Network Clustering along Diabetes Progression in Three Tissues of Goto-Kakizaki Rats

Xinrong Zhou
Tongji Hospital
Huazhong University of Science and Technology
Wuhan, Hubei 430030, China
zhouxinrong8@hotmail.com

Shigeru Saito
Division of Chem&Bio Informatics
INFOCOM Corporation
Tokyo, 150-0001, Japan
sh.saito@infocom.co.jp

Katsuhisa Horimoto
Computational Biology Research Center
National Institute of AIST
Tokyo 135-0064, Japan
k.horimoto@aist.go.jp

Luonan Chen
Key Laboratory of Systems Biology
Chinese Academy of Sciences
Shanghai 200233, China
lnchen@sibs.ac.cn

Huarong Zhou*
SIBS-Novo Nordisk Translational Research Centre for PreDiabetes
Chinese Academy of Sciences
Shanghai 200233, China
hrzhou@sibs.ac.cn

*To whom correspondence should be addressed

***Abstract*—We investigated the macroscopic changes in the regulatory coordination of diabetes progression during three periods in three tissues, adipose, liver and muscle, of Goto-Kakizaki (GK) rats. For this purpose, we performed network clustering by the Newman algorithm for the regulatory networks inferred by a modified path consistency algorithm, and investigated the biological functions of each cluster by an enrichment analysis of the constituent genes. We then compared the network clusters characterized by biological functions with the diabetes progression of GK rats in each of the three tissues. The network structure, the number of clusters, and the number of clusters characterized by biological functions during the three periods showed similar patterns in the three tissues. In contrast, further scrutiny of the biological functions at coordinated clusters revealed characteristic differences between the three tissues along the diabetes progression. In particular, the hypothetical roles of each tissue emerged: adipose and liver function at the cellular and molecular levels at the early stage, respectively, and all three tissues are responsible for diabetes progression, under the control of various transcriptional regulators.**

Keywords-type 2 diabetes progression; tissue relationship; network clustering; enrichment analysis

I. INTRODUCTION

Diabetes is a major healthcare problem, because the disease increases the risk of heart disease, stroke, and microvascular complications, such as blindness, renal failure, and peripheral neuropathy. At the physiological level, defective insulin secretion and resistance emerges as the main cause of diabetes, with associated environmental and lifestyle factors, such as obesity [1]. During the past decade, the genetic bases of various monogenic forms of the disease and their underlying molecular mechanisms have been elucidated, and many genes that increase the risk of type 2 diabetes have also been identified [2].

One way to bridge the gap between the physiological and molecular levels of diabetes is to investigate the coordination of gene expression changes along with type 2 diabetes progression in insulin-responsive tissues in obese animals. Indeed, some results for the coordination mode at the molecular level have been reported. For example, based on the identification of gene groups that undergo expression changes in a highly correlated fashion from six insulin-responsive tissues, the linkage of a cell cycle regulatory gene group in pancreatic islets to diabetes susceptibility [3], and network analyses of expression profiles from liver, revealed several candidates of master regulators from the active regulatory networks along the diabetes progression pathway [4]. However, the roles of the insulin-responsive tissues have not still been characterized clearly, in terms of their macroscopic coordination along diabetes progression.

Here, we analyzed expression profile data measured at three distinctive periods from three tissues, adipose, liver and muscle, in diabetic Goto-Kakizaki (GK) and non-diabetic

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

Wister-Kyoto (WKY) rats [5-7], and investigated the macroscopic changes in the regulatory coordination along diabetes progression between the three tissues, with special focus on the changes in the regulatory gene groups. For this purpose, we first inferred the regulatory networks by a statistical network analysis method, to investigate the relationship between all of the genes at each period in each tissue. To investigate the relationship between the gene regulatory groups, we next performed the network clustering of the inferred networks of the complex structure. To focus on the coordinated regulatory relationships in different tissues along diabetes progression, we selected the connected clusters of many genes, and investigated the biological functions of each cluster by an enrichment analysis of the constituent genes. Finally, the clusters in which the biological functions overlapped were scrutinized, to reveal the similarities and differences of the functional roles in the three tissues along with diabetes progression.

