Anti-clustering of circadian gene expression in mouse liver genome

Bin Kang, Yuan-yuan Li*, Yi-xue Li*
Shanghai Center for Bioinformation Technology
Shanghai 200235, China

***Abstract*—Circadian regulatory system is an evolutionarily ancient biological system. Its prevalence in life kingdoms suggests it has fundamental role in life processes. Although genomic scale of circadian gene expression has been found in various species from cyanobacteria to mammalians, transcriptional patterns and mechanisms of global circadian gene regulation have not yet been revealed. Using high resolution temporal profiling of mouse circadian gene expression, we show that contrary with previously demonstrated clustering tendency of functionally related genes in mammalian genomes, circadian regulated genes display anti-clustering propensity in mouse liver. This unique property does not conform to the notion of domain-wide coordinated gene regulation dictated by acetyl modifications, which is recently identified as a hallmark of circadian regulation. These results suggest that global circadian regulation in mouse liver might involve other structural chromosome interactions irrelevant with clustering regulation.**

Keywords*: anti-clustering, *clustering, circadian, global transcriptional pattern, gene expression

I. INTRODUCTION

The behavior and physiology of most organisms oscillate with daily periodicity, which mirrors the geophysical periodicity of day and night. This behavioral and physiological periodicity is governed by circadian clock, an inherent molecular clock that keeps approximately 24-hour timing. It is now evident that periodic gene expression connects the core circadian clock and downstream periodic behavioral and physiological processes. In accordance with the fact that almost all behavioral and physiological processes are under circadian regulation，more and more genes have been found displaying circadian expression [1]. In cyanobacteria, the majority of its gene transcriptional activity appears to be under circadian control [2-5]. Unlike the case in cyanobacteria, there are no consistent estimates of the number of genes displaying circadian expression in mammalian genomes. Common estimates are that 5-20% of mammalian genes are under circadian regulation. In some cases, the estimated proportion even reaches the upper limit of 100% [1].

Although there exists considerable knowledge about the core oscillator mechanism and the physiological and behavioral processes that are under circadian control, little is known about the connection between them. How the various molecular processes and behaviors are finely coordinated by circadian clock remains an enigma.

In prokaryotic cyanobacteria, an oscilloid model was put forward which suggests that global cyanobacteria circadian gene expression might be driven by rhythmic change of chromosomal status, and specific gene promoters are not essential[6].

Although no equivalent explanatory mechanism for eukaryotic circadian gene regulation has been promoted and accepted by researchers, global patterns and genomic organization of gene expression manifesting the dictating effects of chromosomal structure on gene expression in eukaryotic genomes are well known[7]. One interesting finding is the non-random organization of eukaryotic genomes and coexpression of neighboring genes in different eukaryotic genomes [8-10]. As for circadian gene expression, McDonald et al. demonstrated that rhythmically expressed genes cluster together within the same chromosomal region[11], suggesting regional or chromosomal structural regulation also plays an important role in Drosophila circadian regulation. However, Covington et. al did not found significant clustering tendency of circadian gene regulation in plant genome[12]. Thus mechanism and global organizational characteristics of circadian gene regulation in eukaryotic genomes still remain controversial and need to be furthur clarified.

Genome scale global transcriptonal characteristics of circadian gene expression are normally revealed by microarray technology. To identify circadian genes, microarray experiments are usually designed to collect data every 4 h over a course of 48 h, generating expression profiles with 12 or 13 time-points. This kind of short time series data with low sampling resolution limit the precision and accuracy of statistical analysis for circadian gene identification. Data points can not be increased simply lengthen the sampling time period by collecting data beyond 48 h because circadian rhythm dampens gradually over time. Increasing sampling resolution, however, can increase data points and analytical precision and accuracy. Recently, Hughes et. al reported their circadian transcriptional profiling experiment with higher sampling resolution on mouse liver[12]. Unlike other microarray profiling analyses for circadian gene regulation, in which samples are normally taken every 4 hours, Hughes profiled gene transcription at higher resolution of 1 hour. This unique high sampling resolution dramatically increases the confidence with which

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

circadian cycling genes can be identified and provide unique opportunity to investigate the global pattern, organizational characteristics and underlying mechanisms of eukaryotic circadian regulation. In this paper, we use this high resolution data to investigate the genomic organizational characteristics of circadian cycling genes. At least 10 percent of mouse genes proved to display transcriptional circadian rhythmicity. A striking feature from our findings is that these circadian cycling genes seem to avoid to form cluster and co-expressed neighbor genes tend to be non-cycling genes. The biological significance of this feature is discussed.

