G&I Genomics & Informatics

eISSN 2234-0742
Genomics Inform 2015;13(3):76-80
http://dx.doi.org/10.5808/GI.2015.13.3.76

ORIGINAL ARTICLE

The Construction of Regulatory Network for Insulin-Mediated Genes by Integrating Methods Based on Transcription Factor Binding Motifs and Gene Expression Variations

Hyeim Jung[#], Seonggyun Han[#], Sangsoo Kim*

Department of Bioinformatics and Life Science, Soongsil University, Seoul 06978, Korea

Type 2 diabetes mellitus is a complex metabolic disorder associated with multiple genetic, developmental and environmental factors. The recent advances in gene expression microarray technologies as well as network-based analysis methodologies provide groundbreaking opportunities to study type 2 diabetes mellitus. In the present study, we used previously published gene expression microarray datasets of human skeletal muscle samples collected from 20 insulin sensitive individuals before and after insulin treatment in order to construct insulin-mediated regulatory network. Based on a motif discovery method implemented by iRegulon, a Cytoscape app, we identified 25 candidate regulons, motifs of which were enriched among the promoters of 478 up-regulated genes and 82 down-regulated genes. We then looked for a hierarchical network of the candidate regulators, in such a way that the conditional combination of their expression changes may explain those of their target genes. Using Genomica, a software tool for regulatory network construction, we obtained a hierarchical network of eight regulons that were used to map insulin downstream signaling network. Taken together, the results illustrate the benefits of combining completely different methods such as motif-based regulatory factor discovery and expression level-based construction of regulatory network of their target genes in understanding insulin induced biological processes and signaling pathways.

Keywords: network, regulation, regulator, sequence motif

Introduction

Insulin resistance is a condition in which tissues, such as skeletal muscle, liver, and adipocytes, fail to respond to insulin, leading to type 2 diabetes mellitus. Because skeletal muscle is the primary tissue of insulin-mediated glucose uptake [1], it would be important to study the insulin effects on gene expression in skeletal muscle and to identify the insulin-responsive target genes and regulatory genes.

DNA microarray techniques have been widely used in biological research by enabling comprehensive understanding of molecular and systematic mechanism [2, 3]. In addition, microarray experiments enable researchers to identify gene sets undergoing a response to a specific stimulus as well as to analyze dynamic response at the level of gene regulatory network [4]. Especially, microarray experiments could be designed to study downstream effects of a specific stimulus or diverse stimuli on gene expression at the level of genomic scale, not a single genetic one [5]. Microarray techniques could be a power tool to discover regulatory factors as most of co-expressed genes are co-regulated by the same regulatory factors [6]. Moreover, because genes co-regulated by the same transcription factor (TF) commonly share binding sites for this TF [6], it would be worth to detect TF based on motif discovery method.

Although diverse motif discovery methods have been proposed and refined [7-10], most of them are restricted to using human annotated position weight matrices (PWM). On the other hand, iRegulon, a user-friendly Cytoscape

Received July 31, 2015; Revised September 15, 2015; Accepted September 21, 2015

***Corresponding author:** Tel: +82-2-820-0457, Fax: +82-2-824-4383, E-mail: sskimb@ssu.ac.kr
[#]These authors contributed equally to this work.

plugin, provides approximately nine thousand PWMs that have been collected from various sources and from different species and enable to make regulatory network by linking them to candidate TF [11]. The other approach for pathway analysis is Genomica [12], which yields regulatory modules from gene expression based on probabilistic graphical models and Bayesian networks. It has a benefit in directly mirroring expression of their target genes in a specific condition by comparing expressions between regulators and their target genes [12]. However, this approach requires looping over all regulators, taking too much time in big data.

Here, we tried to construct a regulatory network in order to understand the regulation of insulin-response genes. We identified about 560 genes differentially expressed between before and after treatment in insulin sensitive individuals, and found that those genes act for insulin signaling biological process through Gene Ontology (GO) analysis. Furthermore, we report a reliable regulatory network controlling insulin-mediated genes by integrating motif-based approach using iRegulon and expression-based approach using Genomica; thereby present some benefits for the construction.

Methods

Study samples

We downloaded and re-analyzed a previous published mRNA expression microarray data set measured on Affymetrix Human Genome U95A Array (Affymetrix, Santa Clara, CA, USA) from the Gene Expression Omnibus (GEO) database (GSE22309) [5]. This data was measured in human skeletal muscle cell of human individuals: 20 insulin sensitive, 20 insulin resistant, and 15 diabetic patients for before and after insulin treatment. We used only the insulin sensitive group. Total of 12,626 probes were monitored, and we considered genes which are protein-coding gene.

