S5-1

Novel Statistical Analysis of DNA Microarray Data

CHOI Jung Kiun
Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea

We have established a method for systematic integration of multiple microarray datasets. The method was applied to two different sets of cancer profiling studies. The change of gene expression in cancer was expressed as 'effect size', a standardiz ed index measuring the magnitude of a treatment or covariate effect. The effect sizes were combined to obtain the estimate of the overall mean. The statistical significance was determined by a permutation test extended to multiple datasets. It was shown t h at the data integration promotes the discovery of small but consistent expression changes with increased sensitivity and reliability. The effect size methods provided the efficient modeling framework for addressing interstudy variation as well. Based on t h e result of homogeneity tests, a fixed effects model was adopted for one set of datasets that had been created in controlled experimental conditions. By contrast, a random effects model was shown to be appropriate for the other set of datasets that had be en published by independent groups. We also developed an alternative modeling procedure based on Bayesian approach, which would offer flexibility and robustness compared to the classical procedure.

S5-2

Biologic System: Not Random, But Contextual

KIM Seungchan
Genomic Signal and Information Processing Unit, Molecular Diagnostics and Target Validation Program, Translational Genomics Research Insitute, USA

Cancer is often understood as disruption of homeostasis, therefore, its expression patterns could be deviated from homeostatic (consistent) pattern, resulting in rather heterogeneous pattern. While this is quite true, it is also believed there still may be a regula tory mechanism, probably different from the regulatory mechanism that keeps homeostasis, intact or modified by mutation, driving systematic transition of cellular status.

As biologic system transition from one cellular state to another, its accompanied b y significantly many changes, most likely the governing regulatory mechanism of cellular system being one of them. This transition could be due to developmental evolution of system, systemic disruption, or adaptation to environmental change. This can be termed as biological contextual change, where context is interpreted as the interrelated conditions in which something exists or occurs. Since the behavior of biologic system is also very much dependent upon contextual change, a different cellular status such as cancers can be understood as different contextual biologic status and they must be treated and understood in different context.

Concurrent measurement of global transcriptional activities enables us to monitor cellular systemic behavior and this flux of information has been and is being utilized to identify molecular markers for various diseases. Were these molecules to have causal relevance to a specific disease, they should be a part of those governed by regulatory mechanism that leads to the disease or maintain its status. Hence, capturing this implicit situational information, i.e. *context,* about the system under study based on observational evidence and identifying those genes with behavior specific to this context is critical step toward th e understanding of information flow among its participants, thus, the discovery of its mechanism.

Therefore, hypothesis is that, within a given *biological context*, there is a set of genes that are tightly coordinated to maintain it. If we were to find ge nes that are responsible for this context or its contextual change, those genes would show a certain level of consistency within the context but some randomness outside the context where no coordination among them is exerted. Based on this assumption, a method to identify those genes with deterministic or very consistent behavior within a context has been developed. Were this contextual information not known a prior, this algorithm can be repeatedly applied to find possible contextual information and its corresponding genes. Applying to two different cancer expression profiles reveals sets of genes with interesting expression patterns, supporting the notion of biological context even with huge disruption of the system such as cancer.

A more biologic information becomes available, we keep seeing the evidence of the contextual dependency of biologic system and itd be critical for us to identify and understand the context under which a biologic system is operated, which would reduce randomness that is seem ingly inherent in the system and unfold its core.