

P704

## Global Regulatory Network in *E.coli* as a Model System for Genome Annotation

Jongsun Jung<sup>PC</sup>, Wonseok Han<sup>1</sup>

*Nanormics Inc., Seoul 136-130*

A series of genome sequencing projects has enabled us to predict operon structures and their regulatory network at the genome level. Here we introduce a new procedure for predicting uncharacterized operons and building a global regulatory network based on the predicted operon structures. It is rare that a pair of adjacent genes are found in a different genome. If that happens more than once, it can be assumed that the pair of genes are a co-occurring unit. A cluster of such co-occurring units is referred to here as an operon. This prediction is mainly based on both orthology and proximity of a pair of genes with the probability of co-occurring on 82 non-redundant microbial genomes. The procedure resulted in 588 operons with  $P(>0.98)$  in *Escherichia coli*. Validation has been made using the known operon database and gene expression database. In this study, 85% (255/297) of the known operons that had been experimentally determined for *Escherichia coli* were predicted partially or completely. Of the rest 333, 33 are likely novel operons from experimental evidences and the others are under investigation. The partially predicted operons refer to the ones with either more or less number of genes than the known ones and their number is 84 and 44 respectively. Among those with more number of genes, 10 cases are also likely novel operons from experimental evidences. A new approach of building the global regulatory network was described. In microbial genomes, a set of related or unrelated operons are regulated by a regulator, a set of regulatory genes or stimulus. Therefore, the operons can be categorized into three different sets according to their functional relations; firstly a set of functionally related operons are regulated by a regulator or a set of regulatory genes. In this case, the operons are in close proximity on a genome. Secondly a set of unrelated operons are regulated by a regulator or a set of regulatory genes and their proximity of operons is more like in random. Finally a set of related or unrelated operons are sometime started by stimulus from cell surface such as PH, concentration, heat/cold shock, light and even from lethal mutation. The three different operon sets are defined as regulon, modulon and stimulon, respectively. Two operons are joined together when they are nearby and found in each of experimental or theoretical databases, such as gene clusters(GC), metabolic pathways(MP), gene expression database(GE) and protein interaction database (PI-a, PI-b). PI-a stands for interacting protein pairs from two adjacent operons and PI-b for interacting protein pairs from two different operons that are separated far away. In this procedure, using the operon structures as a basic unit, GC, MP and PI-a were used for regulon, PI-b for modulon and GE for stimulon. One of noticeable features of genome annotation based on the operon grouping is that the 4,289 genes in *E.coli* can be clustered into countable regulons, modulons and stimulons. The number of grouped operons is small enough to make value-added target selection manually or global planning for biosynthesis and anti-bacterial system. Biological consequence and analysis of the regulatory network will be discussed.