II. MATERIALS AND METHODS

A. Data Analyzed in This Study

We analyzed three sets of gene expression data, measured in adipose [5], liver [6], and muscle [7] tissues from GK and WKY rats, which were archived in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/projects/geo/) database (GSE13269, GSE 13271, and GSE13270). The three data sets comprised 31,099 probes, measured using the Affymetrix Microarray Suite 5.0 (Affymetrix), and were further converted into 14,506 genes, for five samples of male spontaneously diabetic GK rats and WKY controls at each of five time points (4, 8, 12, 16, and 20 weeks of age). In this analysis, the five analysis periods were reclassified into three periods: period of 4w, period of 8w and 12w, and period of 16w and 20w. In this reclassification, the periods are composed of 5, 10, and 10 samples which are useful for more robust statistical analysis, and represent pre-diabetes, initial diabetes and advanced diabetes states, respectively. Indeed, this enhanced breakdown was employed because diabetes progression is generally discussed from the simple dichotomy between non-diabetes and diabetes, considering the dynamic ranges of gene expression measurements and the small number of animal samples per period, as reported in a previous diabetes study [3].

B. Overview of the Present Method

An overview of our method is described, as follows. First, the relationships between the genes are basically investigated by a network inference method, the path consistency algorithm [8]. Second, the inferred network structures are divided into groups by a network clustering method, the Newman algorithm [9]. Third, the biological functions are extracted in each cluster by a standard method, an enrichment analysis. Thus, the coordination of the biological functions of the gene groups hidden in the gene expression profiles is revealed visually and automatically, by the network analysis methods and the following functional analyses. The details of each step are described below.

C. Network Inference by Path Consistency (PC) Algorithm with Modifications

The path consistency (PC) algorithm is a method for network inference based on the graphical model [8]. The original PC algorithm is composed of two parts: the undirected graph inference by the partial correlation coefficient and the following directed graph by using the orientation rule. The present method partially exploits the first part of the PC algorithm, because the aim of the present application of the network inference method is to scrutinize the existence of the relationships between the genes, without the influence of causality.

The first part of the algorithm is straightforward. The relationship between two variables is tested from the lower partial correlation coefficient to the higher one. For example, the relationship between the two variables is first tested by the zero-th partial correlation coefficient. If the null hypothesis is accepted; i.e., no association between the two variables, then further tests are not performed for the higher order partial correlation coefficients. If it is rejected, then the relationship between the two variables is tested by the first partial correlation coefficient. In general, the (m-2)-th order of the partial correlation coefficient is calculated between two variables, given (m-2) variables; i.e., $r_{ij,rest}$, between X_i and X_j, given the '*rest*' of the variables, $\{X_k\}$ for k=1, 2, ⋯, m, and $k \neq i, j$, and after calculating the (m-2)-th order of the partial correlation coefficient, the algorithm naturally stops. However, the algorithm does not usually request the (m-2)-th order of the partial correlation coefficient to naturally stop. This is because no adjacent variables will be found after excluding the variables, even in the calculation of the lower order of the partial correlation coefficient. Subsequently, the computational time depends on the number of excluded variables during the calculation process.

In the sample data, the zeroth-order of the correlation coefficient is calculated by Pearson's correlation coefficient, r_{ij}, expressed by

$$r_{ij} = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{var}(X_i)\,\mathrm{var}(X_j)}}, \tag{1}$$

where cov(X_i, X_j), and var(X_i) are the covariance between X_i and X_j, and the variance of X_i. The higher order of the correlation coefficients is the partial correlation coefficient, $r_{kl,rest}$, expressed by

$$r_{kl,rest} = \frac{-r^{kl}}{\sqrt{r^{kk} \cdot r^{ll}}}, \tag{2}$$

where (kl, *rest*) means {1,2, …, p}\{k,l}, and r_{kl} is the k-l element of the inverse correlation coefficient matrix. Note that the dimensions of the correlation coefficient matrix correspond to the orders of the partial correlation coefficients. The n-th-order partial correlation coefficient is calculated from the (n+2) dimension of the correlation coefficient matrix. The partial correlation coefficient is statistically tested by