II. RESULTS

A. Neighboring genes are significantly coexpressed in context of circadian regulation

It has been widely demonstrated that neighboring genes in eukaryotic genomes tend to be coexpressed, and coexpressed genes are likely to form clusters along chromosomes [9, 14-16]. In context of circadian gene regulation, genomic gene expression is coordinated by complex circadian regulatory systems dictated by core clock genes. In this context, the coexpression of neighboring genes has not been particularly confirmed. We addressed this question based on the high resolution temporal expression profiling data released by Hughes et al. [12].

The mean Pearson's correlation coefficient (R) of all pairs of neighboring genes was calculated to measure coexpression extent of neighbor genes. The significance of this value was confirmed using permutation simulation, which compares the value obtained from real data to a distribution of random R values derived from permutated data. Our results gave clear evidence for significant coexpression of neighboring genes across the genome in context of circadian gene expression (R = 0.09, + 20 standard deviations, Fig.1A). These results demonstrate that the local positional effect on gene expression still exists even at the presence of circadian regulation. The R value reduced a little yet still remained significant (R = 0.06, + 11 deviations, Fig.1B) after tandem duplications and divergently organized genes are removed from analysis, indicating tandem duplications and common regulation are not the main cause for coexpression.

B. Co-expressed neighboring genes tend to be unrhythmic in expression

As neighboring genes are coexpressed in the context of circadian regulation, a hypothesis immediate to us is that neighboring genes might be correlated in periodic expression. To test this, we used autocorrelation coefficient (RC) as a measure of expression periodicity of each gene, and checked if RC of neighboring genes show significant correlation. The correlation between RC of neighboring genes was denoted as RR. For genes represented by multiple probesets on the microarray, the largest RC was taken as the RC for that gene. The significance of RR was evaluated using a Monte-Carlo simulation, which compares the RR obtained from real data to

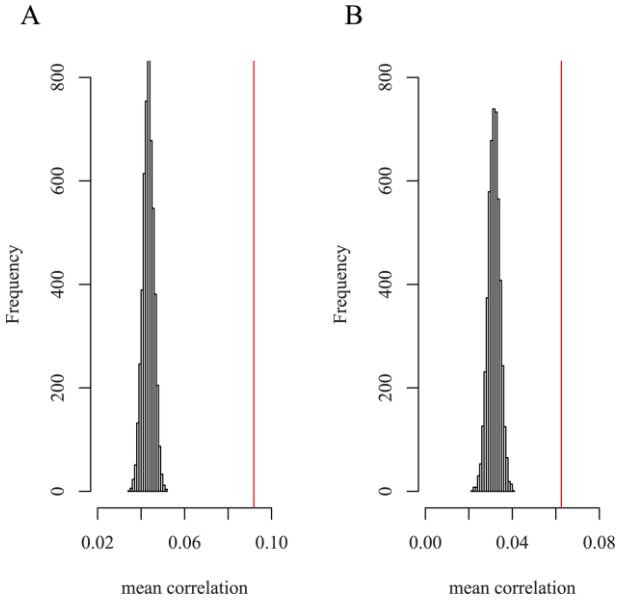

Figure 1. Histogram of neighbor gene coexpression correlation. Mean coexpression correlation of neighboring gene pairs of 10000 randomized genomes are calculated and histograms are created thereupon. Vertical red lines indicate the mean coexpression correlation of neighbor gene pairs in the real genome. (A) Mean coexpression correlation of genomes with tandems and bidirectionally transcribed genes included. (B) Mean coexpression correlation of genomes with tandems and bidirectionally transcribed genes excluded.