Insulin-mediated genes

To select differential genes before and after insulin treatment in 20 insulin sensitive individuals, we carried out paired t test and calculated gene expression fold change. Using local commands of R program, we selected insulin-responsive genes which were significantly dysregulated at p-value (paired t test) < 0.01 (corresponding to FDR [13] ≤ 0.05) and fold change > 1.2 (Fig. 1). We looked for enriched GO terms of the insulin responsive genes using DAVID Functional Annotation Tool [14].

Search for candidate TFs based on regulatory motif and chromatin immunoprecipitation-sequencing

The regulatory factors for insulin-responsive genes were searched using the iRegulon app in Cytoscapev3.2.1. iRegulon detects the TFs and their targets by scanning known TF-binding promoter motifs as well as the predicted motifs discovered from the Encyclopedia of DNA Elements (ENCODE) Project chromatin immunoprecipitation-sequencing data. We set 20-kb upstream for the options "Putative regulatory region," "Motif rankings database," and "Track rankings database." Other options were taken as default. We executed iRegulon and looked for TFs for each down and up regulated genes.

Construction of regulatory networks based on expression value

With TFs searched by iRegulon, we tried to construct regulatory network modules which can explain situation of expressions of target genes based on expression level. This construction was carried out using Genomica [12], which is an analysis and visualization tool for genomic data. Among functions of Genomica, we used the "Create a Module Network." It identifies regulatory network modules from gene expression data based on probabilistic graphical models and Bayesian networks. The options "Max number of modules" and "Min experiments per context" were set at 4.

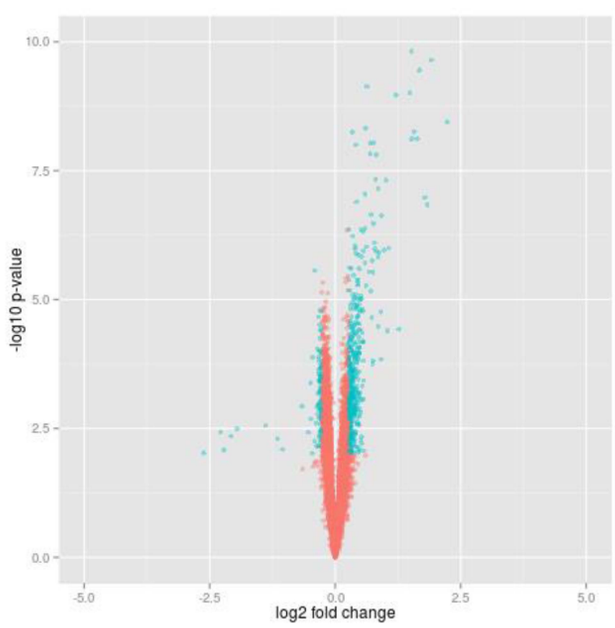

Fig. 1. The volcano plot showing the transcriptional changes between before and after insulin treatment. The $-\log_{10}$(p-value) of paired t test and $\log_2$ gene expression fold changes between before and after the treatment in insulin sensitive individuals were plotted on the x- and y-axis, respectively. The genes colored by cyan are selected using the threshold of t test $-\log_{10}$(p-value) and $\log_2$(FC) of fold change [$-\log_{10}$(p-value) < 2 (corresponding to FDR[19] ≤ 0.05) and $\log_2$(FC) > 0.2630344]. FC, fold change.

Table 1. The enriched GO terms of insulin-responsive genes

GO term[a]	p-value	p_{adj}-value
Negative regulation of biosynthetic process	4.00E-09	1.80E-06
Negative regulation of macromolecule metabolic process	2.90E-09	2.20E-06
Negative regulation of RNA metabolic process	5.80E-07	1.20E-04
Regulation of transcription from RNA polymerase II promoter	1.60E-06	3.00E-04
Enzyme linked receptor protein signaling pathway	5.40E-06	8.70E-04
Positive regulation of biosynthetic process	6.40E-06	9.60E-04
Positive regulation of RNA metabolic process	1.00E-05	1.20E-03
Positive regulation of macromolecule metabolic process	2.50E-05	2.30E-03
Regulation of transcription, DNA-dependent	4.40E-05	3.70E-03
Response to endogenous stimulus	1.10E-04	8.30E-03
Response to hormone stimulus	1.40E-04	1.00E-02
Transmembrane receptor protein tyrosine kinase signaling pathway	1.60E-04	1.20E-02
Regulation of cell proliferation	4.60E-04	3.10E-02

GO, Gene Ontology.
[a]The biological process terms of GO database.

