Prevailing Cancer Transcriptional Networks Revealed by Meta-analysis of Cancer Transcriptomes

Atsushi Niida  Seiya Imoto  Masao Nagasaki
aniida@ims.u-tokyo.ac.jp  imoto@ims.u-tokyo.ac.jp  masao@ims.u-tokyo.ac.jp
Rui Yamaguchi  Satoru Miyano
ruiy@ims.u-tokyo.ac.jp  miyano@ims.u-tokyo.ac.jp

Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

Keywords: microarray data; cancer; expression module, transcriptional network

1 Introduction

In the last decade, microarray technology has revealed transcriptomic diversities underlining various cancer phenotypes; on the other hand, we have yet little knowledge about transcriptional programs controlling them. In our previous study, we proposed a computational method termed EEM, which aims to identify expression modules. An expression module is defined as co-expressed genes under a common regulatory program (e.g., target genes of the same TF), and could be a key to understanding the regulatory program. EEM starts from prescribed gene sets defined by prior biological knowledge like TF binding motifs[1]; for each gene set, EEM first identifies the largest co-expressed subset of genes in an input microarray dataset. Using the size of the subset as a test statistic, EEM then obtains significant co-expressed subsets of genes as expression modules.

Recently, the amount of transcriptome data deposited in public databases is exploding. Although even only separate analysis of each dataset has yielded many significant findings, it is expected that meta-analysis of the massive datasets will provide more global insights into regulatory programs encoded in cancer transcriptomes.

2 Method and Results

In this study, we present a new version of the EEM method which enables efficient screening for expression modules, aiming at meta-analysis of a large number of cancer microarray data sets. Although the original version of EEM employs z-score calculation to evaluate coherence of each gene set, we newly introduce an efficient p-value calculation method based on the extreme value distribution[2].

Moreover, we extend our EEM approach to systematic prediction of cancer transcriptional networks. We predicted transcriptional networks, starting from expression modules identified by EEM and TF binding motif associated with them; we constructs a transcriptional network by connecting motifs based on module overlap, and adding regulatory TFs to the motifs based on the TRANSFAC[3] database and expression correlations. PPI and GO terms enriched in each module were also incorporated to predicted networks.

Furthermore, we attempted network-level meta-analysis to reveal prevailing cancer transcriptional networks, which possibly regulate fundamental oncogenic processes commonly employed by various types of cancers. We downloaded 122 microarray datasets associated the “cancer” keyword from GEO and ArrayExpress, and predicted transcriptional networks for each dataset. We then depicted a
“meta-network” by assembling nodes and edges which were repeatedly predicted across the microarray datasets (Figure 1).

The meta-network contains two major sub-networks. One sub-network includes E2F and DP family TFs, NFYA, and their binding motifs. It has been known, and also confirmed by our GO enrichment analysis, that the E2F and DP families are master regulators of cell-cycle; it is very reasonable that they drive one of the principal transcriptional programs in cancer transcriptomes. The other sub-network contains various TF and motif nodes in addition to two main nodes, IRF and PU.1 binding motifs; they apparently act as hubs in this network and might be drug targets in cancer therapy. This sub-network includes known cancer genes like IRF, ETS, and RUNX and is linked to immune-system by GO analysis.

3 Discussion

Our EEM-based network prediction takes as input various types of fragmented knowledge deposited in the databases and systematically assembles them into networks which presumably functions in the transcriptome of an input microarray dataset. Our predicted network includes motif-motif interaction, PPI, GO terms associated with motifs; they make the predicted networks information-rich, highly interpretable and easy to extract biological knowledge. We applied EEM-based network prediction to more than a hundred cancer microarray datasets and, by superimposing networks predicted for each dataset, constructed a “meta-network”. This study is the first meta-analysis to focus on cancer transcriptional networks, and has opened a way to comprehensive understanding of transcriptional networks in cancer cells.

Figure 1: Prevailing cancer transcriptional networks.

References