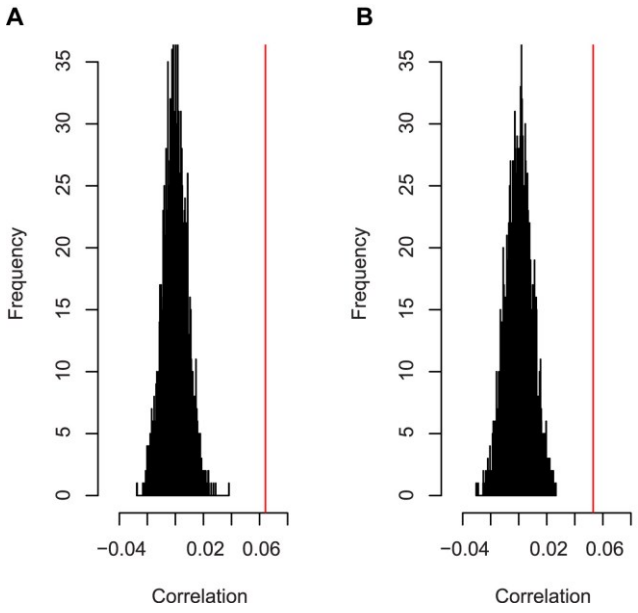

Figure 2. Histogram of periodic expression correlation of neighbor genes. Periodic expression correlation of neighbor genes are calculated for the real mouse genome and 1000 randomized genomes. Histograms are created thereupon. Vertical red lines indicate correlation for the real genome. (A) Periodic coexpression correlation of genomes with tandems and bidirectionally transcribed genes included. (B) Periodic coexpression correlation of genomes with tandems and bidirectionally transcribed genes removed.

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

a distribution of random RR values derived from 1000 permutations of the same data set. Our results again gave clear evidence that there exist small yet significant correlations between RC of neighboring genes (Fig. 2A). This correlation still remained significant after tandem duplicated genes and divergently organized genes were eliminated from the dataset (Fig. 2B).

To investigate whether this significant correlation extends beyond neighbors to distant genes, correlations between genes separated by varying numbers of genes are calculated. Our results show that RR between genes separated by up to 100 genes still remains significant (data not shown). Beyond that the correlation becomes insignificant.

Combined with the fact that neighbor genes tend to be coexpressed in context of circadian regulation, our findings that neighboring or proximal genes are significantly correlated in rhythmic expression prompted us to assume that periodic expression of a gene is significantly affected by the periodicity of the proximal genes and circadian genes might tend to be co-localized along chromosomes. If this is true, neighboring genes having high autocorrelation coefficients (RC) might tend to be highly coexpressed simultaneously. To test above hypothesis, the coexpression coefficient and mean autocorrelation coefficient of neighbor gene pairs and nonoverlapping blocks of three or six genes were calculated and compared. Surprisingly, among 2791 highly coexpressed neighbor gene pairs, only 334 have high mean autocorrelation coefficient (~12%) (Fig.3. A, C). Most highly coexpressed gene pairs have small mean autocorrelation coefficient (~88%). The lowess fit curves in Fig.4.clearly show this tendency. Gene expression profiles support the notion that gene pairs with both high mean autocorrelation and high coexpression coefficient are periodically coexpressed genes (Fig.3B, upper right), while gene pairs with high coexpression and small mean autocorrelation coefficients are non-periodic coexpressed genes (Fig.3B, lower right). For blocks of up to 6 genes, the high coexpression - low autocorrelation phenomenon still exists (Fig.4, upper panel). We suspected that the small proportion of high coexpression - high autocorrelation neighboring gene pairs or blocks might be duplicated genes or genes under common regulation. Contrary to our suspect, removing tandem duplicated genes and divergently organized genes does not eliminate high coexpression-high autocorrelation gene pairs. The proportion of high coexpression - high autocorrelation gene pairs or blocks slightly increased after removal of divergently organized genes and duplicate genes (Fig.4, lower panel). Collectively, These results indicate that not only most highly coexpressed neighbor genes do not display circadian expression， but also most genes exhibiting circadian expression tend to be not coexpressed. A small proportion of neighbor gene pairs or blocks with high coexpression and high autocorrelation are possibly real periodically coexpressed genes.

C. Rhythmically expressed genes do not cluster in mouse genome

Recently, studies in several organisms have demonstrated that tissue specific genes and genes functioning in the same pathway are often clustered in the genome [13, 14]. As for circadian gene expression, McDonald et. al have reported that

