Toward System Biology KnowledgeBase on Transcriptional Regulation in Bacteria.

Pavel S. Novichkov[1] * (PSNovichkov@lbl.gov), Dmitry A. Ravcheev,[2,3] Alexey E. Kazakov,[1] Semen A. Leyn,[2,3] Adam P. Arkin,[1] Inna Dubchak, [1] **Dmitry A. Rodionov**[2,3]* (rodionov@burnham.org)

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [2]Sanford-Burnham Medical Research Institute, La Jolla, CA; [3]Institute for Information Transmission Problems RAS, Moscow, Russia

http://regprecise.lbl.gov

Project Goals:

1. **Develop methods for genome-scale regulon reconstruction utilizing the comparative genomics approach and analysis of RNA and DNA regulatory sites.**
2. **Infer the regulatory interactions and reconstruct transcription regulatory networks (TRNs) in multiple groups of microbial genomes important for DOE mission.**
3. **Integrate the diverse sets of predicted and experimental data on microbial transcriptional regulation into a RegKnowledgeBase.**

Transcriptional regulation of gene expression in response to extracellular and intracellular signals is a key mechanism for successful adaptation of microorganisms to changing environmental conditions. Genes and operons directly co-regulated by the same transcription factor (TF) or by an RNA motif are considered to belong to a *regulon*. All regulons taken together form the transcriptional regulatory network (TRN) of the cell. Availability of complete genomes stimulated wide application of a computational genomics-based approach implemented in the RegPredict Web-server platform for fast and accurate inference of microbial regulons. During the past decade a large number of manually-curated high quality inferences of transcriptional regulons were accumulated for diverse taxonomic groups of bacteria.

We have developed the RegPrecise database (http://regprecise.lbl.gov) for capturing, visualization, and analysis of computationally predicted regulons in microbial genomes. The primary object of the database is a single regulon in a particular genome. Each regulon description contains a regulator, its effector, a set of target genes and operons, and their associated *cis*-regulatory sites. Each TF-operated regulon also has a DNA-binding site model (profile) represented as a nucleotide logo.

Bacterial TRNs are highly flexible in evolution of microbial genomes. The effective large-scale reconstruction of TRNs by comparative genomics requires selection of optimal sets of closely-related genomes. Therefore, the central strategy for regulon analysis in microbial genomes in RegPrecise is based on subdivision of all microbial species into small taxonomic groups that are analyzed independently. The highest level of regulon organization in the database is represented by taxonomic collections of regulons. The current version of RegPrecise contains 13 taxonomic collections of regulons covering major phyla of Bacteria (Proteobacteria, Firmicutes, Cyanobacteria, Actinobacteria etc.).

The total number of TF and RNA-motif regulons in the current release of RegPrecise database exceeds seven thousands. RegPrecise provides three classifications of regulons implemented as controlled vocabularies: (i) biological processes /metabolic pathways; (ii) effectors /environmental signals; (iii) TF protein families. Biological processes attributed to regulons in the database covers a wide spectrum of the cellular metabolism (Fig. 1).The current list of effectors of analyzed TFs includes more than 200 metabolites from the following major classes: amino acids, carbohydrates, nucleotides, lipids and fatty acids, co-enzymes, peptides and antibiotics, secondary metabolites, and inorganic chemicals. Regulons represented in RegPrecise includes ~5400 TFs from >50 TF protein families.

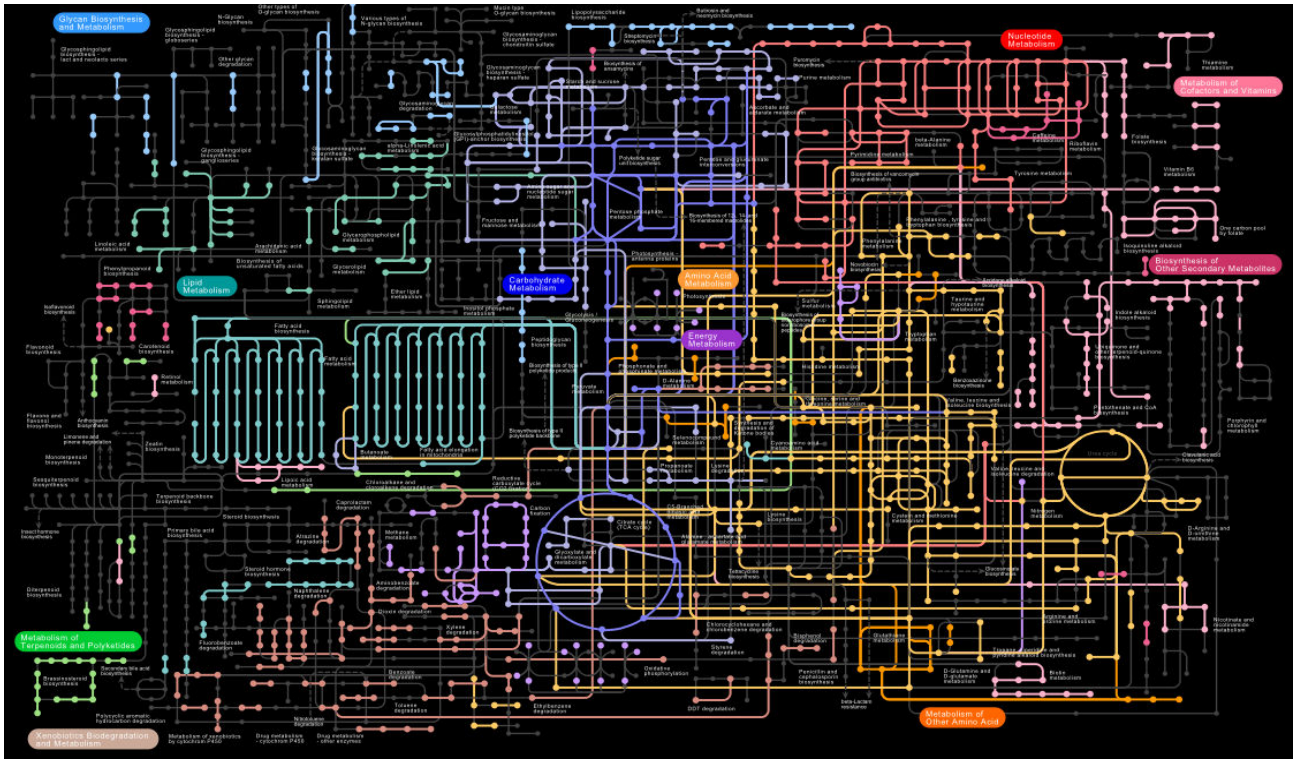

Figure 1. Overview of metabolic pathways covered by reconstructed TF regulons in RegPrecise DB.

The content of RegPrecise database is publically available through the RESTful web-services in JSON format.

In the next release of RegPrecise we will continue extending the regulon content to cover other diverse taxonomic groups. We are also planning to conduct the large-scale assignment of confidence levels to the predicted regulons based on available experimental evidences from literature and external web-resources (EcoCyc, CoryneRegNet, DBTBS, RegTransBase).

RegPrecise is a key component of the upcoming DOE Systems Biology KnowledgeBase. It will provide essential datasets of reference regulons in diverse microbes to enable automatic reconstruction of draft TRNs in newly sequenced genomes.

This research is supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), under contract DE-SC0004999 with Sanford-Burnham Medical Research Institute and Lawrence Berkeley National Lab.