Results

Insulin-targeted gene set

The expression data of 20 insulin sensitive individuals downloaded from GEO (GSE22309) were analyzed to find genes controlled by insulin. In addition, the data contain expression values of transcripts before and after insulin treatment for each sample. Among the total of 12,626 probes, we selected about 9,000 genes which are protein-coding gene. With expression of each gene, we carried out paired t test between before and after insulin treatment and calculated fold change. A total of 560 genes were identified by cutoff using t test p-value ($p_{adj} < 0.05$) and fold change ($FC > 1.2$) in insulin sensitive group (Fig. 1). Furthermore, we looked for enriched GO terms using DAVID functional annotation tool, and several biological process terms of the GO database were searched ($p_{adj} < 0.05$). As the resulting list of terms is redundant, the lists were prunedmanually (Table 1). Most of the terms were compatible with insulin response and overlapped to GO terms which have been known from previous studies [5].

TF discovery based on motif using iRegulon

We tried to discover the regulatory TFs for the insulin-treated gene set using iRegulon in Cytoscape. For each down- and up-regulated gene set, we predicted 25 TFs of the target genes based on motif in upstream 20 kb of the target genes. As shown in Fig. 2, target genes are linked to multiple TFs. A single TF targeted about 138 genes in average.

Construction of regulatory network using Genomica

We tried to explain the expression variations of target genes with those of TFs identified by iRegulon. Firstly, we checked correlation between expression of YY1, which is a predicted TF by iRegulon and known for an insulin signaling gene, and mean of those of target genes. However, the correlation coefficient is rather low (Fig. 3). This prompted us to construct the TF network module to explain the expression variations of target genes using Genomica (Fig. 4). We found most TFs participated in regulatory network constructed by Genomica have been known to have relation-

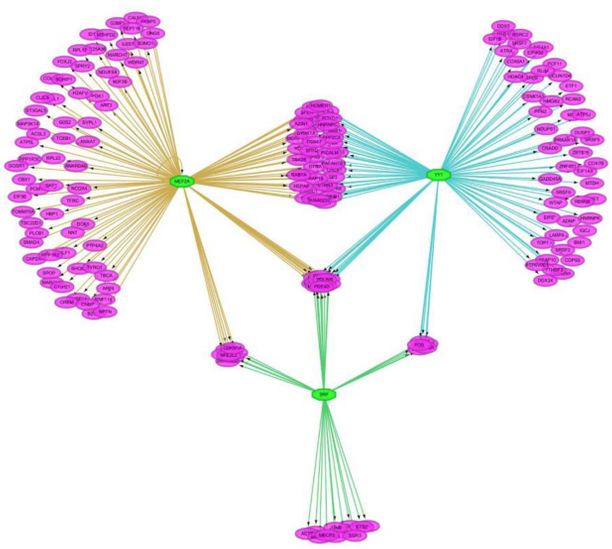

Fig. 2. A representative regulatory network using iRegulon in Cytoscape. The node with purple and green represent the up-regulated genes after insulin treatment and their predicted regulators (MEF2A, YY1, and SRF), respectively. Genes linked to two or three regulators may be regulated by those multiple regulators. The target genes for each transcription factor are portrayed by different edge colors.

ships with insulin from previous studies [15-19]. While iRegulon provides a list of potential regulatory TFs ignorant of their expression conditions, Genomica has the power to filter the list by constructing conditional hierarchical regulatory network where the expression variations of the target genes are explained by the expression conditions of the filtered TFs. We describe a more reliable regulatory network controlling insulin-responsive genes by integrating completely different methods for constructing regulatory network.