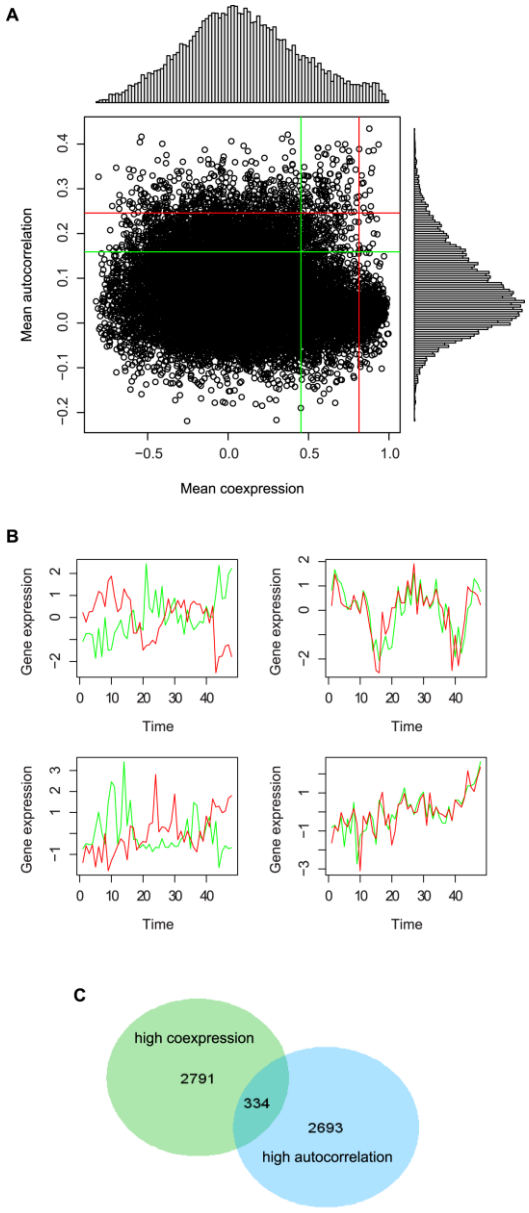

Figure 3. Relationship between coexpression and periodic expression of neighboring gene. Coexpression correlation coefficient and mean autocorrelation coefficient of neighbor gene pair are calculated. While the former value measures coexpression extent of neighbor genes, the latter measures the propensity two neighboring genes both display periodic expression. (A) Scatter plot of coexpression correlation coefficient and mean autocorrelation coefficient for all neighbor pairs. The green and red lines indicate positions of 1 and 2 standard deviations for corresponding values. (B) Representative gene expression profiles of neighbor gene pairs from corresponding quadrants of A. Starting from the upper left corner and following the clockwise direction, the profiles represent gene pairs of high mean autocorrelation coefficient-low coexpression, high mean autocorrelation coefficient-high coexpression, low mean autocorrelation coefficient-low coexpression and low autocorrelation coefficient-high coexpression. (C) Venn diagram of numbers of neighbor gene pairs with high coexpression coefficient and high mean autocorrelation coefficient. Coexpression coefficient values bigger than mean plus 1 deviation is defined as high coexpression coefficient and high mean autocorrelation coefficient is defined similarly.

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

periodically expressed genes clustered in drosophila genome [11]. However, Our results in this study have demonstrated that most highly coexpressed neighboring or proximal genes do not display circadian expression and most neighboring or proximal genes displaying circadian expression are not highly correlated in expression. This suggested that circadian regulated genes in mouse genome show a propensity of anti-clustering rather than clustering. It is thus interesting to determine whether circadian expressed genes are locally clustered or anti-clustered in mouse genome.

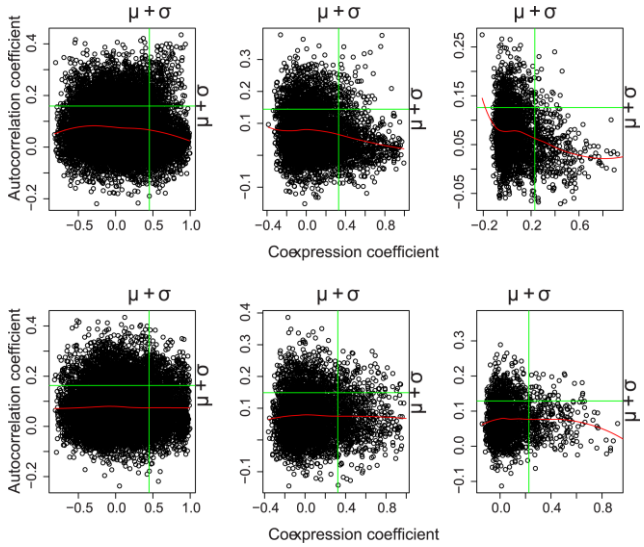

Figure 4. Scatterplot of coexpression coefficient and mean autocorrelation coefficient for neighboring gene pair(left panel), three gene block(middle panel) and six gene block(right panel). Data of upper panel are calculated before removal of tandems and bidirectionally transcribed genes, data of lower panels are calculated after removal of tandems and bidirectionally transcribed genes. Green lines indicate the position of 1 standard deviation. Red lines are lowess fit curves for autocorrelation coefficient and coexpression coefficient.

