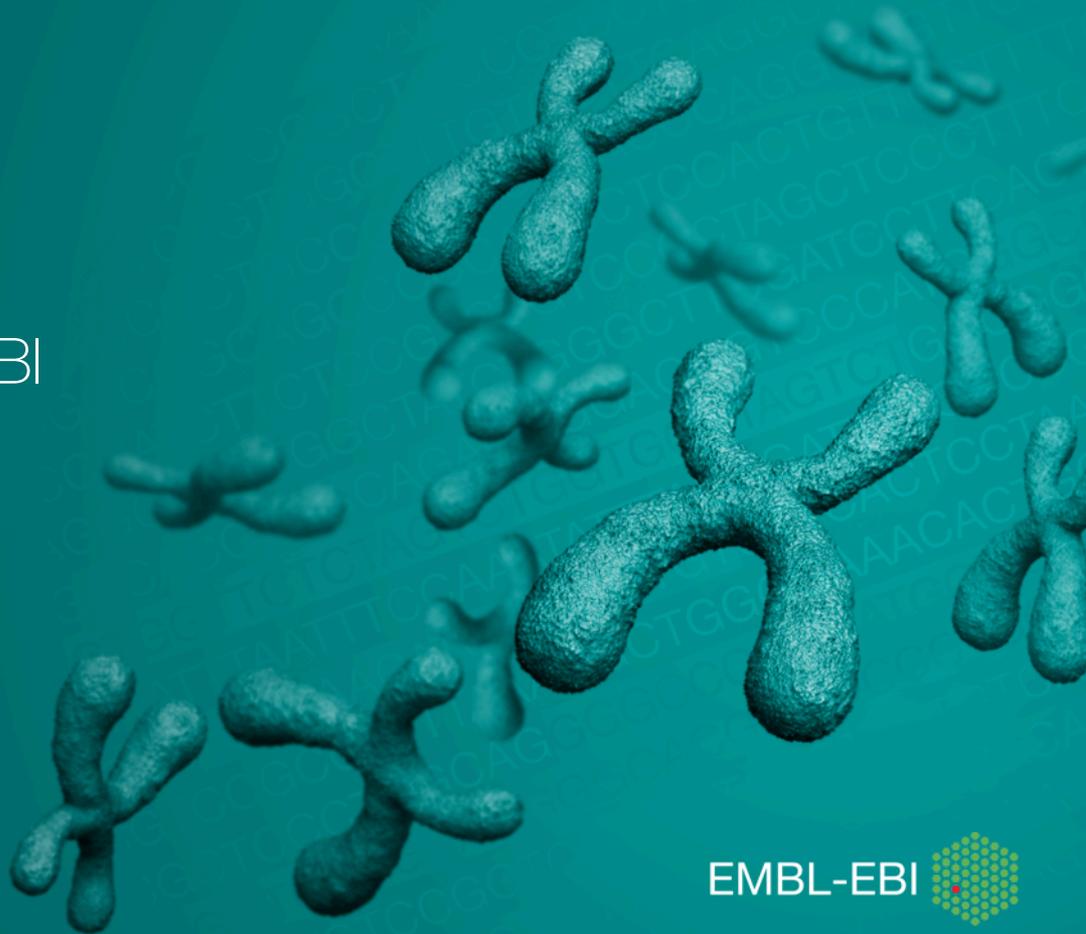


# Roche CV mapping project

Automating alternative views of the GO

David Osumi-Sutherland, EBI



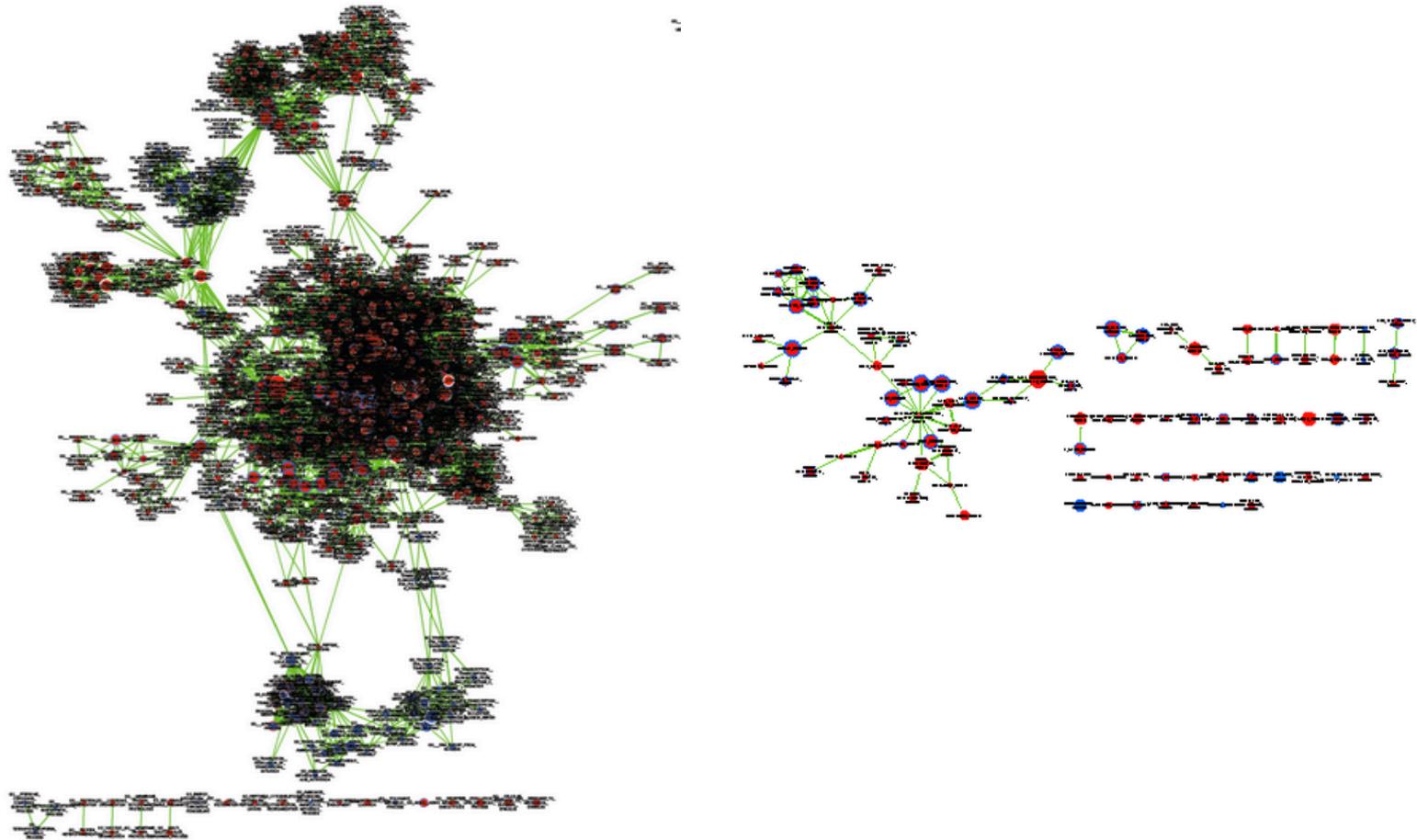
# ROCHE CV

- 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

📖 README.md

## 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.

# GitHub interaction – tsv results files

## **BDNF** pathway RCV\_000002

- Key class: [brain-derived neurotrophic factor receptor signaling pathway](#)
- Pattern: [is\\_a\\_OR\\_part\\_of\\_OR\\_regulates](#)
- map summary: Roche\_cvt: RCV\_000002; class\_expression 1C127FE1-B049-4E09-8DCC-8B323644160F some GO\_0031547; manual\_list\_count 7, generated\_list\_count 8
- [Results](#)

