TRANSFORMING HIGH-THROUGHPUT OMICSDATA TO BIOMEDICINE

BIN YAN

Ph.D, School of Biomedical Sciences, LKS Faculty of Medicine, the University of Hong Kong

Abstract - Advance of high throughput technologies such as next generation sequenceshasgenerated tens of millions of biological data ("Omics") from bench work. Efficiently processing, analyses and interpretation of these high-throughput information will not only help our understanding of biologicalmechanisms controlling human cancer and diseases, but only provide a basis for transform the derived results to bedside (clinical needs), for example, molecular diagnosis and prognosis, targeted therapeutics or drug development, etc. However, meanwhile, it raises a major challenge for biomedical scientists: development of computational methodologies or software that is able to integrate and interpret these multi-dimensional Omics data effectively. My major interests are the discovery of biological importance from the integrated analysis of big datasets, and facilitation of translational medicine.

I. DISSECTING BIOLOGICAL COMPLEXITY IN GENE REGULATION

Human disease or cancer progression is driven by genetic alterations associated with a large set of gene programs. Altered expression of these genes is controlled by transcription factors (TFs), epigenetic and post-translational regulators in a complex manner. Integrative analyses of high throughput data describing the associated biological processes can help for identification of distinct gene regulatory programs underlying human disease or cancer progression. We have developed several systems biology-based approaches, CMGRN, PTHGRN and Logic TRN to infer gene regulatory networks. Our methods are able to identify regulatory patterns for dissecting biological complex interactions between multilevel regulators and the downstream genes. The resulting discovery would provide an insight into understanding molecular basis of disease or cancer, and further facilitate regenerative medicine. We first developed a web-based framework, post-translational hierarchical gene regulatory network (PTHGRN) to unravel relationships among Post-translational modifications, TFs, epigenetic modifications and gene expression [1]. PTHGRN utilizes a graphical Gaussian model with partial least squares regression-based methodology, and is able to integrate protein-protein interactions, ChIP-seq and gene expression data and to capture essential regulation features behind high-throughput data. It provides an integrative platform for users to analyze ready-to-use public high-throughput Omics resources or upload their own data for systems biology study. Users can choose various parameters in the method, build network topologies of interests and dissect their associations with biological functions. Application of the software to stem cell and breast cancer demonstrates that it is an effective tool for understanding regulatory mechanisms in biological complex systems. PTHGRN web server is publically available at web site http://www.byanbioinfo.org/pthgrn.

Regulation of gene expression is mainly carried out by interactions between multiple TFs and TF binding DNA sequences. To dissect the nature of transcriptional gene regulation, we introduced an integrative method, Logic TRN, to decode TF-TF interactions that form TF logics in regulating target genes [2]. The newly developed method can quantitatively characterize logic relations between TFs by combining cis-regulatory logics and transcriptional kinetics in one single model framework. The derived TF logics are then used to infer the putative TF cooperation in regulating target genes so as to reconstruct transcriptional regulatory networks. We conducted simulation and comparison studies to evaluate the performance of Logic TRN. Then we applied this method to analyze datasets representing the estrogen-induced breast cancer. The derived networks are able to explore the nature of transcriptional gene regulation with biological meanings, which are consistent with previously experiments.

II. TARGETS-SEARCH, A COMPREHENSIVE RESOURCE CAN PRIORITIZE TARGETING CELL SURFACE PROTEINS FOR EXPLORING MULTI-INTERACTIONS AMONG MOLECULES, DRUGS, AND CANCER/DISEASES

Integration and analysis of high content omics data have been critical to the investigation of molecule-interactions (e.gs., DNA-protein, protein-protein, chemical-protein) in biological systems. Human proteomic strategies that provide enriched information on cell surface proteins can also be utilized for re-purposing of drug targets and discovery of disease biomarkers. Although several published resources have proved useful to the analysis of these interactions, our newly developed web-based platform Targets-search has the capability of integrating multiple types of omics data to unraveltheir association with diverse molecule

interactions and disease. Through our platform, we can carry outintegrative and systemic exploitation of potential drug targets among these cell surfaceproteomes, which can further be used to analyze cell surface protein-cell types, protein networks, and biomarkers targeting diseases and cancers. After searching against the multi-dimensional datasets,we identified a set of seven surface proteins showing their linkage with approved-drug targets, markers for detecting embryonic stem cells or cancer/disease cells, and relatively high abundant in protein level. Our analysis demonstrates that surface transporter ABCC4and receptor ERBB2 in particular play a major role in drug repurposing andbiomarkers targeting different types of cells. We then experimentally illustrate this finding by examining embryonic stem cells and found that ABCC4 is preferentially present in these cells and inhibition of this transporter reduced the viability of these cells.

REFERENCE

[1] **Yan B**, Guan D, Wang C, Wang J, He B, Qin J, Boheler KR, Lu A, Zhang G, Zhu H. An integrative method to decode regulatory logics in gene transcription. **Nature Communications** 2017, 8:1044

[2] Guan D, Shao J, Zhao Z, Wang P, Qin J, Deng Y, Boheler KR, Wang J, **Yan B**. PTHGRN: unraveling post-translational hierarchical gene regulatory networks using PPI, ChIP-seq and gene expression data. **Nucleic Acids Res,** 2014, 12:W130-136.

★★★