Using method of Discrete Fourier Transformation (DFT) and permutation test [15], over 2000 genes are identified to be rhythmically expressed genes. Among these genes, about 1500 genes fluctuated with a circadian period of 24 hour. These circadian expressed genes are split into 24 groups according to circadian phase of their peak expression. For each phase group, cluster score is calculated as described by Lee et. al [14]. The significance of each cluster score is evaluated by comparing the score value to a distribution of random cluster score calculated from 1000 random sampling from mouse genome. 3 of the 24 phase groups (~13%) had significant cluster scores while the majority of the groups did not have significant cluster score(Fig.5, upper panel), indicating that most groups of genes do not show propensity for clustering. After removing the tandem duplicates from mouse genome, only 1 group of 69 genes (~4%) remain significant(Fig.5, lower panel). These results suggest that although neighboring gene pairs are significantly coexpressed and circadian expression of neighbor or proximal genes are significantly correlated, circadian expressed genes do not display clustering propensity at genomic scale.

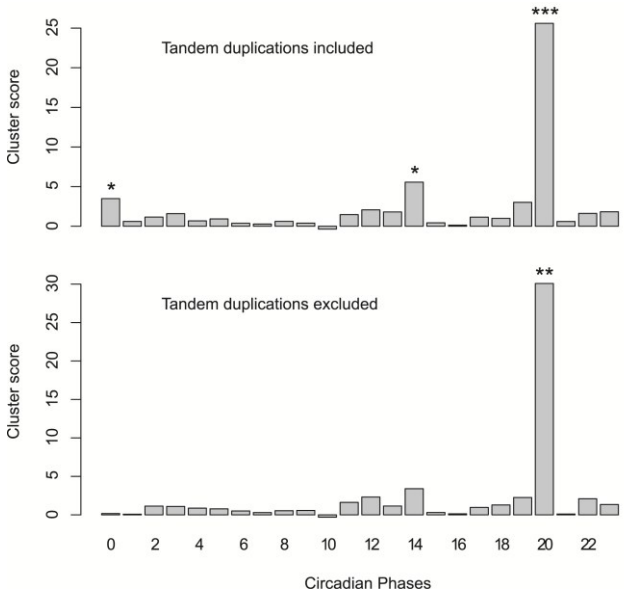

Figure 5. Clustering propensity of genes displaying circadian expression. Clustering propensity is quantitatively analyzed as described in material and methods section. Upper panels data are calculated before removal of tandems and bidirectionally transcribed genes, lower panels data are calculated after removal of tandems and bidirectionally transcribed genes. The height of each bar represents the cluster score for a group of genes with corresponding peak expression phase. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

III. DISCUSSION

Microarray and RNA sequencing technologies make it possible to analyze global gene expression and detect genome wide transcriptional patterns in various different genomes. Accumulating evidence supports the idea that gene order in eukaryotic genome is not random， and gene location and organization have significant impacts on gene expression[16]. Similarly expressed genes, including tissue specific genes, highly expressed housekeeping genes and functionally linked genes normally show localized clustering along chromosomes[8, 13, 17]. This clustering feature may be an evolutionary conserved gene regulatory mechanism with unrevealed biological significance. In this study, we investigated the global transcriptional organization of circadian gene expression in mouse liver. Our results show that most locally coexpressed genes are not circadian regulated genes and circadian regulated genes tend to be anti-clustered in mouse liver genome.

With the advent of genomic expression profiling techniques, global genomic organization characteristics has attracted much attention in recent years. Interesting properties of eukaryotic genomic organization have been revealed. Regional similarity in expression has been found in human (Caron et al. 1995; Lercher et al. 2002), Drosophila (Cohen et al. 2000; Boutanaev et al. 2002; Spellman and Rubin 2002;), yeast (Cohen et al. 2000), and Caenorhabditis elegans (Lercher et al. 2003). This seems to be a rule rather than special cases for eukaryotic genomes. In addition to regional similarity in expression, regional clustering of functionally related genes is another organizational characteristics of

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

eukaryotic genomes. The selection of these features by evolution seems rational and conforms to the principle of biological economy because placing functionally related genes in a region with similar expressional property will reduce regulatory cost.

Circadian gene regulation seems to submit to the dominion of the rule of regional expressional similarity. Our results based on high temporal resolution circadian profilings from Hughes et. al showed that neighboring genes show significant coexpression (Fig. 1). This indicates that the rule of regional similarity in expression still holds in the context of circadian regulation. We also found weak but statistically significant correlation of periodic expression between neighboring or proximal genes (Fig. 2). The magnitude of this periodicity correlation is comparable to but weaker than that of neighborhood coexpression correlation, suggesting that the periodic expression of neighboring genes might be affected by both neighborhood coexpression and other regulatory mechanisms.