name	ID	manual	auto	checked	blacklisted
brain-derived neurotrophic factor receptor signaling pathway	GO_0031547	1	1	1	0
regulation of brain-derived neurotrophic factor receptor signaling pathway	GO_0031548	1	1	1	0
negative regulation of brain-derived neurotrophic factor-activated receptor activity	GO_0031552	1	1	1	0
positive regulation of brain-derived neurotrophic factor-activated receptor activity	GO_0031553	1	1	1	0
negative regulation of brain-derived neurotrophic factor receptor signaling pathway	GO_0031549	1	1	1	0
positive regulation of brain-derived neurotrophic factor receptor signaling pathway	GO_0031550	1	1	1	0
regulation of brain-derived neurotrophic factor-activated receptor activity	GO_0031551	1	1	1	0
brain-derived neurotrophic factor-activated receptor activity	GO_0060175	0	1	1	0

# GitHub interaction

- Version control
  - Git => very solid versioning system
  - GitHub allows easy user interaction via web

9 lines (9 sloc) | 0.803 kb

Raw

Blame

History



Search this file...

name	ID	manual	auto	checked	blacklisted
brain-derived neurotrophic factor receptor signaling pathway	GO_0031547	1	1	1	0
regulation of brain-derived neurotrophic factor receptor signaling pathway	GO_0031548	1	1	1	0
negative regulation of brain-derived neurotrophic factor-activated receptor activity	GO_0031552	1	1	1	0
positive regulation of brain-derived neurotrophic factor-activated receptor activity	GO_0031553	1	1	1	0
negative regulation of brain-derived neurotrophic factor receptor signaling pathway	GO_0031549	1	1	1	0
positive regulation of brain-derived neurotrophic factor receptor signaling pathway	GO_0031550	1	1	1	0
regulation of brain-derived neurotrophic factor-activated receptor activity	GO_0031551	1	1	1	0
brain-derived neurotrophic factor-activated receptor activity	GO_0060175	0	1	1	0

# Issue tracker

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



**dosumis** commented 11 days ago 

This ticket is intended for discussion of the definition and mapping of BDNF pathway RCV\_000002