Discussion

Most genes have networks of interactions for appropriate biological functions. Furthermore, regulators do not act alone. Describing a combination of regulatory genes is important to more accurately explain their target genes and understand biological functions. In this regard, construction of regulatory networks can give an insight in the biological interpretation of differential gene set between normal and complex diseases. In order to understand insulin-responsive system, we tried to construct regulatory network with microarray mRNA expression data monitored before and after insulin treatment in skeletal muscle of insulin sensitive samples. For the construction of a reliable network, we carried out integration of two different methods, one based on TF-binding motif and the other on expression variations. Firstly, we defined insulin-mediated gene set using threshold of t test p-value and expression fold change, about 560 genes among total of 12,626 genes were identified. GO enrichment analysis indicated that the gene set is enriched with the biological process terms related with insulin signaling pathway. Especially, among those GO terms, "response to hormone stimulus" and "trans membrane receptor protein tyrosine kinase signaling pathway" have well known links directly to the insulin [8]. To understand regulation of those genes made up of the insulin signaling pathway, 25 TFs were predicted by iRegulon, based on the enrichment of TF-binding motif. iRegulon can infer TFs of target genes, and visualize interactions between TFs and target genes with previously known evidences that what TFs bind to a target gene with sequence information. In this case, expression values of target genes are not considered, and this regulatory network cannot explain how those regulators

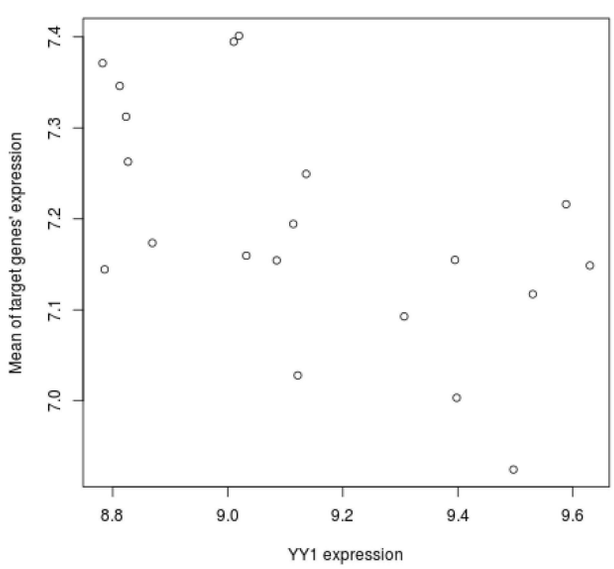

Fig. 3. The expression variations of a regulator YY1 and its target genes. For each individuals, the expressions of the up-regulated target genes by insulin treatment were averaged and plotted on y-axis, and those of YY1 were plotted on x-axis. The expressions of YY1 do not resemble those of their target genes (correlation coefficient = −0.5878721).

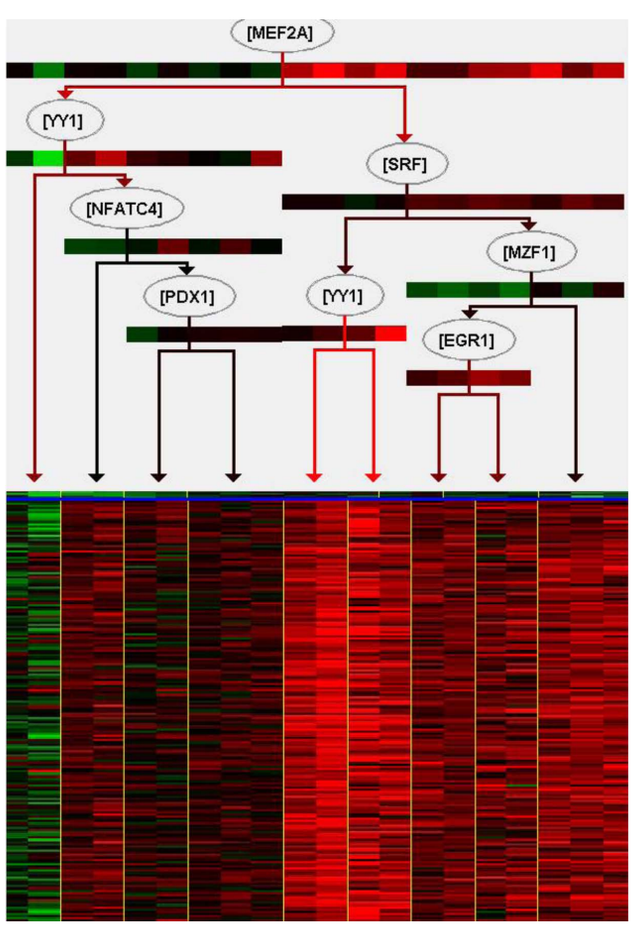

Fig. 4. The regulatory network, upper section is a dendrogram for regulators' expression and bottom section is a heat map for target genes' expression, generated using Genomica, based on expression profiling. The upper part of the figure represents a dendrogram for a combination of regulators and their expressions for each individual. The lower part shows a heat map for the expression profiling of up-regulated target genes by insulin treatment (row and column indicate genes and individuals, respectively).