However, circadian gene regulation in mouse liver seems to be a breaker of the rule of clustering of functionally related genes in eukaryotic genomic organizational rule. Scatterplotting of mean autocorrelation coefficient and co-expression coefficient of neighboring genes shows that only ~12% of highly coexpressed neighbor genes also have high mean autocorrelation coefficient and vice versa. Most neighboring genes displaying high co-expression are not periodically expressed, while those neighboring genes displaying periodic expression normally do not co-express. This means that unlike other coexpressed or functionally linked genes, periodically expressed circadian genes seem to avoid to lumped together and show anti-clustering characteristics in mouse liver genome. This feature totally contradicts with clustering expression, which is a repeatedly identified feature of eukaryotic genomic expression. Whether this anti-clustering characteristic exists in other tissues and species needs to be validated when comparable high resolution circadian profiling data are available in the future.

Quantitative analysis of the clustering tendency of mouse liver circadian genes using the scoring approach proposed by Lee and Sonnhammer[14] confirmed the notion that circadian genes do not cluster within mouse liver genome. Our result of anti-clustering of circadian regulated genes in mouse liver genome seems to be different from that of McDonald's study in drosophila[11], which found that many circadian regulated genes with similar functions form clusters in certain chromosomal regions of drosophila genome. Detailed inspection, however, revealed that our result is not controversial to that of McDonald. Among 134 Drosophila cycling genes identified in McDonal's study, only 22 form clusters along chromosomes. The clustering genes only constitute a small proportion of circadian cycling genes (~16%). And furthermore, most of the genes forming clusters are members of gene families. It seems that such clustering phenomenon in drosophila genome is just a local effect of gene family member clustering, and may not be a global characteristic of drosophila circadian transcriptional organization. Small clusters of 2-4 circadian genes are also found in our study (data not shown). These clustering circadian genes again only constitute a small proportion of mouse circadian genes (~ 10%,). It is clear from these results that mouse liver circadian regulated genes do not show significant global clustering propensity although clustering expression is a frequently identified eukaryotic genomic feature in other circumstances.

The investigation of global gene transcriptional organizational pattern in context of circadian regulation is doubly interesting comparing to that in general context since it may bear insights into the mechanisms of coordinated circadian gene regulation. Although there exists considerable knowledge about the core oscillator mechanism and the physiological and behavioural processes that are under circadian control, little is known about the connection between them. How the various molecular processes and behaviors are finely coordinated by circadian clock remains an enigma. In prokaryotic cyanobacteria, an oscilloid model was put forward which suggests that global cyanobacteria circadian gene expression might be driven by rhythmic changes of chromosomal status, and specific gene promoters are not essential[6]. Experimental evidence also confirmed that the topological status of cyanobacteria chromosome undergoes circadian fluctuation[2, 4]. This regulatory relationship between chromosomal structure and circadian gene expression seems to maintain in eukaryotic genome during evolution. Core clock gene CLOCK has been revealed to have intrinsic acetyltransferase activity[18]. NAD dependent deacetylase Sirt1 has also been demonstrated to be important regulatory factor in circadian gene regulation[19]. Still further, our analysis of mouse circadian gene expression showed that 35.9% genes related to chromatin structure regulation display circadian gene expression (data not shown). In eukaryotic genomes, association of various chromatin regulatory proteins leads to the formation of multigene chromatin domain. Genes within a chromatin domain share the same molecular environment and exhibit concerted transcription[20]. Although this domain wide concerted regulation of adjacent genes has been extensively demonstrated in various eukaryotic genomes, our finding of anti-clustering of circadian genes in this study show that this kind of domain wide concerted regulation seems to be avoided by mouse genome in context of circadian regulation. Domain wide concerted regulation and clustering of functionally linked genes are biologically economic because they can reduce regulatory complexity. Nonetheless, this kind of regulation undoubtedly will pay the price of flexibility. The deviation of circadian regulation in mouse liver from the biologically economic principle might reflect an unusual requirement of flexibility in circadian gene regulation in eukaryote genomes because the eukaryotic physiology system became more and more complex and the circadian regulating system need much more flexibility to coordinate this growing complexity during the course of evolution[21].

Since higher-order genome structures are increasingly recognized as the key components contributing to gene transcription [22], we propose that circadian genes may subject to transcriptional regulation based on other kind of spatial chromosome interaction rather than domain wide coexpression. Handoko[23] provided CTCF as an example of this kind of regulation. CTCF functions as genome organizer and bring together distant genes by intra- and

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

interchromosomal interactions. We speculate that some CTCF-like or other kind of genome organizer be activated in circadian transcriptional programs and distant genes, including genes on different chromosomes, be brought together into "transcriptional factory" to be co-expressed. In this regulatory scenario, co-expressed circadian genes will not display clustering characteristics.