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

using the *Z*-statistic. First, the *z*-transforms of the partial correlation coefficients are calculated, by the following equation:

$$z_{kl} = \frac{1}{2}\ln\left(\frac{1+r_{kl}}{1-r_{kl}}\right), \quad (3)$$

where r_{kl} is the absolute value of the partial correlation coefficient. Then, the *z*-statistic is obtained from the following equation:

$$z = z_{kl}\sqrt{n-3-p}, \quad (4)$$

where n is the number of samples and p is the conditioning order of the partial correlation coefficient. The *z*-statistic follows the standard normal distribution, $N(0,1)$, and the significance probability can be set according to this distribution.

The key point in the present network inference is two modifications of the original PC algorithm, for its application to actual data, such as gene expression profiles [10,11]. The first modification is the correction of the algorithm in the calculation of the partial correlation coefficient. Since many genes frequently show very similar expression profiles, the difficulty increases in the numerical calculation of the partial correlation coefficients, due to the multi-colinearity between the variables. The original PC algorithm accidentally stops, if only one partial correlation between a pair of variables violates the numerical calculation, against the high similarity of the descriptors. To avoid the accidental stops by the highly associated gene pairs, the original PC algorithm is modified as follows: if the calculation of any order of the partial correlation coefficient between the variables is violated, then the corresponding pair of variables is regarded as being dependent. The second modification is to perform a permutation test. This is because the network structure inferred by the original algorithm naturally depends on the order of the variables used for calculating the partial correlation coefficients. To reduce the uncertainty of the network structure, further modifications are needed to infer robust gene relationships.

In this study, we set the significance probability of the partial correlation coefficient to 0.01, and regarded each gene relationship as being established when the relationship appeared more than 50 times among 100 permutations.

D. Network Clustering

In networks, the vertices are often clustered into tightly knit groups, with a high density of within-group edges and a lower density of between-group edges. This property is called a "community structure", and the computer algorithms for identifying the community structure are based on the iterative removal of edges with high "betweenness" scores, which identify such structures with some sensitivity. Here, we applied one of these algorithms to group the genes in the inferred network [9].

This method is based on the modularity that is measured by a parameter, the *Q*-value. The *Q*-value is defined as follows:

$$Q = \sum\left(e_{ii} - a_i^2\right), \quad (5)$$

where e_{ii} means the fraction of edges in cluster i with respect to all edges in the network, and a_i means the fraction of the number of edges that end in cluster i. First, this method considers each node as a cluster. In each subsequent step, two clusters are combined to maximize the increment of the *Q*-value, ΔQ. ΔQ is calculated as follows:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2\left(e_{ij} - a_i a_j\right) \quad (6)$$

where e_{ij} means half of the fraction of edges between clusters i and j, with respect to all edges in the network. The complexity of the calculation is on the order of $O(N)$, where N is the number of nodes in the network, and we combine two clusters at most (N-1) times; therefore, in sparse networks, the clustering is complete after $O(N^2)$ times.

E. Enrichment Analysis

The constituent genes of each cluster are characterized by known gene sets with biological functions. Here, we adopted the gene sets in CP (canonical pathways from KEGG, BIOCARTA, and REACTOME) of C2 (curated gene sets) in MSigDB [12], which contained 871 gene sets that were manually curated from previous reports to define robustly biological functions. We then estimated the enrichment probability of the genes in the expression signature for each gene set, based on the hyper-geometric probability. When the gene set is composed of k genes, and l genes are detected in the cluster, the probability is obtained by

$$P(X \le l) = \sum_{i=0}^{l} \frac{\binom{M}{i}\binom{N-M}{k-i}}{\binom{N}{k}}, \quad (7)$$

where M and N are the total number of genes in the cluster and the total number of genes in the gene sets, respectively. In this study, we set the significance probability to 0.01.