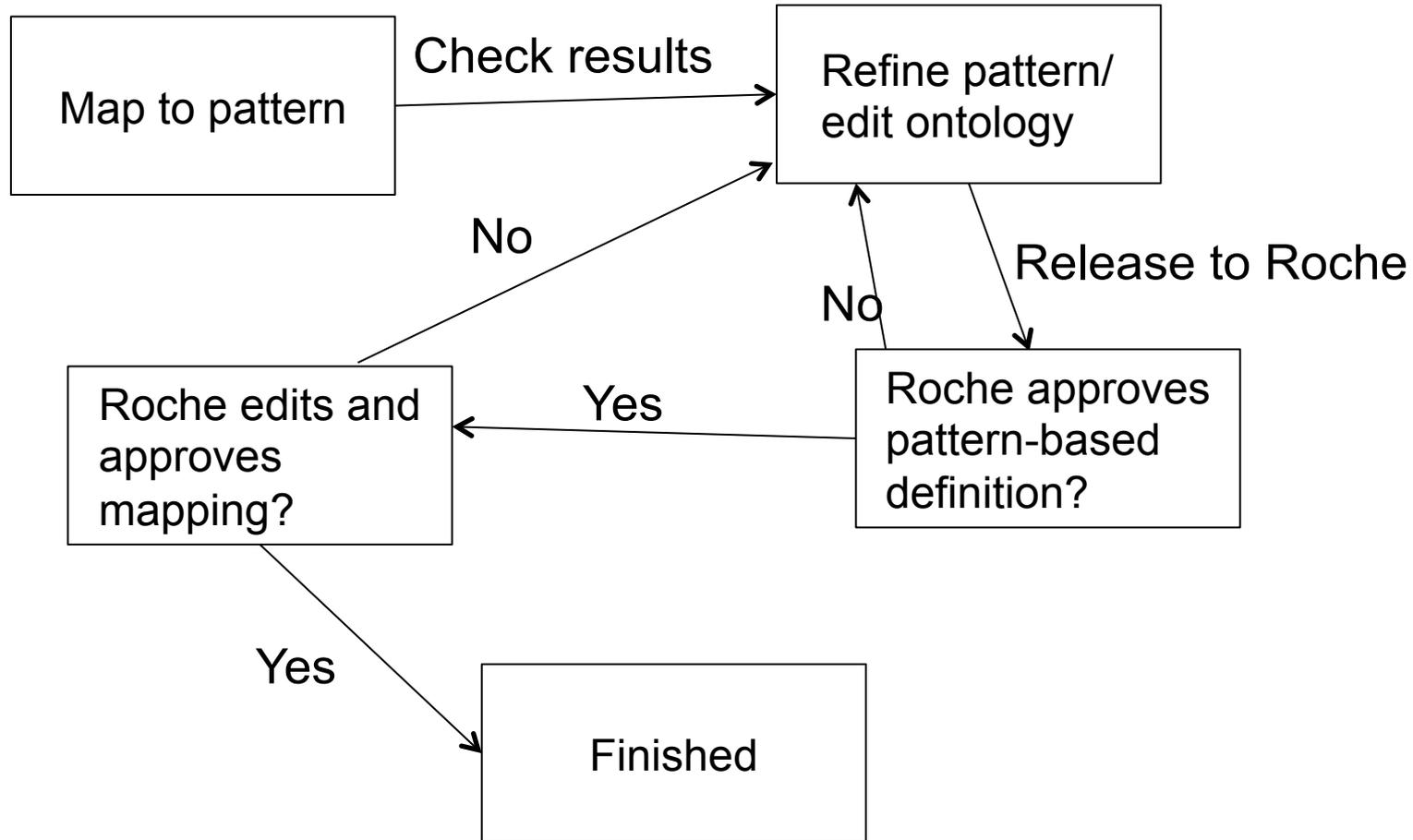
- [Summary of mapping](#)
- [Results table for mapping](#)



