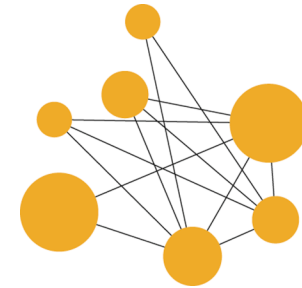
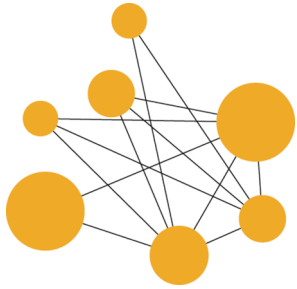


Finding Informative Sentences in Full-text Journal Articles

Zhiyong Lu*, William A. Baumgartner Jr., Gregory Caporaso, K. Bretonnel Cohen, Lawrence Hunter
Computational Bioscience Program
University of Colorado School of Medicine

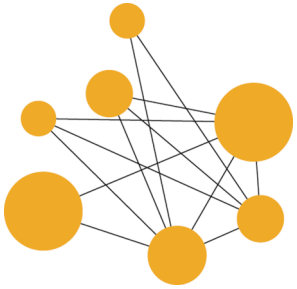


Zhiyong.Lu@uchsc.edu
http://compbio.uchsc.edu/Hunter_lab/Zhiyong



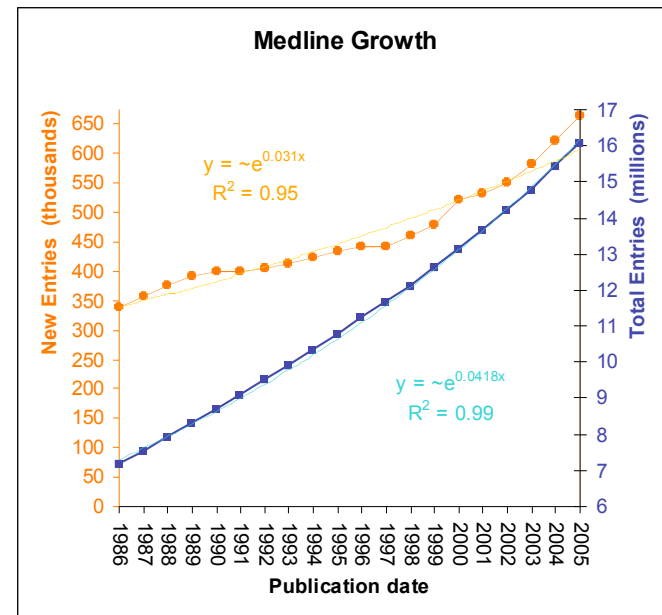
Introduction

- “Informative”: make assertions about a gene’s function
- Examples:
 - Positive: The in vivo interaction between CIPK23 and CBL1 or CBL9 was confirmed using BiFC assays as shown in Figure 6F. [PMID: 16814720]
 - Negative: We do not yet know how these protein complexes activate or inhibit the kaiBC promoter. [PMID: 12441347]



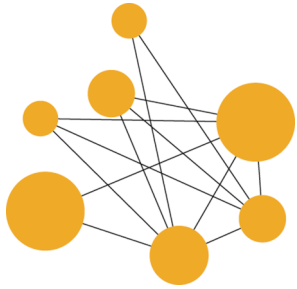
Motivation

- Information Overload
 - Double-exponential growth of peer-review literature
 - Breakdown of disciplinary boundaries



[Hunter and Cohen, Mol Cell. Mar 2006]

- Identifying informative sentences can:
 - Provide a simple mechanism for aggregating gene function information
 - Provide evidence sentences for database annotation
 - Provide basis for generating gene summarizations



Related Work

- Gene References Info Function (GeneRIFs) in the Entrez Gene database

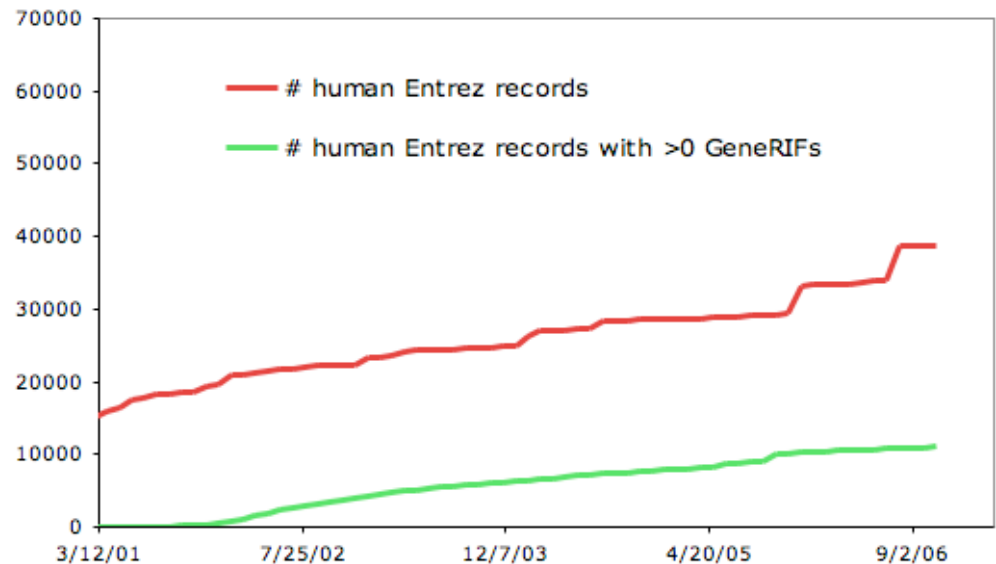
GeneRIFs: Gene References Into Function

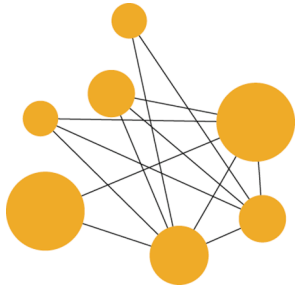
[What's a GeneRIF?](#)