III. RESULTS AND DISCUSSION

A. Networks at Three Periods in Three Tissues

The networks inferred by the modified path consistency algorithm are depicted in Figure 1. Although the network structure can be described by some quantitative values such as network density, diameter, and clustering coefficient (data not shown), the difference is clear by visual inspection. Indeed, as seen in the figures, a few large sub-networks and many smaller sub-networks emerged for the three tissues at the three periods. However, the network sizes, and subsequently the number of small sub-networks, were distinctive between the

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

networks at 4w and those at 8w-12w and 16w-20w, regardless of the different tissues. Since the rats are not very diabetic at 4w, while they are highly diabetic at 8w-12w and 16w-20w [4-7], the difference between the sub-network structures suggests the hypothesis that the coordination of regulatory genes proceeds along diabetes progression. In other words, as a first approximation, while the partial regulations occur independently at the normal stage, the entire regulations of many genes are organized to function coordinately along with diabetes progression. Although the identification of the networks that are activated during diabetes progression is beyond the present work, it may be possible by comparing activated networks in three periods that are estimated by our network screening procedure [4].

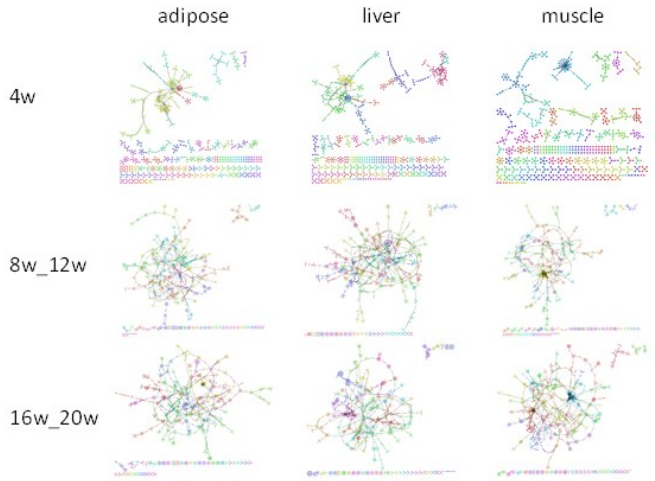

Figure 1. Networks inferred by the modified path consistency algorithm [10,11]. The regulatory networks were inferred by the modified path consistency algorithm for the three tissues (adipose, liver, and muscle) during the three periods (4 weeks, 8 and 12 weeks, and 16 and 20 weeks).

B. Network Clustering at Three Periods in Three Tissues

To further analyze the network structures of each tissue, we performed the network clustering of the large sub-networks for the three tissues during the three periods. The clusters in the sub-networks are depicted in Figure 2. As easily seen in the figure, there were few differences in the clustering patterns between the three tissues. Indeed, regardless of gene content of each cluster, the numbers of clusters at each period were similar in the three tissues: 12 in adipose, 12 in liver, and 5 in muscle, at 4w; 26 in adipose, 26 in liver, and 26 in muscle, at 8w and 12w; and 31 in adipose, 29 in liver, and 30 in muscle, at 16w and 20w. This indicated that the network structures themselves have similar connection patterns between the genes, as well as similar sizes, when gene contents in each cluster are neglected.

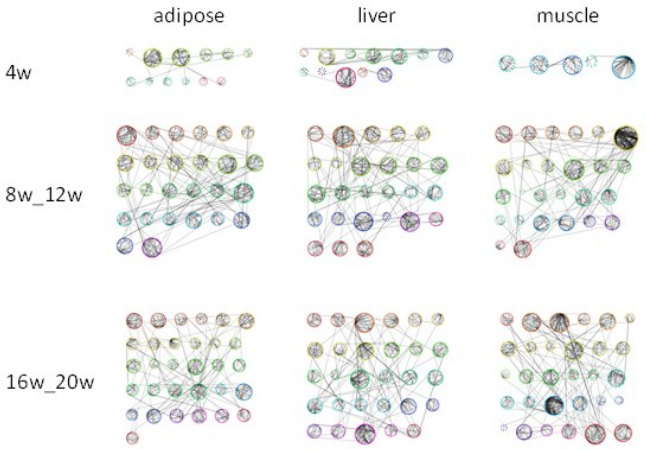

Figure 2. Network clustering in the inferred networks. The networks with more than 30 genes in Figure 1 were clustered by the Newman algorithm [9].