affect expression variation of target genes. We tried to explain expression variation of target genes with those of TFs identified by iRegulon. However, the expression variation of a TF did not resemble those of target genes. It suggests that expression variations of multiple TFs should be considered collectively to understand those of target genes. With the TFs predicted by iRegulon, we constructed a hierarchical regulatory network among them using Genomica, explaining the expression variation of the target genes. As shown Fig. 4, a combination of multiple TFs instead of a single TF explained the expression variations of target genes. For example, MEF2A which is at the top of the regulatory TF network had the largest influence on the expression variations of the target genes, and YY1, the second highest ranking TF, did not influence as much as MEF2A. However, YY1 affects to increase expression of target genes when MEF2A has low expression value. In fact, MEF2A, YY1, SRF, PDX1, and NFATC4 are known for having relationships with insulin [15-19]. Although expression of downstream TFs in the regulatory network is not perfectly fitted to that of their target genes, the construction based on expression data helps to understand more clearly regulatory network of target genes. Altogether, in this paper, we presented a reliable regulatory network controlling insulin-mediated genes as well as the approach integrating two different methods, iRegulon and Genomica, may allow making a reliable regulatory network.

Acknowledgments

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01167402)", Rural Development Administration, Republic of Korea.

References

1. Cline GW, Petersen KF, Krssak M, Shen J, Hundal RS, Trajanoski Z, *et al*. Impaired glucose transport as a cause of decreased insulin-stimulated muscle glycogen synthesis in type 2 diabetes. *N Engl J Med* 1999;341:240-246.
2. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, *et al*. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999;23:41-46.
3. Krylov AS, Zasedateleva OA, Prokopenko DV, Rouviere-Yaniv J, Mirzabekov AD. Massive parallel analysis of the binding specificity of histone-like protein HU to single- and double-stranded DNA with generic oligodeoxyribonucleotide microchips. *Nucleic Acids Res* 2001;29:2654-2660.
4. Mutarelli M, Cicatiello L, Ferraro L, Grober OM, Ravo M, Facchiano AM, *et al*. Time-course analysis of genome-wide gene expression data from hormone-responsive human breast cancer cells. *BMC Bioinformatics* 2008;9 Suppl 2:S12.
5. Wu X, Wang J, Cui X, Maianu L, Rhees B, Rosinski J, *et al*. The effect of insulin on expression of genes and biochemical pathways in human skeletal muscle. *Endocrine* 2007;31:5-17.
6. Do JH, Choi DK. Clustering approaches to identifying gene expression patterns from DNA microarray data. *Mol Cells* 2008;25:279-288.
7. Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* 2004;32:1372-1381.
8. Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* 2004;5:276-287.
9. Aerts S. Computational strategies for the genome-wide identification of cis-regulatory elements and transcriptional targets. *Curr Top Dev Biol* 2012;98:121-145.
10. Bulyk ML. Computational prediction of transcription-factor binding site locations. *Genome Biol* 2003;5:201.
11. Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, *et al*. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol* 2014;10:e1003731.
12. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, *et al*. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 2003;34:166-176.
13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol* 1995;57:289-300.
14. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44-57.
15. Mora S, Pessin JE. The MEF2A isoform is required for striated muscle-specific expression of the insulin-responsive GLUT4 glucose transporter. *J Biol Chem* 2000;275:16323-16328.
16. Verdeguer F, Blattler SM, Cunningham JT, Hall JA, Chim H, Puigserver P. Decreased genetic dosage of hepatic Yin Yang 1 causes diabetic-like symptoms. *Mol Endocrinol* 2014;28:308-316.
17. Jin W, Goldfine AB, Boes T, Henry RR, Ciaraldi TP, Kim EY, *et al*. Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. *J Clin Invest* 2011;121:918-929.
18. Stoffel M, Stein R, Wright CV, Espinosa R 3rd, Le Beau MM, Bell GI. Localization of human homeodomain transcription factor insulin promoter factor 1 (IPF1) to chromosome band 13q12.1. *Genomics* 1995;28:125-126.
19. Yang TT, Suk HY, Yang X, Olabisi O, Yu RY, Durand J, *et al*. Role of transcription factor NFAT in glucose and insulin homeostasis. *Mol Cell Biol* 2006;26:7372-7387.