**dosumis** commented 11 days ago  

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

# Workflow

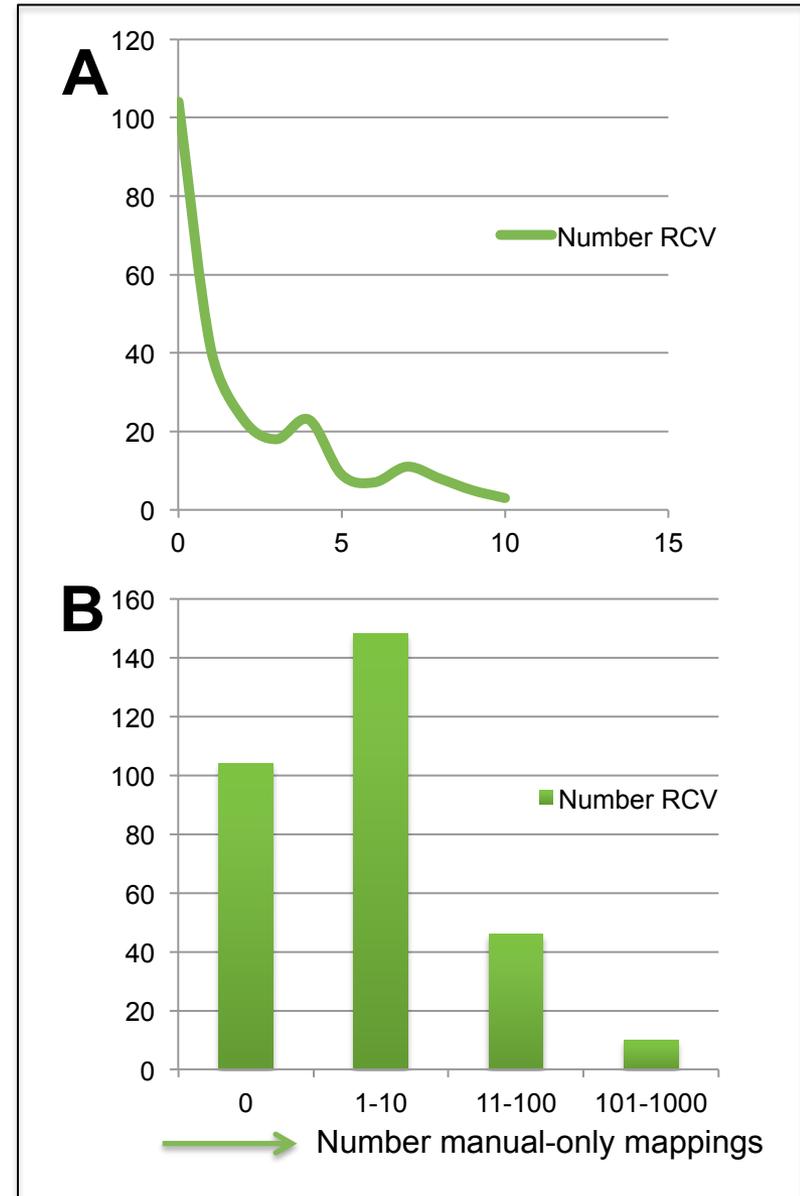


# Results

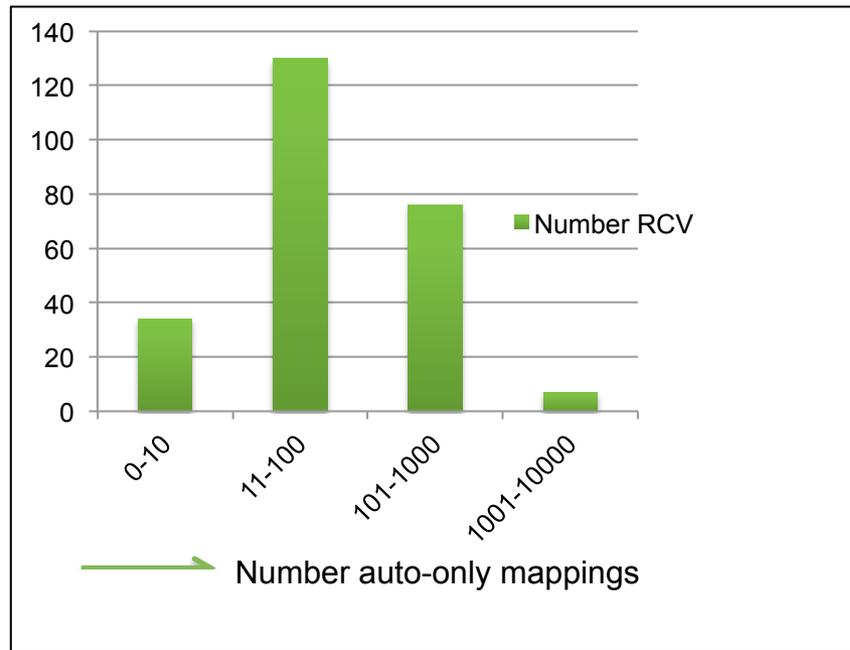
- 308/364 RCV terms:
  - Mapped to OWL queries
  - Results reviewed and accepted by Roche
  - New mapping now in daily usage at Roche
- Many gaps and errors found in GO and corrected

# Manual mappings still required

- Most terms require only small numbers of additional, manual mappings.



# Automation identified many missing RCV mappings



# 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