C. Cluster Functions at Three Periods in Three Tissues

To investigate the regulatory coordination in each network, we estimated the biological functions of each cluster in the large networks by an enrichment analysis.

The network clustering of gene expression data assumes that each cluster is composed of groups with functional regulatory relationships; i.e., a transcriptional factor and its regulated genes. As expected, all of the clusters, except for 6 of 31 clusters in adipose at 16w and 20w, were characterized by the enrichment analysis with a significant probability ($p<0.01$).

To investigate the clusters in terms of biological function, we counted the number of clusters that share the same function during different periods in each tissue (TABLE I). As seen in the table, the fractions of the number of clusters sharing functions during different periods to the total number of clusters were similar between the three tissues. Since the large networks of clusters contain most genes, this result from the cluster number may imply that a common functional mode exists along with the diabetes progression in the three tissues.

TABLE I. NUMBER OF CLUSTERS SHARING FUNCTIONS AT DIFFERENT PERIODS IN EACH TISSUE

	adipose	liver	muscle
4w	4 (12)	3 (12)	2 (5)
8w_12w	12 (26)	12 (26)	10 (26)
16w_20w	11 (25)	11 (29)	13 (30)

The total numbers of clusters for each period and tissue are indicated in parentheses.

D. Overlapped Functions During Three Periods in Three Tissues

To investigate the functional coordination along diabetes progression in each tissue, we first listed the functional terms in each cluster, and then compared the functional terms to count the overlapped terms between the clusters at the different periods in each tissue (TABLE II). As a first approximation, the number of overlapped terms may suggest

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

the coordination degree of gene regulation between the different periods in each tissue.

TABLE II. OVERLAPPING FUNCTIONAL TERMS FOUND IN CLUSTERS ALONG DIABETES PROGRESSION WITHIN EACH OF THE THREE TISSUES.

	adipose			liver			muscle		
	4w	8w_12w	16w_20w	4w	8w_12w	16w_20w	4w	8w_12w	16w_20w
4w	55	11	23	89	32	38	19	3	3
8w_12w	0.068	106	49	0.160	111	38	0.026	96	46
16w_20w	0.141	0.229	108	0.217	0.193	86	0.027	0.243	93

The number of terms and the fraction of its number to the total term number in the two periods are indicated at the upper and lower of diagonals for each tissue, respectively

Interestingly, the patterns of overlapped functional terms were distinctive between the three tissues. In adipose, the number of overlapped terms was small between the 4w and 8w-12w periods, and in contrast, was large between the 8w-12w and 16w-20w periods. Although a relatively large number of terms are shared between the 4w and 16w-20w periods, the coordination mode is different between the 4w period and the latter two periods. This coordination mode was more clearly seen in muscle. Indeed, fewer functional terms at the 4w period are shared with those at the 8w-12w and 16w-20w periods. In contrast, similar fractions were seen in the overlapped terms between the 4w and 8w-12w periods and between the 8w-12w and 16w-20w periods, and even between the 4w and 16w-20w periods. As a result, similar coordination of gene regulation was seen in adipose and muscle, and this coordination is tight during the late period of 8w-12w and 16w-20w, while the coordination in liver was mainly seen during all of the periods, from 4w, 8w-12w, and 16w-20w, and was different from the coordination in adipose and muscle.

To scrutinize the functional term patterns in the three tissues, we listed the terms found through all three periods (TABLE III). In the table, 8 and 18 terms were found in adipose and liver, respectively, and no terms were found in muscle, due to the small numbers of terms between the 4w and 8w-12w periods (3 terms in TABLE II) and between the 4w and 16w-20w periods (also 3 terms). Furthermore, 5 terms were overlapped between adipose and liver, and were also found in the functional terms common between the 8w-12w and 16w-20w periods (43 terms) in muscle (data not shown).