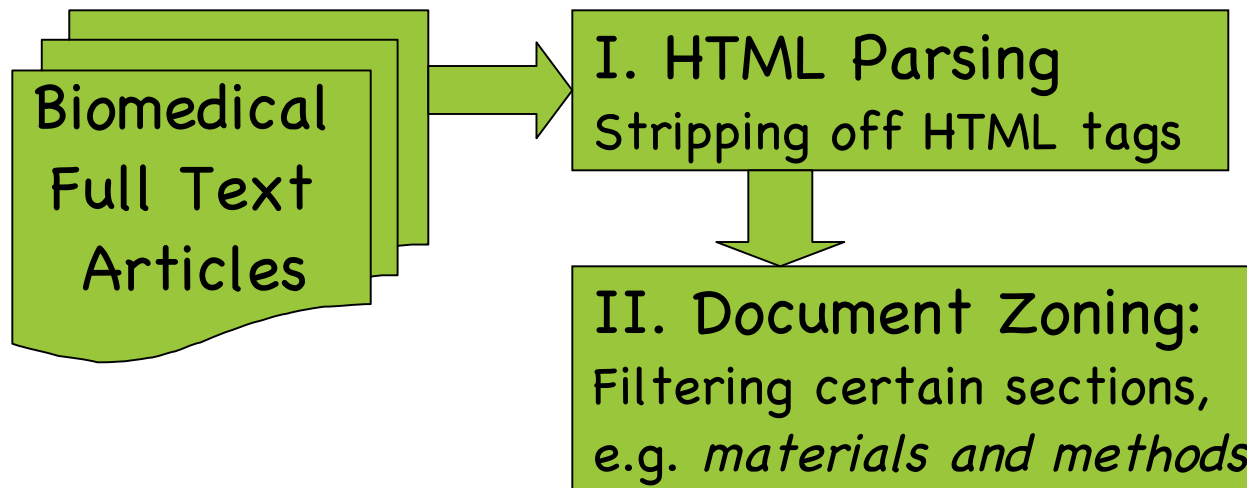
3. Data show that treatment of tumor cell lines with the epidermal growth factor receptor-specific inhibitor cetuximab results in paradox phosphorylation of tyrosine 1173 in the receptor.

- Two Problems
 - Many Entrez genes have no GeneRIFs
 - GeneRIFs were mostly pulled from abstracts rather than the body of the article





System and Method

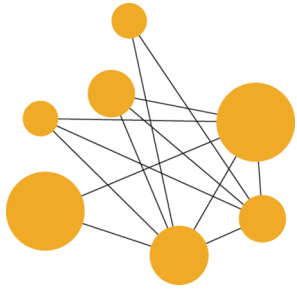


The in vivo interaction between CIPK23 and CBL1 or CBL9 was confirmed using BiFC assays as shown in Figure 6F. [PMID: 16814720]

III. Sentence Selection

Scoring each sentence according to its:

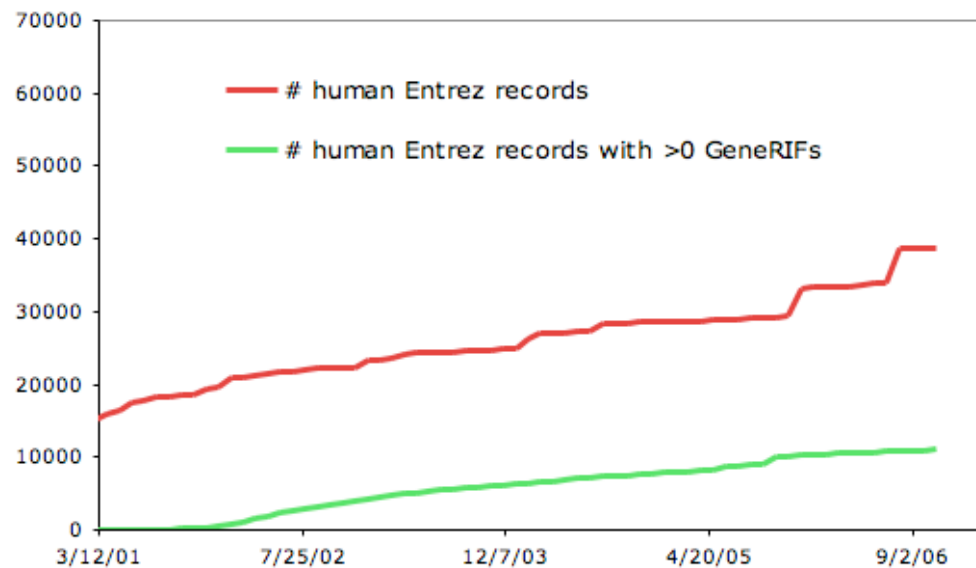
1. keywords of interest [user specific]
2. location
3. mentions of gene/protein names
4. summary-indicative cue words
5. mentions of experimental methods
6. relation with figures/tables



Two Applications

- Finding More GeneRIFs for Entrez Genes (Lu et al., Pac Symp Biocomput, 2006)
 - 20% more accurate than other methods
 - Predicted GeneRIFs for over 8,000 human genes

- Finding Interactions
- An i
- Find data



stein

ing teams

ΛINT