IV. MATERIALS AND METHODS

A. Data source

The circadian transcriptional profiling data used in this study were downloaded from Gene Expression Omnibus under the accession number GSE11923. This dataset was genome scale RNA temporal profiling of mouse liver[12]. In comparison with other genome scale circadian profiling datasets, this dataset has a higher time resolution of 1 hour, which makes it possible to identify rhythmic genes with higher precision and accuracy.

B. Data processing and annotation

The data file downloaded from Gene Expression Omnibus contains Affymetrix Microarray Suite 5.0 computed expression level for each targeted genes on the microarray. All the expression data were log transformed before our subsequent analysis. Gene annotations were based on R package mouse4302.db. For genes represented by multiple probes, the expression profile having the maximum autocorrelation coefficient was used for subsequent analysis.

C. Detecting locale coexpression and rhythmicity correlation

Local coexpression was defined as Pearson's correlation coefficient of neighbor or proximal genes. Autocorrelation coefficient of temporal (RE) profile of each gene was used as a measure of rhythmic expression. Locale rhythmicity correlation was defined as the Pearson's correlation coefficient (RR) of autocorrelation coefficient (RC) of gene pairs separated by zero to 100 genes. Either the locale coexpression coefficient RE or the locale rhythmicity correlation coefficient RR calculated from the real dataset was compared with corresponding values calculated from permutated datasets, in which the order of the genes in mouse genome are randomized.

D. Removal of tandem duplicates and bidirectionally transcribed genes

To identify tandem duplicates in mouse genome, all mouse protein sequences were downloaded from NCBI (ftp://ncbi.nlm.nih.gov/db/) and each pair of the protein sequences from the same chromosomes was compared using BLAST algorithm. Any pair of genes within 10 genes having an e-value of less than 0.2 was counted as a tandem duplicate according to [24]. One member of each pair of tandem duplicates was removed from the analysis. A total of 2679 duplicated genes were removed. To remove bidirectionally transcribed genes, bidirectionally transcribed gene pairs were download from head-to-head gene pair database (http://lifecenter.sgst.cn/h2h/). 1431 mouse head-to-head gene pairs were found in this database and 688 genes were removed.

E. Clustering analysis

Peak expression phase of each probeset was determined according to maximal correlation position between probeset expression profile and a cosine curve with 24-hour periodicity. Gene expression periodicity was analyzed by spectral analysis. Briefly, each expression profile Y was converted from a time domain series to a frequency domain periodogram IY(ω) by Discrete Fourier Transformation (DFT). For each expression profile, 1000 random profiles were generated by permutation. Each permutated profile was transformed to frequency domain periodogram as original profile. Significance of the peak frequency of the original periodogram was estimated by Monte-Carlo simulation p value as:

$$p = \frac{N}{1000}$$

Where N is the number of permutated series for which the peak of the periodogram is greater or equal to that of the original series.

Significant series with circadian frequency were selected. For genes with multiple probsets, the series with the smallest p value was taken as the representative series. A total of 2632 genes were identified as circadian regulated genes by this analysis. 1553 remained after duplicated genes and bidirectionally transcribed genes were removed. The remaining 1553 genes were grouped according to peak expression phase. For each phase group genes, gene cluster analysis was performed as described[17]. Briefly, for each group of the 24 groups of genes, the clustering score was calculated by pairwise analysis of all genes belonging to it. For gene pairs on the same chromosome, the score was calculated by the following equations:

$$\text{pair score} = \frac{\text{average length of chromosomes in genome}}{\text{distance between genes}}$$

F

or gene pairs on different chromosomes, the score was calculated by the following equation:

$$\text{pair score} = \frac{\text{average length of chromosomes in genome}}{\text{average length of chromosomes the genes are located on}}$$

The clustering score for each group of genes was the sum of all the pairwise score in that group.

ACKNOWLEDGMENT

B.K thanks Dr. Bing Liu for reading and providing valuable ideas in the process of manuscript preparation.

REFERENCES

[1] Ptitsyn, A.A., and Gimble, J.M.: 'True or false: all genes are rhythmic', Ann Med, 43, (1), pp. 1-12

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

[2] Woelfle, M.A., Xu, Y., Qin, X., and Johnson, C.H.: 'Circadian rhythms of superhelical status of DNA in cyanobacteria', Proc Natl Acad Sci U S A, 2007, 104, (47), pp. 18819-18824

[3] Ito, H., Mutsuda, M., Murayama, Y., Tomita, J., Hosokawa, N., Terauchi, K., et al.: 'Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus', Proc Natl Acad Sci U S A, 2009, 106, (33), pp. 14168-14173

[4] Vijayan, V., Zuzow, R., and O'Shea, E.K.: 'Oscillations in supercoiling drive circadian gene expression in cyanobacteria', Proc Natl Acad Sci U S A, 2009, 106, (52), pp. 22564-22568

[5] Liu, Y., Tsinoremas, N.F., Johnson, C.H., Lebedeva, N.V., Golden, S.S., Ishiura, M., et al.: 'Circadian orchestration of gene expression in cyanobacteria', Genes Dev, 1995, 9, (12), pp. 1469-1478

[6] Woelfle, M.A., and Johnson, C.H.: 'No promoter left behind: global circadian gene expression in cyanobacteria', J Biol Rhythms, 2006, 21, (6), pp. 419-431

[7] Elizondo, L.I., Jafar-Nejad, P., Clewing, J.M., and Boerkoel, C.F.: 'Gene clusters, molecular evolution and disease: a speculation', Curr Genomics, 2009, 10, (1), pp. 64-75

[8] Michalak, P.: 'Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes', Genomics, 2008, 91, (3), pp. 243-248

[9] Williams, E.J., and Bowles, D.J.: 'Coexpression of neighboring genes in the genome of Arabidopsis thaliana', Genome Res, 2004, 14, (6), pp. 1060-1067

[10] Li, Y.Y., Yu, H., Guo, Z.M., Guo, T.Q., Tu, K., and Li, Y.X.: 'Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance', PLoS Comput Biol, 2006, 2, (7), pp. e74

[11] McDonald, M.J., and Rosbash, M.: 'Microarray analysis and organization of circadian gene expression in Drosophila', Cell, 2001, 107, (5), pp. 567-578

[12] Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., et al.: 'Harmonics of circadian gene transcription in mammals', PLoS Genet, 2009, 5, (4), pp. e1000442

[13] Li, Q., Lee, B.T., and Zhang, L.: 'Genome-scale analysis of positional clustering of mouse testis-specific genes', BMC Genomics, 2005, 6, pp. 7

[14] Lee, J.M., and Sonnhammer, E.L.: 'Genomic gene clustering analysis of pathways in eukaryotes', Genome Res, 2003, 13, (5), pp. 875-882

[15] Ptitsyn, A.A., Zvonic, S., Conrad, S.A., Scott, L.K., Mynatt, R.L., and Gimble, J.M.: 'Circadian clocks are resounding in peripheral tissues', PLoS Comput Biol, 2006, 2, (3), pp. e16

[16] Hurst, L.D., Pal, C., and Lercher, M.J.: 'The evolutionary dynamics of eukaryotic gene order', Nat Rev Genet, 2004, 5, (4), pp. 299-310

[17] Versteeg, R., van Schaik, B.D., van Batenburg, M.F., Roos, M., Monajemi, R., Caron, H., et al.: 'The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes', Genome Res, 2003, 13, (9), pp. 1998-2004

[18] Doi, M., Hirayama, J., and Sassone-Corsi, P.: 'Circadian regulator CLOCK is a histone acetyltransferase', Cell, 2006, 125, (3), pp. 497-508

[19] Nakahata, Y., Kaluzova, M., Grimaldi, B., Sahar, S., Hirayama, J., Chen, D., et al.: 'The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control', Cell, 2008, 134, (2), pp. 329-340

[20] Gierman, H.J., Indemans, M.H., Koster, J., Goetze, S., Seppen, J., Geerts, D., et al.: 'Domain-wide regulation of gene expression in the human genome', Genome Res, 2007, 17, (9), pp. 1286-1295

[21] Masri, S., and Sassone-Corsi, P.: 'Plasticity and specificity of the circadian epigenome', Nat Neurosci, 13, (11), pp. 1324-1329

[22] Misteli, T.: 'Beyond the sequence: cellular organization of genome function', Cell, 2007, 128, (4), pp. 787-800

[23] Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., et al.: 'CTCF-mediated functional chromatin interactome in pluripotent cells', Nat Genet, 43, (7), pp. 630-638

[24] Lercher, M.J., Urrutia, A.O., and Hurst, L.D.: 'Clustering of housekeeping genes provides a unified model of gene order in the human genome', Nat Genet, 2002, 31, (2), pp. 180-183

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