TABLE III. FUNCTIONAL TERMS COMMONLY FOUND AT THE THREE PERIODS IN EACH TISSUE.

adipose	**KEGG_PATHWAYS_IN_CANCER**
	REACTOME_APOPTOSIS
	REACTOME_CELL_CYCLE_MITOTIC
	REACTOME_ELONGATION_AND_PROCESSING_OF_CAPPED_TRANSCRIPTS
	REACTOME_FORMATION_AND_MATURATION_OF_MRNA_TRANSCRIPT
	REACTOME_GENE_EXPRESSION
	REACTOME_HIV_INFECTION
	REACTOME_TRANSCRIPTION_COUPLED_NER
liver	BIOCARTA_CARM_ER_PATHWAY
	KEGG_PATHWAYS_IN_CANCER
	KEGG_RNA_POLYMERASE
	REACTOME_ABORTIVE_ELONGATION_OF_HIV1_TRANSCRIPT_IN_THE_ABSENCE_OF_TAT
	REACTOME_DIABETES_PATHWAYS
	REACTOME_ELONGATION_AND_PROCESSING_OF_CAPPED_TRANSCRIPTS
	REACTOME_FORMATION_AND_MATURATION_OF_MRNA_TRANSCRIPT
	REACTOME_FORMATION_OF_THE_EARLY_ELONGATION_COMPLEX
	REACTOME_GENE_EXPRESSION
	REACTOME_HIV_INFECTION
	REACTOME_HIV1_TRANSCRIPTION_ELONGATION
	REACTOME_MRNA_PROCESSING
	REACTOME_ORC1_REMOVAL_FROM_CHROMATIN
	REACTOME_RNA_POL_II_CTD_PHOSPHORYLATION_AND_INTERACTION_WITH_CE
	REACTOME_RNA_POLYMERASE_II_TRANSCRIPTION
	REACTOME_SIGNALING_BY_WNT
	REACTOME_TAT_MEDIATED_HIV1_ELONGATION_ARREST_AND_RECOVERY
	REACTOME_TRANSCRIPTION

No terms were commonly found for the three periods in muscle. The terms commonly found in adipose and liver are indicated by bold characters.

Interestingly, the above 5 terms are related to the function of gene transcription. Indeed, 3 gene sets, "REACTOME_ELONGATION_AND_PROCESSING_OF_CAPPED_TRANSCRIPTS", "REACTOME_FORMATION_AND_MATURATION_OF_MRNA_TRANSCRIPT" and "REACTOME_GENE_EXPRESSION", are closely related to gene transcription according to their names, and the remaining 2 gene sets, "KEGG_PATHWAYS_IN_CANCER" and "REACTOME_HIV_INFECTION", include many transcription factors. As indicated in induced pluripotent stem cells [13], changing the expression of transcription factors affects the properties of the entire cell. Indeed, the mutations of transcription factors are also related to cell property changes associated with diabetes progression [2]. With these situations in mind, these results imply the existence of master regulators even in diabetes progression, as suggested from our previous network analysis for the gene expression data in liver from GK and WKY rats [4].

Another remarkable feature, in addition to the transcriptional events, is that the terms in liver refer to the events related to diabetes progression at the molecular level, such as "REACTOME_DIABETES_PATHWAYS" and "REACTOME_SIGNALING_BY_WNT" [14], while the terms in adipose refer to those at the cellular level, such as "REACTOME_APOPTOSIS" [15] and "REACTOME_CELL_CYCLE_MITOTIC" [3]. One of the interpretations of the above results may be that the events on a large scale occur in adipose, and those on a local scale occur in liver. This might imply that some molecular events in adipose are locally transferred to those in liver, as suggested in our recent study of functional communication between the two types of cells [16]. At any rate, the detailed investigation of the biological functions revealed a difference between the three tissues, against the similarity of the network structure and the number of functional terms.

E. Concluding and Further Remarks

To extract a macroscopic relationship between biological functions in different cells, we presented a procedure that combines network inference and network clustering. In its application to the gene expression data at three different periods from three tissues of GK and WKY rats, we revealed that the pattern changes in the inferred network structure and the following cluster structure, in addition to the numbers of functional terms, at the three periods were similar in the three tissues. In contrast, the overlapping functional terms between the periods suggested distinctive roles between the three tissues. From a methodological viewpoint, considerations of only the mathematical manipulations and the evaluation of numerical values would produce superficial results leading to a misunderstanding of the differences between the three tissues. Therefore, the following biological interpretation of the results was important, as the first step to pursue the molecular mechanisms underlying the phenomena at the phenotypic level.

