DARPA Big Mechanism
R3: Reading, Remembering, Revising

PI: Mark Burstein, Burstein@sift.net

SIFT, LLC – Minneapolis, MN

Identification and Significance of Innovation

- Automatically and rapidly build structure from large body of scientific journals using **Deep Natural Language Processing (NLP)**
- read Scientific articles and extract semantic descriptions to augment, revise, or contract with previously represented knowledge of mechanisms involved in cancer-related signaling pathways
- Construct semantic representation at different levels of abstraction
- Reason about connects, ambiguity and reference resolution
- Understand in context – read with respect to previously acquired knowledge/models

Collaborators:

- Larry Hunter (co-PI) – University of Colorado Denver Medicine School
- James Pustejovsky - Brandeis University
- Brent Cochran – Tufts U Medical School
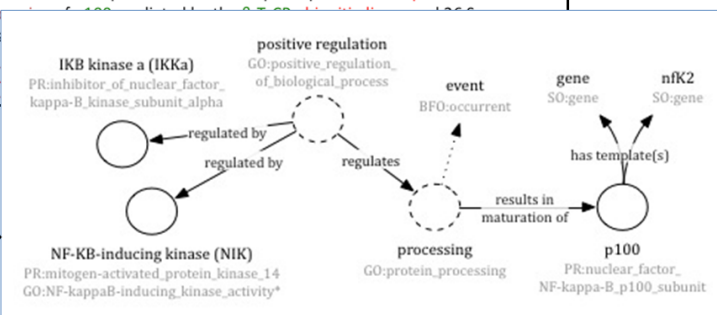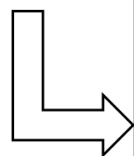- Melissa Haendel, Oregon Health and Sciences University

The processing of the nfkb2 gene product p100 to generate p52 is a regulated event, which is important for the instrumental function of NF-κB. We previously demonstrated that this tightly controlled event is regulated positively by NF-κB-inducing kinase (NIK) and its downstream kinase, IκB kinase α (IKKα). However, the precise mechanisms by which NIK and IKKα induce p100 processing remain unclear. Here, we show that, besides activating IKKα, NIK also serves as a docking molecule recruiting IKKα to p100. This novel function of NIK requires two specific amino acid residues, serine 866 and serine 870, of p100 that are known to be essential for inducible processing of p100. We also show that, after being recruited into p100 complex, activated IKKα phosphorylates specific serines located in both N- and C-terminal regions of p100 (serines 99, 108, 115, 123, and 872). The phosphorylation of these specific serines is the prerequisite for ubiquitination and subsequent proc... ...p100 ... ...26 S proteasome, res...

Ontologies: Ge...

Technical Objectives and Work Plan

Engineer methods to automatically read scientific articles at multiple abstract levels:

- Inferential methods to identify most appropriate terms for a given under-specified NL description
- Introduce abstractions to enable inferences
- No ungrounded terms
- Resolve lexical ambiguity, distribution over conjunctions, and co-references
- Use human subject matter experts to validate generated models

Applications

- Use free from text to build structured ontologies for knowledge discovery in large research area such as oncology research
- identify different types of relationships, areas with many or little contributions, and conflicting scientific findings
- can be applied to a number of scientific and non-scientific knowledge domains

Firm Contact: Linda Holje, SIFT, LLC - Phone: 612-226-5061
319 N 1st St Suite 400 Minneapolis, MN 55401