The present analyses have revealed a hypothetical scenario for the macroscopic roles of the three tissues along diabetes

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

progression, as follows. Adipose functions commonly at the cellular level, even from the 4w period, and is responsible for additional and common roles in the 8w-12w and 16w-20w periods. Liver functions commonly at the molecular level at the 4w period, and this role is maintained until the late periods. Muscle has fewer functions at the 4w period, in comparison with adipose and liver, but at later periods, it is responsible for many roles in diabetes progression. Furthermore, each role of the three tissues during the three periods is guaranteed by master transcriptional regulators. Although the identification of the concrete gene set and the following experimental verification of the present hypothesis is beyond this work, the present results at the meta-physical level, based on network inference and clustering, provide some insights into the coordination of the three tissues along with diabetes progression.

ACKNOWLEDGMENTS

The work was partly supported by the National Natural Science Foundation of China (No. 81070657, H.Z.)

REFERENCES

[1] P. Rorsman and E. Renström, "Insulin granule dynamics in pancreatic beta cells. Diabetologia," vol. 46, pp. 1029–1045, 2003.

[2] F. M. Ashcroft and P. Rorsman, "Diabetes Mellitus and the β Cell: The Last Ten Years," Cell, vol. 148, pp. 1160–1171, 2012.

[3] M. P. Keller, Y. G. Choi, P. Wang, et al., "A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility," Genome Res., vol. 18 pp. 706–716, 2008.

[4] G. Piao, S. Saito, Y. Sun, et al., "A computational procedure for identifying master regulator candidates for diabetes progression in Goto-Kakizaki rat," BMC Sys. Biol., in press.

[5] B. Xue, S. Sukumaran, J. Nie, et al., "Adipose Tissue Deficiency and Chronic Inflammation in Diabetic Goto-Kakizaki Rats," PLoS ONE, vol. 6, e17386, 2011.

[6] R. R. Almon, D. C. DuBois, W. Lai, et al., "Gene expression analysis of hepatic roles in cause and development of diabetes in Goto-Kakizaki rats," J. Endocrinol., vol. 200, pp. 331–346, 2009.

[7] J. Nie, B. Xue, S. Sukumaran, W. J. Jusko, D. C. DuBois and R. R. Almon, "Differential muscle gene expression as a function of disease progression in Goto-Kakizaki diabetic rats," Mol. Cell. Endocrinol. vol. 338, pp. 10–17, 2011.

[8] P. Spirtes, C. Glymour and R. Scheines, Causation, Prediction, and Search. 2nd ed. Cambridge, MIT Press, 2001.

[9] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," Phys. Rev. E., vol. 69, 066133, 2004.

[10] S. Saito and K. Horimoto, "Co-Expressed Gene Assessment Based on the Path Consistency Algorithm: Operon Detention in Escherichia coli," Proceedings of IEEE International Conference on Systems, Man and Cybernetics, October 2009, San Antonio US. IEEE Xplore, pp. 4280-4286, 2009.

[11] S. Saito, T. Hirokawa and K. Horimoto, "Discovery of Chemical Compound Groups with Common Structures by a Network Analysis Approach," J. Chem. Inf. Model., vol. 51, pp. 61-68, 2011

[12] A. Subramanian, P. Tamayo, V. K. Mootha, et al., "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," Proc. Natl. Acad. Sci. USA, vol. 102, pp. 15545-15550, 2005.

[13] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," Cell, vol. 126, pp. 663-676, 2006.

[14] T. Jin, "The WNT signalling pathway and diabetes mellitus," Diabetologia, vol. 51, pp. 1771-1780, 2008.

[15] P. A. J. Krijnen, S. Simsek and H. W. M. Niessen, "Apoptosis in diabetes," Apoptosis, vol. 14, pp. 1387-1388, 2009.

[16] Y. Sun, S. Saito, K. Horimoto and H. Zhou, "Functional Networks in Diabetes-Progression by Comparison of Gene Expression in Three Tissues of Goto-Kakizaki Rats," Current Bioinformatics, in press.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE