Supporting Information

Chen et al. pnas

Materials and Methods

Cell Culture and Crosslinking of Chromatin. Human foreskin fibroblasts from a normal karyotyped male individual (Cat # CRL-2522, ATCC, Manassas, VA) were propagated in growth medium, composed of MEM medium (cat #11095-098, Life Technologies, Grand Island, NY), 10% fetal bovine serum (FBS, cat #10082-147, Life Technologies), 1 X non-essential amino acids (NEAA, cat #11140-050, Life Technologies), and 1 X antibiotic/antimycotic (cat #15240-062, Life Technologies).

To prepare cultures for time series Hi-C and RNA-seq experiments, we trypsinized fibroblasts propagated in T225 flasks with 0.25% trypsin (Cat #25200-056, Life Technologies). Dissociated cells were plated in 150mm cell culture plates for Hi-C experiments, and in 6-well plates for RNA-seq experiments. Cells were cultured in growth medium for 36 hours before proceeding to cell cycle synchronization. To synchronize the cell cycle, the cells were incubated in serum-free MEM medium supplemented with 1 X NEAA and 1 X Antibiotic/antimycotic for 48 hours. Under this condition, it has been reported that more than 95% of the cells are in G0/G1 phase [1].

We also synchronized the circadian clock in the cell cycle-synchronized cultures with dexamethasone treatment for one hour [2]. At the 48th hour of cell cycle synchronization, we added dexamethasone to the cultures at a final concentration of 100nM for the treatment group. For the control group, we added ethanol to cultures at a final concentration of 0.001% (dexamethasone solvent concentration in the treatment cultures). At the end of circadian clock synchronization, cells were rinsed twice with PBS, fed with growth medium, and time point zero was set.

Base line samples from dexamethasone treated plates and control plates were taken at the end of the 1-hour circadian clock synchronization without feeding of growth medium. Thereafter, time series sampling of cells treated with dexamethasone was performed at 8-hour intervals for a total of 56 hours. All samples for Hi-C experiments were cross linked in situ. Total RNA was extracted directly from 6-well plates (see RNA isolation and RNA-seq).

Approximately 25 X 10^6 cells were cross linked with 1% formaldehyde (Cat #BP531-25, Fisher Scientific, Pittsburgh, PA) in serum free-medium for 10 min at room temperature, and then quenched with glycine (Cat #G8898-500g, Sigma-Aldrich, St. Louis, MO) added to a final concentration of 0.125 M. Cross linked cells were harvested and flash frozen in liquid nitrogen, and then stored at -80°C until the construction of Hi-C libraries.

RNA Isolation and RNA-seq. We used TRIzol® Reagent (Cat # 15596-018, Life Technologies) to extract RNA from cells grown in 6-well plates. All extracted RNA was treated with RNAse-free DNaseI (Cat # 79254, Qiagen), then submitted to the University of Michigan Bioinformaics Core lab for library construction and RNA sequencing (RNA-seq) on the Illumina Hiseq-2000 platform. Single-end 50-base sequence reads were generated at a multiplex of 4 per sequencing lane.

During RNAseq data processing, we checked the raw reads with FastQC [3] (version 0.10.0), to identify potential quality problems in the reads data (eg. low quality scores, over-represented sequences, inappropriate GC content). We used Tophat [4] (version 2.0.9) and Bowtie [5] (version 2.1.0.0) to align the reads to the reference transcriptome (HG19). We used default parameter settings for alignment, with the exception of: "–b2-very-sensitive" telling the software to spend extra time searching for valid alignments, as well as "–no-coverage-search" and "–no-novel-juncs" to limit the search to known transcripts. We then performed a second round of quality assessment using FastQC on the aligned reads. Data was found to be of excellent quality overall. We used Cufflinks/Cuffdiff[6, 7, 8, 9] (version 2.1.1) for expression quantification and differential expression analysis, using UCSC hg19.fa as the reference genome and UCSC hg19.gtf as the reference transcriptome. For the CuffDiff analysis, we used parameter settings: "–multi-read-correct" to adjust expression (FPKM) calculations for reads that map in more than one locus, as well as "–compatible-hits-norm" and "–upper-quartile-norm" for normalization of expression calculations across samples. We used a locally developed Perl script to format the Cufflinks output.

Bru-seq Analysis of Nascent RNA Species. Fibroblast cultures were synchronized at G0/G1 phase with incubation in serum-free medium for 48 hours. Dexamethasone was then added to the cells (see Cell culture and crosslinking of chromatin). Thirty minutes after the addition of Dexamethasone, Bromouridine (Aldrich) was added to the media to a final concentration of 2 mM, and cells were incubated at 37 C for 30 min. Cells were then washed three times in PBS and collected directly (nascent RNA, Bru-Seq). RNA was isolated by using Trizol reagent (Invitrogen), and Bru-labeled RNA was isolated from the total RNA by incubation with anti-BrdU antibodies (BD Biosciences) conjugated to magnetic beads (Dynabeads, Goat anti-Mouse IgG; Invitrogen) under gentle agitation at room temperature for 1 h. Isolated Bru-labeled RNA was used to prepare strand-specific DNA libraries by using the IlluminaTruSeqKit (Illumina) according to the manufacturer's instructions. Sequencing of the cDNA libraries prepared from nascent RNA was performed at the University of Michigan Sequencing Core by using the Illumina HiSEq-2000 sequencer.

Generation of Hi-C Libraries. We adopted the methods for Hi-C library construction from Belton et al. (Methods 58:268, 2012). For each Hi-C library, approximately 25 X 10^6 cells were resuspended in 1mL ice-cold lysis buffer, consisting of 10mM Tris-HCl, 10mM NaCl, 0.2% Igpel (Cat # 8896-50mL, Sigma-Aldrich), and 10 mL protease inhibitor cocktail (Cat # P8340-1ml, Sigma-Aldrich). All resuspended cells were incubated on ice for 15 min. Cells were homogenized in a Dounce homogenizer on ice with pestle A, and the lysate was transfer to a 1.7mL tube. Cells were collected by spinning for 5 minutes at 2,000xg, and then washed twice in 500 μL of ice cold 1x NEB buffer 2. Cells were distributed between 4 individual 1.7 ml centrifuge tubes (50 μL per tube). Chromatins in each tube were digested with 400u of restriction enzyme HindIII (Cat # R0104M, New England BioLabs, Ipswich, MA) in 1x NEB buffer 2 at 37°C overnight on a spin wheel.

After HindIII digestion, restriction site overhanging ends were filled and labeled with biotin using DNA polymerase I large (Klenow) fragment (Cat # M0210L, New England BioLabs) in a reaction containing dATP, dGTP, dTTP, and

biotin-14-dCTP (Cat # 19518-018, Life Technologies) in each of the 4 HindIII digestion tube. DNA fragments labeled biotin-14-dCTP from each of the 4 tubes were ligated at 16^{o}C for 4 hours in an 8.23 mL reaction containing 1x ligation buffer, 1% Triton X-100 (Cat # T8787-250ML, Sigma-Aldrich), 1 mg/ml bovine serum albumin (BSA)(Cat # BP9706-100, Fisher Scientific), 10 mM ATP (Cat # A9187-1g, Sigma-Aldrich), and 50u T4 DNA ligase (Cat # 15224-025, Life Technologies). Reverse cross linking was performed in two steps. First, 50 μL of 10 mg/ml proteinase K (Cat # 25530-015, Life Technologies) were added to each ligation reaction tube and incubated at 65^{o}C for 4 hours. Then, another 50 μL of proteinase K were added to each tube and continued incubating at 65^{o}C overnight. Next, DNA was extracted with saturated phenol : chloroform (1:1) (Cat # 1100631, Fisher Scientific), and desalted by using AMICON® Ultra Centrifugal Filter Unit (Cat # UFC503024, Millipore, Billerica, MA) with 1 x TE buffer. The final volume of desalted DNA was adjusted to 100 μL in 1 x TE buffer.

Removal of Biotin from un-ligated ends was carried out in 8 individual reactions each of 50 μL containing 5 μg of Hi-C DNA, 1 mg/ml BSA,1X NEB buffer 2, 25 nM dATP, 25 nM dGTP, and 15 u T4 DNA polymerase at 20^{o}C for 4 hours. The Hi-C DNA was then pooled in a single tube, purified with single phenol extraction, and precipitated by ethanol. The DNA was re-dissolved in 105 μL of water, and transferred to a microTUBE AFA tube (Cat # 520045, Covaris, Woburn, Massachusetts). DNA fragmentation was performed in a Sonicator (Covaris S2, Covaris). The DNA fragments of size 200–400 bp were recovered with Agencourt AMPure XP mixture (Cat # A63880, Beckman Coulter, Indianapolis IN) following the manufacturer's protocols.

DNA fragment ends were repaired in a 70 μL reaction containing 1 X ligation buffer (Cat # B0202, New England BioLabs), 0.25 mM of dNTP mixture, 7.5 u of T4 DNA polymerase (Cat # M0203L, New England BioLabs), 25 u of T4 polynucleotide kinase (Cat # M0201S, New England BioLabs), 2.5 u of DNA polymerase I large fragment at 20^{o}C for 30 min. The reaction is purified with a MinElute column (Cat 28204, Qiagen, Valencia, CA). The DNA is eluted in 32 μL of elusion buffer for A-tailing, which was performed in a 50 μL reaction containing purified DNA (5 μg), 1 X NEB buffer 2, 0.2 mM dATP, 15 u Klenow fragment (3'→5' exo-) (Cat # M0212L, New England BioLabs). The reaction was incubated at 37^{o}C for 30 min, then at 65^{o}C for 20 min to inactivate Klenow (exo-).

For Streptavidin pull-down of biotinylated Hi-C ligation products, the biotinylated Hi-C ligation products are mixed with MyOne C1 streptavidin bead solution (Cat # 65001, Life Technologies) for binding of biotinylated Hi-C fragments. Non-specifically binding DNA was removed by washing with 1 X binding buffer (5 mM Tris-HCl (pH8.0), 0.5 mM EDTA, and 1 M NaCl), then with 1 X T4 Ligation buffer (Cat # 46300-018, Life Technologies). The DNA-bound beads were resuspended in 38.75 μL of 1 X ligation buffer for adapter ligation.

Illumina adapter ligation was performed in a 50 μL reaction by adding to the DNA-bound beads suspension of 1 X T4 ligation buffer, 90 pM of Illumina paired end adapter, 3 u of T4 DNA ligase (Cat # 15224-025, Life Technologies). The reaction was incubated at room temperature for 2 hours. The beads were reclaimed, and the supernatant discarded. The beads were washed twice in Tween Wash Buffer (5 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, 1 M NaCl, 0.05% Tween 20), and once in 1 X binding buffer (5 mM Tris-HCl (pH 8.0), 0.5 mM EDTA, and 1 M NaCl), and twice in 1 X NEB buffer 2. After the last wash, the beads were resuspended in 20 μL of 1X NEB buffer 2.

The Hi-C DNA sample was amplified by 15 PCR cycles (optimized in the log amplification phase) for Illumina HiSeq sequencing. Each PCR reaction in 25 μL, 1.5 μL of Bead-bound Hi-C DNA, 0.35 μL of PE primer 1.0, 0.35 μL of PE primer 2.0, 0.2 μL of 25mM dNTP, 2.5 μL of 10X PfuUltra buffer, 19.6 μL of H_2O, and 0.5 μL of PfuUltra Fusion DNA polymerase. The PCR cycling parameters were 98^{o}C for 30 seconds, followed by 15 cycles at 98^{o}C for 10 seconds, 65^{o}C for 30 seconds, and 72^{o}C for 30 seconds, and a final extension at 72^{o}C for 7 minutes. PCR products pooled from the supernatant of multiple reactions were subjected to AMpure XP beads purification to remove primer dimers. A standard quality control (QC) procedure was performed on the purified PCR products (Hi-C library). Each Hi-C library passed the QC procedure and was then sequenced in a single lane of a flow cell on a HiSeq 2000 sequencer to generate paired-end sequence reads at 100 bases per end read.

Generation of Hi-C Matrices. We standardized a pipeline to process Hi-C sequence data at the University Bioinformatics Core facilities. With this pipeline, raw sequence reads were processed with FastQC (`http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/`) for data quality control. Paired-end reads with excellent quality were mapped to the reference human genome (HG19) using Bowtie2 [10], with default parameter settings and the "–very-sensitive-local" preset option, which produced a SAM formatted file for each member of the read pair (R1 and R2). HOMER (`http://homer.salk.edu/homer/interactions/`) was used to develop the contact matrix with "makeTagDirectory" with the "–tbp 1" setting, and with "analyzeHiC" with the "-raw" and "-res 1000000" settings to produce the raw contact matrix at 1MB resolution.

Supplementary Text

Cell Cycle Synchronization Analysis. We aim to initiate the experiments with a cell population arrested at G0/G1 phase of the cell cycle. Serum-starvation has been shown as an effect method for this purpose (Exp. Cell Res. 179:40, 1988; and 'Part III Serum Starvation (G0/G1 block)' in the protocol, "State of the art in human cell synchronization" from the Biotech Research and Innovation Centre, University of Copenhagen (`http://www.docstoc.com/docs/26023877/Synchronization-protocols-for-human-cells---DOC---DOC#`), and has been used previously (Science 283:83, 1999). Our experiment is designed to investigate genome form and function over time following re-stimulation of the initially arrested at G0/G1 cell population into proliferation, but not to target specific cell cycle phases.

We used the ratio of near interactions to far interactions to estimate the time point when the first wave of cell division likely occurs (see Sec. 2.2 for details). Our RNA-seq data corroborates this, though we acknowledge that the peaks of chr21 are offset by 8 hours. From our RNA-seq data, expression of cell cycle genes aligns with cell cycle phase in the first division wave after serum stimulation, i.e. genome replication between 8-16 hours, then entering mitotic division. However, this pattern is difficult to detect thereafter (Figure S5A).

We used mitotic cell counting to estimate that the first wave of mitotic division begins at 20 hours and peaked at 32 hours after serum stimulation (Figure S5B). The Gaussian distribution centered around the 32 hr mark strongly suggests cell

synchronization has occurred. An un-synchronized population would demonstrate a fairly constant percentage of cells in M-phase at each time point over the sampling period. Mitotic cells were defined as those showing a nuclear plate or rosette, or a set of two daughter cells before chromosome decondensation (late metaphase, anaphase, and telophase), with the full M-phase generally accepted to last ∼1hr in an ∼24 hr cell cycle. Though it is not possible to use these counts to accurately estimate the total percentage of cells undergoing division, if an arbitrary bin window of 20 minutes for these sub-stages of the M phase of the cell cycle is selected, a rough estimate of cells undergoing mitosis from the 16th to 52nd hrs is 86.5%, with 76.0% in the 22nd to 42nd hr period (Figure S5B).

The overall mitotic pattern of Figure S5B compares favorably to the pattern of S-phase human diploid fibroblast cells shown by flow cytometry of BrdU and propidium iodide – stained cells in 'Part III Serum Starvation (G0/G1 block)' in the protocol, "State of the art in human cell synchronization" from the Biotech Research and Innovation Centre, University of Copenhagen (`http://www.docstoc.com/docs/26023877/Synchronization-protocols-for-human-cells---DOC---DOC#`) . In hindsight, a cell sorting analysis would have provided corroborating data for the gene expression and mitotic cell counting to further demonstrate cell synchronization.

M-Phase Analysis. The genome can be viewed as a dynamic graph, of which the structure is time variant. The dynamics of the graph structure can be translated to the changes of Hi-C counts. At a large scale, the most significant overall structural change of chromosome can be expected at the Mitosis phase (M-phase) where the pairs of chromosomes condense (implying higher contact frequencies). Literature shows that at M-phase the captured intra-chromosome Hi-C matrices become more concentrated along the diagonal, which can be expected since chromosomes naturally line up in physical space during this phase. Due to the mixture of inherently imperfectly synchronized cells and the short duration of M-phase, we did not specifically capture Hi-C matrices in this phase. However, in order to verify that at some instances there may be some cells that are in the M-phase in the population, we calculate the ratio between the sum of counts close to the diagonal (diagonal bins + bins within 5% of chromosome length) and the sum of the rest of this intra-chromosome matrix. If some cells at M-phase exist in the population, this ratio would be increased. We plot the change of this ratio over time for all chromosomes in Figure. S5C. The peaks are generally found at time points 24 and 48, or over 16-24 and 40-48, during which periods the cells are supposed to pass M-phase.

Gene expression Clusters Analysis. We present a dynamical view of gene expression in human fibroblasts based on RNA-seq analysis [11, 12] of a 56-hour time course. We analyzed three replicates on each condition from cells initially with cell cycle and circadian clock synchronized. Time zero includes Dexamethasone treatment samples (D) and corresponding base line controls (E) without exposure to serum. The rest are sampled at 8-hour intervals counting from time zero after exposing to serum. Details of sampling scheme are described in material and methods.

We identified a set of 7786 genes which significantly varied in expression levels between any two-time points in the time course studied (Table S2). The selection of this set of significant genes was based on the following criteria: (i) showed at least 1.5 fold difference between any two time points, (ii) over time mean expression RPKM value >1, and (iii) Ratio between variance over time and variance within each time point replicate larger than 1.5.

To have a global view of the expression patterns of these 7786 genes (4824 increased expression levels, 2962 decreased levels), we performed clustering analysis based on the correlation matrix of their expression levels over time, and consider it as the weighted adjacency matrix. Variance normalized spectral clustering is then applied on this adjacency matrix with the cluster number set to 6, 8, 16, or 32. We found that a two-step spectral clustering process worked best, first to group the genes into 8 top clusters (Figure S2A), then to re-cluster each top cluster into 4 sub-clusters for total of 32 sub-clusters, which mostly represent the expression patterns of the significant genes. Overall, the expression variation over time dominantly stands out at specific time points, however, the overall changes between D and E is modest compared to that of other time points. The differentially expressed genes are considered as responding to Dexamethasone treatment (see Dexamethasone treatment responsive module).

We observed several characteristic expression patterns in the cells after serum stimulation (Figure S2A). Compare to base line D & E levels, the mean expression levels in top clusters C1, C 2, C5, and C6 are increased over time and peaked at hour 8 (C2), hour 16 (C5), hour 24 (C2), or hour 48 (C6). A second peak can be seen at hour 32 (C6) or hour 48 (C2). Top cluster C7 shows spiking levels at hour 8, then return to slightly below D & E over time. C8 shows the opposite changes compared to C7, with sharp repressed expression at hour 8, and then returns to base line level over time. C3 expression is significantly repressed at hour 8 & 16, and then remains at a level slightly below D & E. C4 shows dramatically decreased expression levels from hour 8 through 56.

We performed Gene Ontology (GO) analysis [13] of the genes in each sub-cluster for enrichment under GO terms (Table S3, FDR<0.05). A summary of the top 65 significant GO terms enriched with genes (Bonferroni p <0.05) from each of the 32 sub-clusters is presented in Table S4.

Among the top 65 significant GO terms, we found that the most often used terms were cell cycle related terms such as: cell cycle phase, cytokinesis, cell division, DNA-dependent DNA replication, DNA repair, spindle checkpoint, (mitotic) spindle organization and biogenesis, M phase, mitosis, and regulation of mitosis. Other terms include: amino acid metabolic process, biopolymer metabolic process, cytoskeleton organization and biogenesis, nucleobase, nucleoside, nucleotide and nucleic acid metabolic process, organelle organization and biogenesis, translation, and wound healing (Figure S2B-C).

RNA-seq Functional Module Analysis.

Cell cycle Module

We are interested in identifying genes related to cell cycle that are novel findings from this study. We found that among the 7786 significant gene, 724 are functionally defined as cell cycle genes based on 1) GO term "cell cycle" or 2) genes identified by Whitfield et al. [14]. Among the 724 genes (Table S5), our spectral clustering correctly groups cell cycle phase known genes in corresponding sub-clusters that are correlated with the time points for S, G2/M, or M phase after serum stimulation (Figure S5A). According to the expression pattern of known cell cycle phase-expressed genes such as DNA replication/repair (S phase) genes, (Figure S5A) these genes are repressed until an activation spike at the 16-hour time point. This provides further evidence that our initial cell population is indeed synchronized at G0/G1, and that DNA replication occurs at around 16 hour as observed by others [15, 16]. In

each cluster corresponding to cell cycle phase, in addition to the well-known cell cycle genes, e.g., *MCM2*, *MCM5*, *POLD1*, *POLE*, and *RAD51* (DNA replication and repair); *CCNA2* and *CDC25C* (promoting G1/S or G2/M transition); *BUB1B*, *CCNB1*, *CCNB2*, and *CDC20* (G2/M transition), there are a large number of genes that have not been fully documented for their role in cell cycle. Because their expression patterns are very similar to those known to be cell cycle regulated, we speculate that they are functionally similar to the known cell cycle genes.

Circadian Clock Module

In mammalian cells, the core circadian genes, e.g., *CLOCK*, *ARNTL*, *PER1*, *PER2*, *PER3*, *CRY1*, *CRY2*, *NPAS1*, *TIMELESS*, are known to govern the autonomous circadian clock [17]. The circadian clock oscillators (the cells) in a culture can be synchronized (phase resetting) with the treatment of Dexamethasone [2]. In our experiment, the cells were treated with Dex to reset the circadian clock. We analyzed our time course RNA-seq data with the software JTK-CYCLE designed for identifying gene expression rhythmic periodicity [18]. We obtained a set of 1040 genes which appeared to have rhythmic period of 16, 24, or 32 hours (adjusted P<0.01 and amplitude>0.5) (Table S6). Gene ontology (GO) analysis of these rhythmic genes identified a set of genes (*ARNTL*, *CRY1*, *CRY2*, *CSNK1D*, *NR1D1*, *NR1D2*, *PER1*, *PER2*, *PER3*) significantly enriched under KEGG pathway "Circadian rhythm" which are known components of the core circadian genes [11]. These 9 genes were also identified in the 7786 genes list described above. Other components of known core circadian genes had either no significant changes (i.e. *CSNK1A1*, *CSNK1D*, *CSNK1E*, *RORA*) in expression levels or were not detected in this experiment (*FBXL21*, *RORB*, *RORC*). We also identified *C1orf51* (CIART) has a 32-hour period. This gene encodes an E-box binding protein (CHRONO) known to modulate circadian gene expression and function as a core component of the mammalian circadian clock [19]. The large number of genes appears to have rhythmic periods suggesting that additional genes play important roles in the mammalian circadian clock.

Wound Healing Module

We are interested in genes responsive to serum stimulation (wound healing genes). Genes in human fibroblasts responding to serum are thought to play important roles in wound healing physiology and cancer pathophysiology [20, 15]. We obtained a set of 939 wound healing genes from publicly available data [20, 15], and found that our set of 7786 genes contained 709 of the known 939 wound healing genes (75.5%, Fisher exact test P<1E-7) (Table S7). This suggests that the majority of genes known to be related to wound healing and cancer pathophysiology can be reliably identified. Furthermore, we observed thousands of genes that changed expression levels in this 56-hour time course study after serum stimulation, which indicates a more complex situation that involves many more genes with undefined roles in the wound healing process.

Dexamethasone (Dex) Treatment Responsive Module

Dexamethasone has been shown to reset the circadian clock in peripheral tissue and culture cells [2, 21]. In our experiment we not only synchronized the cells at G0/G1 phase, but also reset the circadian clock with Dex treatment (See Materials and Methods). At time zero, D and E samples were not exposed to serum. We identified a set of 953 genes with significantly altered expression levels (P<0.05, FC>10%) in D samples contrast to E controls (Table S8), and 524 of them (55%) were identified in the set of 7786 significant genes that varied over time. We applied Bru-seq [22] to identify genes with instant transcriptional changes due to Dex treatment (See Materials and Methods). We identified 921 newly initiated transcripts that show significant changes in expression levels 1.5 fold or greater in the treated sample compared to the control (Table S9). Among these 921 genes, 193 genes show significant changes in response to Dex treatment at time zero (between the D and E samples). The union set of 193 genes has significant overlapping (Fisher exact test P = 1E-7) and shows correlated expression changes (R2 = 0.77, and 93% in same direction change) between Bru-seq and RNA-seq.

The Bru-seq technique targets instant transcription, however, our RNA-seq experiment only interrogates processed mRNA transcripts. Since both RNA-seq and Bru-seq samples were collected at the same time, the non-overlapping changes in 728 genes (921-193) between the Bru-seq and time zero sets may suggest that the instant nascent RNA transcripts detected by Bru-seq may be not fully processed within the sampling time window for RNA-seq. The other possibility is that there are a portion of false discovered genes in both sets of significant genes. In addition, considering a potential processing lag between nascent transcripts and mature mRNAs, there may be a higher percentage overlapping between Bru-seq and RNA-seq. We compared the 921 Bru-seq significant genes with the 7786 genes with high variance changes detected by RNA-seq, and found that 695 of the 921 Bru-seq significant genes (75.5%) were identified in the set of 7786 RNA-seq high variance genes. This data indicates that a substantial proportion of the 7786 genes may be detected in later hours and are potentially Dex responsive genes, though a potential influence from serum stimulation on the expression changes of these genes cannot be completely excluded.

Hi-C Matrix Normalization.

Toeplitz Normalization (O/E Normalization)

It is expected that nearby loci in linear base pairs distance are more likely to be ligated than distant pairs. This makes a Hi-C matrix highly diagonally dominant and conceals the contact pattern embedded in the matrix. In order to alleviate this effect, we normalize the counts by their contact probability as a function of the linear distance, namely, each entry of the matrix is normalized by its expected contact value. This is equivalent to normalization of the Hi-C matrix by a Toeplitz structure whose diagonal constants are the mean values calculated along diagonals of the observed matrix. Considering Hi-C matrix $\mathbf{H}^{(i)}$ of chromosome i with length $L^{(i)}$, the normalization is performed by:

$$\mathbf{H}^{(i)}_{\mathrm{TP}} = \mathbf{H}^{(i)} \oslash \mathbf{E}^{(i)} \qquad [1]$$

where $\oslash$ represent the Hadamard division of matrices, and entries of $\mathbf{E}^{(i)}$ are given by

$$[\mathbf{E}^{(i)}]_{k\ell} = \frac{1}{\mathrm{card}(\mathcal{D}_{k\ell})} \sum_{m,n \in \mathcal{D}_{k\ell}} [\mathbf{H}^{(i)}]_{mn} \qquad [2]$$

with the set $\mathcal{D}_{k\ell} = \{m, n \,|\, m - n = k - \ell, 0 < k, \ell < L^{(i)}\}$, and $\mathrm{card}(\cdot)$ denoting the cardinality of its set argument.

Modified Toeplitz Normalization for Large Chromosomes

Centromeres separate chromosomes into two arms. The contacts between two arms of the same chromosome are significantly lower than contacts within the arms, and behave in a manner closer to the inter-chromosome contacts. This phenomenon leads to the Toeplitz normalized matrix containing much lower values at the off-diagonal entries close to the centromere in large chromosomes, as compared to the rest of the matrix. In order to alleviate this effect, the intra-chromosome Hi-C matrix can then be segmented into four divisions with boundaries determined by the centromere position. The Toeplitz normalization is then applied to each of these divisions.

Stratified Toeplitz Normalization for High Resolution Matrices

Hi-C matrices at higher resolution (e.g. 100kb resolution) may present a sparse pattern for off-diagonal entries, especially when we do not have a large number of total reads. The contact probability at far off-diagonals can then be very small due to the sparse nature of the region. Applying the Toeplitz normalization on such a matrix will largely amplify the non-zero entries at off-diagonals. This will lead to a normalized matrix distorted by a band structure, and largely prevent us from investigating its long-range structures. We thus propose a stratified variant for Toeplitz normalization by using the following expected matrix instead:

$$[\mathbf{E}^{(i)}]_{k\ell} = \frac{1}{\mathrm{card}(\widetilde{\mathcal{D}}_{k\ell})} \sum_{m,n\in\widetilde{\mathcal{D}}_{k\ell}} [\mathbf{H}^{(i)}]_{mn} \qquad [3]$$

with the set $\widetilde{\mathcal{D}}_{k\ell} = \{m, n \,|\, m - n = k - \ell, 0 < k, \ell < L^{(i)}, [\mathbf{H}^{(i)}]_{mn} \neq 0\}$.

Normalization for the Inter-chromosome Matrices

The inter-chromsome matrices can be normalized by dividing their (stratified) mean contact probability, so that they have similar visibility compared with intra-chromosome matrix. It can be observed that inter-chromosome structures are inherited from intra-chromosome Hi-C matrices.

Spectral Graph Theory. Graph theory plays an important role in modeling and analyzing the genome architecture and function. Relevant concepts are reviewed for supporting and framing our analysis [23, 24]. We define a graph $\mathcal{G} = (V, E)$ where $V = \{v_1, v_2, \ldots, v_N\}$ is a finite set of vertices with the cardinality N, and E is the edge set consisting of elements of the form $\{v_i, v_j\}$. The adjacency matrix $\mathbf{A}(\mathcal{G})$ (or simply $\mathbf{A}$ for short) is the symmetric $N \times N$ matrix encoding the adjacency relationships in the graph $\mathcal{G}$, such that $[\mathbf{A}(\mathcal{G})]_{ij} = 1$ only if $\{v_i, v_j\} \in E$, otherwise $[\mathbf{A}(\mathcal{G})]_{ij} = 0$, with $[\cdot]_{ij}$ denoting the ijth entry of its matrix argument. The degree of a given vertex, denoted by $d(v_i)$, is the cardinality of the neighborhood set of v_i. This degree is equivalently expressed via the adjacency matrix by $d(v_i) = \sum_{j\in\mathcal{N}} [\mathbf{A}(\mathcal{G})]_{ij}$. The degree matrix, $\mathbf{D}(\mathcal{G})$, is defined as a diagonal matrix with the ith diagonal entry given by $d(v_i)$. The Laplacian of $\mathcal{G}$ is defined by

$$\mathcal{L}(\mathcal{G}) = \mathbf{D}(\mathcal{G}) - \mathbf{A}(\mathcal{G}) \qquad [4]$$

and the normalized variant is given by

$$\mathcal{L}(\mathcal{G}) = \mathbf{D}(\mathcal{G})^{-\frac{1}{2}} \left(\mathbf{D}(\mathcal{G}) - \mathbf{A}(\mathcal{G})\right) \mathbf{D}(\mathcal{G})^{-\frac{1}{2}} \qquad [5]$$

For a connected graph, let the ordered eigenvalues of $\mathcal{L}(\mathcal{G})$ be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$. The relation $0 = \lambda_1 \leq \lambda_2 \leq \ldots, \leq \lambda_N$ holds. The second smallest eigenvalue λ_2 is called the Fiedler number, or Fiedler value of the graph, $\mathcal{G}$. The associated eigenvector is called the Fiedler vector. According to the entry signs of the Fiedler vector (+, or -) vertices of a graph can be grouped into two clusters, with each cluster having relatively stronger within-cluster connections and weaker between-cluster connections [25]. More generally, instead of considering binary connections between pairs, weights can be assigned to each edge such that $[\mathbf{A}(\mathcal{G})]_{ij} = w_{ij}$ only if $\{v_i, v_j\} \in E$ (otherwise $[\mathbf{A}(\mathcal{G})]_{ij} = 0$) to characterize the connection strengths. The associated degree matrix, $\boldsymbol{D}(\mathcal{G})$, and Laplacian, $\mathcal{L}(\mathcal{G})$, are still defined in the same way as previously presented. Fiedler number and clustering based on the Fiedler vector can also be defined and performed identically. In Fig. 2C we reported the correlation coefficient for each chromosome between the signed Fiedler vector and thresholded gene expression (RNAseq counts) where the threshold was selected to maximize the correlation.

The Fiedler vector can also be used for extracting TADs. The sign pattern a Fiedler vector can divide the chromosome to local units, which can be considered as TADs at relatively large scale. In order to get finer structures, we can recursively compute the Fiedler vector to split given obtained TADs until the Fiedler value of the region is higher than some threshold. The TAD structures in Figure S1E were obtained with this method.

Topological Domain Identification via the Graph Theory. We propose an algorithm to extract topological domains with the consideration of overall organization. At the first step, we consider the weighted graph with edge weights defined by the Toeplitz normalized matrix $\mathbf{H}_{\mathrm{TP}}$. The Fiedler vector, denoted by $\boldsymbol{v}^{(1)}$, of this graph is computed and segments the graph into two clusters relying on the signs of Fiedler vector entries. A number of locally compact structures are then given by the sets of vertices with the same sign on the largest range of continuous indices from i to j, i.e., the region $\mathcal{D}_{i\text{-}j}$ defined by $v_i^{(1)}, v_{i+1}^{(1)}, \ldots, v_j^{(1)}$ having the same sign. Each domain is characterized by highly local compactness compared with the neighboring domains. The sizes of these domains vary from 100kb to several megabases. They can naturally be defined as topological domains at the first layer. We can identify over 2000 TADs determined by the Fiedler vector derived from $\mathbf{H}_{\mathrm{TP}}$, which is consistent with the number reported in the literature. We will see that compared with the gene expression represented by RNAseq counts, regions within each domain generally behave in a binary manner, all active, or all inactive. Further, all domains with the same sign behave similarly. From the spectral graph theory, we know that domains with the same sign are labelled as in the same cluster, and have more contacts than a different cluster. This result is obtained from $\mathbf{H}_{\mathrm{TP}}$ that considers the overall contact organization of the chromatin, and is biologically consistent with the two chromosome compartments.

Topological domains are likely to have hierarchical structure, and domains at finer scale can have more compact local organization. In order to identify domains at finer scale, we propose to calculate the Fiedler vector for a obtained domain $\mathcal{D}_{i\text{-}j}$ to split the domain into sub-domains. The Fiedler value of the obtained domains are calculated and compared with a predefined threshold λ_{thr} to determine whether they are compact enough, or can further be split. Since finer domains rely more on the local compactness, these recursive steps will be performed on sub-blocks of the original matrix $\mathbf{H}$ in log scale,

instead of $\mathbf{H}_{\mathrm{TP}}$. Table 1 (in this document) provides a comparison of the proposed algorithm with the literature ones.

Gene Resolution Contact Map.

Fragment Level Contact Map

Each row of the fragment read file (.csv file) indicates a ligated pair of fragments from the genome, with the coordinates of both fragments given. This information enables us to know the exact locations of the contacts existing in the genome. Then, given any two regions defined by exact coordinates, the contacts of these two regions can be identified by searching over the fragment read file and counting the number of pairs that are located within these regions.

Construction of Gene Resolution Contact Map

Convectional Hi-C matrices are generated by binning fragments into fixed resolution bins (e.g., 100kb, 1Mb). However, we found that the analysis of gene spatial relations can be difficult to perform using this method due to the large variability of gene sizes (from several 1kb to 2Mb). This motivated us to use original fragment read pairs for the construction of the gene-resolution contact map. We generate a matrix with each bin representing a gene, the number of inter-contacts between a pair of genes is determined by summing up the fragment pairs located within the region defined by these genes. The diagonal entries are the number of contacts within each gene itself. This matrix gives us an exact contact relation for all genes. We denote this original gene-resolution contact matrix by $\mathbf{G}$.

Normalization of Gene Contact Matrix

In order to further study the gene level contact map, normalization is needed on the original contact matrix $\mathbf{G}$. A general trend exists in which longer genes tend to have more HindIII cutting sites, and therefore more contacts. We thus need to normalize the inter-contact of two genes i and j, given by $[\mathbf{G}]_{ij}$ by the lengths of the two contributing segments, namely,

$$[\mathbf{G}_{\mathrm{N1}}]_{ij} = \frac{[\mathbf{G}]_{ij}}{L_i L_j} \qquad [6]$$

where L_i and L_j are the lengths of genes i and j respectively. The value $[\mathbf{G}_{\mathrm{N1}}]_{ij}$ gives the density of contacts within a region of the area $L_i \times L_j$.

Next, similar to a Hi-C matrix, it is expected that nearby genes in linear base pairs distance are more likely to be ligated than distant pairs. This makes the gene-resolution contact matrix highly diagonally dominant and conceals the contact pattern embedded in the matrix. In order to present the relative contact probability with respect to the linear distance, we normalize the contact density $[\mathbf{G}_{\mathrm{N1}}]_{ij}$ by the mean contact of the gene pairs at the similar distance. Since gene distance (defined by the linear distance between the middle points of two genes) are not uniform, we binned the distances between all pairs of genes on the chromosome under study into M intervals (e.g.: $M = 1000$ in our analysis), then we calculate the mean contact density of the pairs within each interval, then normalize the contact density matrix by

$$[\mathbf{G}_{\mathrm{N2}}]_{ij} = \frac{[\mathbf{G}_{\mathrm{N1}}]_{ij}}{\mathcal{M}(g_i, g_j)}. \qquad [7]$$

where $\mathcal{M}(g_i, g_j)$ denotes the mean contact density of all the gene pairs with the distance in the interval where g_i and g_j are located.

Phase Plane for 4DN.

Construction of Phase Plane

The concept of phase plane is introduced to characterize features of 4DN, and it is a useful visual display for dynamical systems. For the 4DN, one axis is a measure of the form, denoted by $\mathcal{F}_1(\mathbf{H})$, the other is a measure of function, denoted by $\mathcal{F}_2(\boldsymbol{r})$, where $\mathcal{F}_1$ and $\mathcal{F}_2$ are the summarization functions for form and function, respectively, and $\mathbf{H}$ and $\boldsymbol{r}$ are the (normalized) Hi-C matrix and gene expression vectors at a given scale. At time instant n, the state of the form and function of a unit at this scale can be represented by a point $\big(\mathcal{F}_1(\mathbf{H}(n)), \mathcal{F}_2(\boldsymbol{r}(n))\big)$ on the phase plane.

- The phase plane in Figure 1G (right) was constructed by considering the Fiedler value (as $\mathcal{F}_1$) of Toeplitz normalized Hi-C matrix with desolate gene regions excluded (as $\mathbf{H}$) for form, and the mean square root (as $\mathcal{F}_2$) of the RNAseq counts of the chromosome regions associated with $\mathbf{H}$ (as $\boldsymbol{r}$) for function.
- The phase plane in Figure 1G (middle) was constructed by considering the Fiedler value (as $\mathcal{F}_1$) of each TAD in the Toeplitz normalized Hi-C matrix (as $\mathbf{H}$) for form, and the mean square root (as $\mathcal{F}_2$) of the RNAseq counts in each TADs for function.
- The phase plane in Figure 1G (left) was constructed by considering the Fiedler value (as $\mathcal{F}_1$) of the matrix from the fragment contact of a gene (as $\mathbf{H}$) for form, and the RNAseq counts of this gene (as $\boldsymbol{r}$) for function.

Domain of Form and Function

For the fibroblast cells, we collected the data for $N = 8$ time points. These points can define a domain of form and function (DFF) that is cell-type specific. In this work, we considered elliptic regions for DFF. To determine the ellipse, the two eigenvalues of the 2×2 covariance matrix obtained from the

	HMM	DP	Laplacian
Measure	Directionality index	Reads (normalized)	Fiedler value (normalized)
Key methods	State estimation with Hidden Markov chain	Optimization with dynamic programming	Spectral clustering with graph Laplacian
Characteristics	Method in original paper	Optimal in region total reads	Good physical interpretation
Resolution	Partially related with the length for DI	Related with γ	Related with λ_{thr}
Hierarchical	No	No	Yes
Robustness	Sensitive to initialization	Unique solution	Unique solution
Complexity	Moderate	High	Low

coordinates of the 8 points were calculated, with $\lambda_{\max}$ and $\lambda_{\min}$ denoting the large one and the small one, respectively. The center of the ellipse is given by the centroid of the 8 points. The orientation is determined by the direction of the eigenvectors. The long axis and the short axis are determined by $\delta\sqrt{\lambda_{\max}}$ and $\delta\sqrt{\lambda_{\min}}$, where δ is the parameter to define the allowance of the deviation. In our figures, δ was set to 3.

Graph Attached to the Genome. The components within the nucleus interact with each other in various manners and form a dynamic system. Defining these components and the interactions among them as a graph allows one to assign them with quantifiable values and to study the associated properties over time. Consequently, this dynamic view permits one to study the process of differentiation as well as nuclear organization and its effect on cell properties. Genes can be considered as nodes in such a graph. Their interacting relationship can be defined from several aspects. Gene expression data (RNAseq) provides the basis for constructing a transcriptome graph based on co-regulated genes. The Hi-C method probes the 3D architecture of whole genomes by coupling proximity-based ligation with massively parallel sequencing. Spatial proximity maps of the genome can be constructed using Hi-C at different resolution. Hi-C contact frequencies and the way they changes over time can be used to define the graph with physical contact relation.

Spectral Clustering. Spectral clustering is one of the most popular modern clustering algorithms. It can be solved efficiently by standard linear algebra techniques and very often outperforms traditional clustering algorithms. Assuming that data consists of N points, we can construct the weighted adjacency matrix $\mathbf{A}$ based on the similarities among these data points. We can then cluster these n data points into two clusters determined by the entry signs of the Fiedler vector, and multiple clusters can be obtained by clustering again on obtained clusters, or by the following general spectral clustering algorithm [26]:

- Compute the (normalized) Laplacian $\mathcal{L}$ of the weighted adjacency matrix.
- Compute the eigenvectors $\boldsymbol{u}_2, \ldots \boldsymbol{u}_{K+1}$ of $\mathcal{L}$.
- Let $\mathbf{U} \in \mathbb{R}^{N\times K}$ be the matrix constructed with column vectors $\boldsymbol{u}_2, \ldots \boldsymbol{u}_{K+1}$.
- Construct matrix $\overline{\mathbf{U}}$ by normalizing the rows to unit norm, that is, $[\overline{\mathbf{U}}]_{ij} = [\mathbf{U}]_{ij}/(\sum_{k=1}^{N}[\mathbf{U}]_{ik}^2)^2$.
- Take each row as an input of the k-means algorithm, and get clusters $C_1, \ldots, C_K$.

Spectral Clustering on Hi-C Matrix. Hi-C matrices naturally provide weighted adjacency relation between genome loci. The Toeplitz normalized Hi-C matrix is thus considered as $\mathbf{A}$, and spectral clustering is then applied to segment the loci into two or more clusters.

Spectral Clustering on Gene RNA-seq Profiles. N' genes with relatively larger temporal variability are selected out of the total 28000 possibilities. The correlation matrix of these N' gene expression variation profiles is computed and shifted by 1 to ensure nonnegative value for entries. Then this matrix is considered as the adjacency $\mathbf{A}$. This gene is then divided into clusters via the spectral clustering algorithm. The center of each cluster is then the mean of the centralized and variance normalized RNAseq counts.

Contact Matrix and Spatial Reconstruction for a Gene. A fragment read map is capable for illustrating how the fragments within a gene contact with each other. This super-resolution information enables us to (approximately) reconstruct the gene spatial structure. Several methods are proposed for this computation in literature. In order to maintain consistency with the mathematical framework and graph theory, we use the technique–Laplacian map to reconstruct these spatial structures [27], using the fragment read level contact information for a given gene and the enzyme cutting site locations within the gene region. The major steps are:

- Summarize the pair reads into a matrix according to the cutting site locations; namely, (ij)th-entry of the matrix is given by the total number of reads with two end coordinates within $(i, i+1)$ and $(j, j+1)$ respectively.
- Data transformation, such as square root transformation can be applied to shape the matrix data. Calculate the normalized Laplacian of the above matrix.
- Use the eigenvectors associated with the three smallest nonzero eigenvalues as the output coordinates.
- Spline interpolation can be applied to smooth the data points.

Criteria for Extracting Dynamic Genes. We establish several criteria to extract gene that have significant dynamics of transcription over time. These genes can be considered as genes that are highly expressed in Fibroblast cells and have response to serum stimulation. The criteria are listed here-below:

- At least 1.5 fold difference in expression between any two time points (q value < 0.01).
- Mean expression over time RPKM value larger than 1.
- Variance over time larger than 0.5.
- Ratio between variance over time and variance within each time point replicate larger than 1.5.

Extracting Possible Common Transcription Factors. We propose a method to extract possible common transcription factors of two genes, g_1 and g_2, via the motif data from ENCODE database and gene expression data.

Prerequisite:

- Genes g_1 and g_2 have similar transcription evolution over time (as given from RNA-seq data).
- The contacts between g_1 and g_2 (from Hi-C data) show similar trend with their transcription.

(These two conditions ensure they may share common transcription factors in a co-localization manner).

Extraction steps:

- Determine the location of g_1 and g_2 . Suppose they are given by chr-$n_1 : x_1 - x_2$, and chr-$n_2 : y_1 - y_2$ respectively.
- Extend the regions at both sides of the genes, in order to cover their regulation regions. Let the extended location of g_1 and g_2 be denoted by chr-$n_1 : x_1' - x_2'$, and chr-$n_2 : y_1' - y_2'$, typically, with $x_1' < x_1$, $y_1' < y_1$ and $x_2' > x_2$, $y_2' > y_2$.
- Search in the ENCODE database and get the CMTmotif set $\mathrm{MT}^{(1)} = \{mt_1^{(1)}, mt_2^{(1)}, \ldots, mt_{N1}^{(1)}\}$ in the region chr-$n_1 : x_1' - x_2'$, and the motif set $\mathrm{MT}^{(2)} = \{mt_1^{(2)}, mt_2^{(2)}, \ldots, mt_{N2}^{(2)}\}$ within the region chr-$n_2 : y_1' - y_2'$.
- Assign scores for each motif in the sets $\mathrm{MT}^{(1)}$ and $\mathrm{MT}^{(2)}$, given by $S^{(1)} = \{s_1^{(1)}, s_2^{(1)}, \ldots, s_{N1}^{(1)}\}$ and $S^{(2)} = \{s_1^{(2)}, s_2^{(2)}, \ldots, s_{N2}^{(2)}\}$. The scores can be associated with some related information of motifs within the region (e.g. times of presence). Taking the motifs with largest score, or using a thresholded score, we can extract small subsets $\mathrm{SMT}^{(1)} = \{smt_1^{(1)}, smt_2^{(1)}, \ldots, smt_{N1?}^{(1)}\}$ and $\mathrm{SMT}^{(2)} =$

$\{smt_1^{(2)}, smt_2^{(2)}, \ldots, smt_{N2'}^{(2)}\}$ from $\mathrm{MT}^{(1)}$ and $\mathrm{MT}^{(2)}$, with $\mathrm{SMT}^{(1)} \subset \mathrm{MT}^{(1)}$ and $\mathrm{SMT}^{(2)} \subset \mathrm{MT}^{(2)}$.

- Get the intersection set: $\mathrm{CMT} = \mathrm{SMT}^{(1)} \cap \mathrm{SMT}^{(2)} = \{cmt_1, cmt_2, \ldots, cmt_N\}$.
- Check the transcription level of genes that produce the transcription factors CMT. If a gene is not active in the current cell type data, remove the associated transcription factor from CMT.
- The remaining factors are considered as common transcription factors for these two genes.

Circadian Genes. In this paper we place a keen interest on circadian rhythms within mammalian cells; specifically the CLOCK gene pathways. In this biological pathway, CLOCK and ARNTL (BMAL1) genes transcribe proteins that form a heterodimer protein in the cytoplasm. This heterodimer protein then reenters the nucleus in areas with E-Boxes (5'-CACGTG- 3' and 5'-CACGTT-3') to activate genes that contain this motif structure. Key genes in this pathway include the cryptochrome genes (CRY1 and CRY2) and period genes (PER1 and PER2). The CRY and PER proteins transcribed from these genes dimerize, and in turn inhibit CLOCK and ARNTL transcription. This activation-inhibition pathway results in a continuous periodic cycle, equaling ∼24 hours in mammalian cells. The gene position within the nucleus and RNA activity of CLOCK, ARNTL, PER2, and CRY1 has been further explored in this paper.

FISH Gene Selection. The co-location, (mRNA) expression, and nuclear distribution of four circadian clock genes was investigated over time after G0 cell cycle arrest by serum starvation and circadian rhythm reset by dexamethasone treatment in cultured human fibroblasts. The four genes selected, Arntl, (aka Bmal1, Mop3), Per2, Clock, and Cry1 are located on four different chromosomes, allowing a relative degree of freedom for each gene to distribute within the nucleus.

The Bmal1:Clock heterodimer binding to the e-box cis-regulatory enhancer sequences of RORs, REV-ERBS, Per1/Per2, and Cry1/Cry2, increase their expression, and, in the case of ROR expression, increases the expression of Bmal1, or, for the REV-ERBs, blocks Bmal1 expression. The Per:Cry heterodimer, in conjunction with their phosphorylation by CK1e/d, blocks Bmal1:Clock activity [28]. It has been shown that Bmal1 is essential in this process [29] Per2 shows a very low Expressed Sequence Tag profile in human skin (Bioinformatics Lab, Wilmer Institute, John Hopkins University). If mRNA expression levels in the present experiment confirm this in fibroblasts, Per2 may serve as an internal control to determine if this 'silenced' gene tends to gather at the nuclear periphery.

We are investigating whether genes tend to be closer in physical space when they express in-phase (and share similar transcription machinery), while genes which express anti-phase tend to be further apart. Silenced genes should be statistically closer to the nuclear periphery [30].

Oligo Creation. Oligonucleotide tags (MYtags, MYcroarray, Ann Arbor, MI) for the genes Clock, Cry1, Per2 and Bmal1 were designed using MYcroarray's proprietary software [31]. Input target sequences were cut into candidate tags staggered every nucleotide along the target sequence. Tag length was adjusted between 43 and 47 nucleotides to fit into the narrowest possible Melting Temperature (Tm) range. Each tag candidate was blasted against the reference genome previously masked for the target sequences to prevent self-hit. A Tm of cross-hybridization was predicted for every blast hit. Tag candidates having no significant potential cross-hybridization at the hybridization temperature were selected and synthesized for use as FISH probes. Fluorescent labeling of the probes were performed in our laboratory using techniques adapted from Murgha YE, et al (proprietary). The choice of fluorescent labels for each probe set was based on the dichroic filters and lasers available to us and advice from Zeiss and MYcroarray representatives on excitation/emission separation and efficiency of labeling. Atto488, MAX, Atto 550 and Atto633 were chosen from a list of available fluorophores (ATTO-TEC GmbH; FreedomTM dyes IDT®).

The final dye-probe concentrations were acceptable, averaging 8 pmoles/μL.

FISH Hybridization and Imaging. BJ ATCC® CRL-2522 homosapien skin fibroblasts were plated in large culture dishes (for Hi-C), and 6 well culture plates containing 10 mm circular coverslips (for FISH) in complete media (DMEM, 10% FBS, NEAA, antibiotic/antimycotic) overnight, followed by G0 cell growth arrest by serum starvation for 48 hrs [32, 33] in FBS-deficient media. This was followed by a 1 hour treatment with the addition of 100 nM dexamethasone [34].The control sample was treated with carrier, 0.001% ethanol. Removal of the dexamethasone/carrier was accomplished by two washes in PBS. Complete media was added, and a total of 16 time points were collected; at hr 0 and at 4 hr intervals afterward, totaling 60 hours and over 2 full circadian cycles.

Fluorescent In Situ Hybridization (FISH): The 10mm coverslip fibroblasts for FISH were preserved using 4% PFA following the 'Support Protocol 1: Preparing Tissue Culture Cells for Interphase FISH)' of Beliveau et al [35]. After 24 hrs stabilization in 2xSSCT/50% formamide, the cells were dehydrated in graded 70%, 90%, 100% cold ethanol at 5 minutes per step. Storage was in 100% ethanol at -20°C.

Samples for hybridization were re-hydrated with reverse graded cold ethanol (100%, 90%, 70%) followed by a 1 min wash in PBS at room temperature. The 'Alternate Protocol 1: Fast 3D-FISH Using Oligopaint Probes' [32] was then followed, with modifications of hybridization buffer volumes and probe amounts to reflect the 22x22mm to 10mm round coverslip area difference. Individual probe volumes were mixed, reduced to dryness in a 40°C heating block and reconstituted in 1μL of DNase free water and 7μL of hybridization buffer per sample.

Imaging was performed on a Zeiss LSM 710 confocal microscope using a 63x objective. Settings for laser intensities, pinhole settings, photomultiplier gains, etc. were stored and used consistently for all samples.

Manual selection of nuclei for analysis was based on clear labeling of two areas for each of the four probes at each time point. Processing to reduce signal cross-talk and obtain distance measurements between the labeled probes, nuclear centroid, nuclear envelope and radial distance was performed with image processing techniques. See Figure S4DI for the entire workflow, which consists of modules with different roles. Considering the data size and the requirement of processing time, simple but robust algorithms were applied for each module. In what follows we briefly present the functions of these modules and the used algorithms, starting with the raw image data I_O obtained from the microscope.

Pre-processing: The collected data is a three-dimensional data such that I_O with width w and height h, and n denoting the number of observed slices, and C denoting the number of color channels. In the multispectral mode of our system, five channels are involved in the observation, namely, $C = 5$. One channel is used for forming the nucleus shapes, while the

others are used for labeling four pairs of genes. Considering the natural disk-like shape of fibroblast cells we generate the maximum intensity projection I of the image I_O, to reduce the data dimension.

Subregion extraction: Among a number of cells in the observed scenario, only several one are considered as "good" by observing the quality of the cells. The good cell center information, i.e., (X, Y) coordinates of the center of the good cells, are currently provided manually. We then generate a series of square sub-images I_S centered at these centers and with side length R_S.

Threshold determination: Since we need to segment the nuclei away from the background in the following steps, it is necessary to determine a segmentation threshold ν. In order to avoid manually selection of the threshold in each scenario, this module provides us with a value of nv by selecting a point that is located between two peaks of the histogram of the channel associated with nucleus coloring dye (DAPI).

Endmember extraction: In this experiment's context, it is reasonable to make the assumption that pure dye spectra exists in images. The nucleus dye (DAPI) has little crosstalk with the others dyes, and gene positions are overlapped with low probability if only several genes are colored. Due to these conditions, we can assume the existence of pure pixels. Furthermore, we can assume that pure pixels have more significant contribution on its observation channel than the overlap from the others dyes. The endmembers M are selected by the brightest pixel in each channel.

Main cell segmentation: The subimages I_S may contain not only the cell to be analyzed but also parts of other cells that lie in the current square. We thus need to determine a cell shape mask to extract the desired cell and to eliminate the influence from the others in the cell shape determination. This mask will also be used in the following unmixing step to reduce the number of pixels to be processed. In this module several operations are performed, e.g.: thresholding the subimage, labeling the connected regions, and selecting the major region of interest. Finally, image morphology operations are also applied to eliminate small isolated objects.

Data unmixing: With the estimated endmember matrix and the observations at hand, we are enabled to conduct the linear unmixing by solving the nonnegativity constrained least-square problem.

Gene position determination: After evaluating the abundance distribution A, we localize the positions $(\tilde{x}_i, \tilde{y}_i)$ of two genes on each channel by seeking two maximum intensity points in the abundance maps.

Ellipse coordinate establishment: We model the shape of a nucleus by an ellipse. The long axis and short axis of the ellipse are determined by PCA performed on the nucleus shape mask. The genes are then marked in the obtained ellipse.

Cell alignment: Since the cells might be rotated and flipped, it is better to align the cells to facilitate the comparison and further computation. After aligning the nuclei by setting their long axis to the vertical direction, we also flip/rotate cells to so that their gene positions are maximally matched with a reference nucleus. Then the coordinates of genes are denoted by (x_i, y_i).

Measures Used for Characterizing Gene Locations in FISH Data.

- **Mean closest distance (MCD) between two genes**: MCD is defined as the shortest distance between two different gene locations (given that there are two homologous genes per target, there are 4 different "distances" between two gene locations; take the shortest distance). MCD is meant to be used as a measure of relative spatial interaction between two genes at a given time point, with a lower distance correlating to a higher likelihood of genomic contact.
- **Distance Matrix**: The matrix with entries denoting the distance in Euclidean space between two genes.
- **Index of Stability**: Derived from the Fiedler number of normalized laplacian of the distance matrix. This can be computed for any size distance matrix (i.e. number of genes) to characterize relative structure. Our analysis used the normalized laplacian to create a scale invariant measure of stability.

Rationale for MCD. MCD was developed as a simple measurement of inter-gene distance. Genes measured with MCD in our analysis are biallelic, meaning each gene has a copy that also is transcriptionally active. As a result of these gene copies, when measuring the distances between two genes there are 4 possible combinations. The shortest distance derived from the 4 possible combinations, MCD, does not take into account the gene copies, but was chosen to give a general picture of how close two genes may be at certain time point.

Cell Normalization. After further analysis of our data we discovered that nucleus area over time had a periodic nature as well, with peaks aligning with our peaks in MCD data. It is believed that this is largely due to natural fluctuations in cell size in the cell cycle [36], but to eliminate this variable from our analysis we also reviewed our data with normalized nucleus size. To normalize our data we calculated the average major axis (10um) and minor axis (5.4 um) at the first time point and translated each nucleus to fit this form, while preserving gene position within the nucleus. We then recompiled our data to see how our findings were affected.

Fig. S3E-F show MCD between CLOCK and PER2 genomic locations, much like Fig. 3B, with normalized data. A cyclic pattern can still be seen, on similar amplitude, with a slightly greater period (28 hrs). In our relatively small time window, the overall relationship between MCD and RNA transcription levels remains the same when cells are normalized.

1. Wang L, Li L, Cheng G, Zhou H (2011) Cell cycle regulation of human foreskin fibroblasts. African Jounral of Biotechnology 10:11797–11801.
2. Balsalobre A, et al. (2000) Resetting of circadian time in peripheral tissues by glucocorticoid signaling. Science 289:2344–2347.
3. FastQC (2011) A quality control tool for high throughput sequence data (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc).
4. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short dna sequences to the human genome. Genome Biology 10:R25.
5. Trapnell C, Pachter L, Salzberg SL (2009) Tophat: discovering splice junctions with rna-seq. Bioinformatics 25:1105–1111.
6. Roberts A, Pimentel H, Trapnell C, Pachter L (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27:2325–2329.
7. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biology 12:R22.
8. Trapnell C, et al. (2010) Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28:511–51.
9. Trapnell C, et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology 31:46–53.

10. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9.
11. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by rna-seq. Nature methods 5.
12. Wang Z, Gerstein M, Snyder M (2009) Rna-seq: a revolutionary tool for transcriptomics. Nature reviews Genetics 10:57–63.
13. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4.
14. Whitfield ML, et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Bio Cell 13:1977–2000.
15. Iyer VR, et al. (1999) The transcriptional program in the response of human fibroblasts to serum. Science 283.
16. Tobey RA, Valdez JG, Crissman HA (1988) Synchronization of human diploid fibroblasts at multiple stages of the cell cycle. Experimental Cell Research 179.
17. Koike N, et al. (338) Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. 2012 349–354.
18. Hughes ME, Hogenesch JB, Kornacker K (2010) JTK_CYCLE an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. J. Biol. Rhythms 25:372–380.
19. Anafi RC, et al. (2014) Machine learning helps identify chrono as a circadian clock component. PLoS biology 12:e1001840.
20. Chang HY, et al. (2004) Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS biology 2:E7.
21. Nagoshi E, et al. (2004) Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. Cell 119:693–705.
22. Paulsen M, et al. (2014) Use of bru-seq and bruchase-seq for genome-wide assessment of the sysnthesis and stability of rna. Methods 67.1:45–54.
23. Mesbhi M, Magnus E (2010) Graph theoretic methods in multiagent networks (Princeton University Press).
24. Chung FR (1997) Spectral graph theory (American Mathematical Soc.) Vol. 91.
25. Shi J, Jitendra M (2000) Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22:888–905.
26. Andrew N, Michael Y, Jordan I, Weiss Y (2002) On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems 2:849–856.
27. Cayton L (2005) Algorithms for manifold learning., (Univ. of California at San Diego), Technical report.
28. Ko CH, Takahashi JS (2006) Molecular components of the mammalian circadian clock. Human Molecular Genetics 15:R271–R277.
29. Bunger MK, et al. (2000) Mop3 is an essential component of the master circadian pacemaker in mammals. Cell 103:1009–1017.
30. Shaklai S, Amariglio N, Rechavi G, , Simon AJ (2007) Gene silencing at the nuclear periphery. FEBS Journal 274:1383–1392.
31. Murgha YE, Rouillard JM, , Gulari E (2014) Methods for the preparation of large quantities of complex single-stranded oligonucleotide libraries. PLoS One 9:e94752.
32. Beliveau BJ, et al. (2012) Versatile design and synthesis platform for visualizing genomes with oligopaint FISH probes. Proceedings of the National Academy of Sciences 109.
33. Griffin MJ (1976) Synchronization of some human cell strains by serum and calcium starvation. In Vitro 12:393–398.
34. Balsalobre A, Damiola F, Schibler U (1998) A serum shock induces circadian gene expression in mammalian tissue culture cells. Cell 93.
35. Beliveau BJ, Apostolopoulos N, Wu CT (2014) Visualizing genomes with oligopaint FISH probes. Current Protocols in Molecular Biology pp 14.23.1–14.23.20.
36. Cookson N, Cookson S, Tsimring L, Hasty J (2010) Cell cycle-dependent variations in protein concentration. Nucleic acids research 38.3:2676–2681.

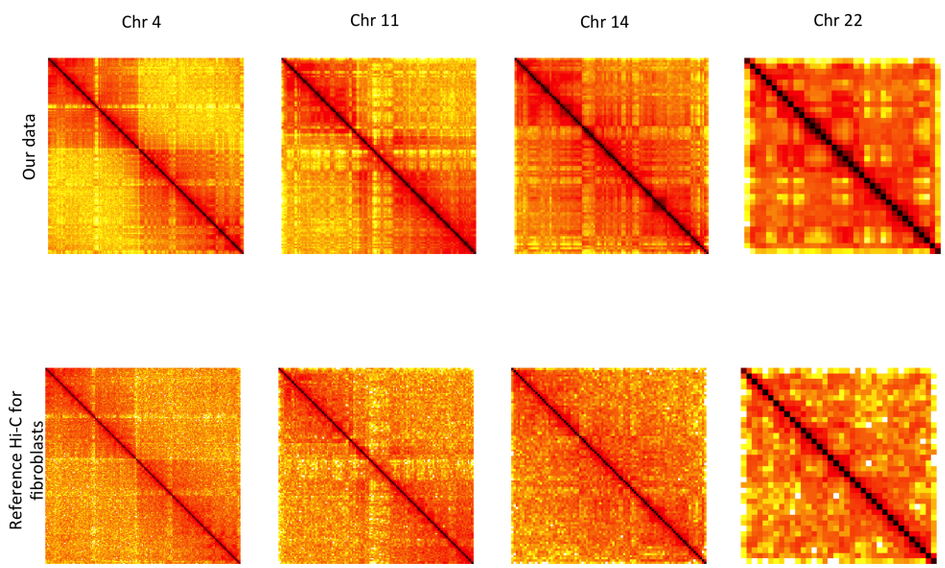


Fig. S1A. This figure shows the reproducibility of our data with previously published Hi-C data at 1M resolution for fibroblasts (Genome Research 23:260-269, 2013) for chromosomes 4, 11, 14, 22. It can be observed that our data has superior quality with clearly structured organization.

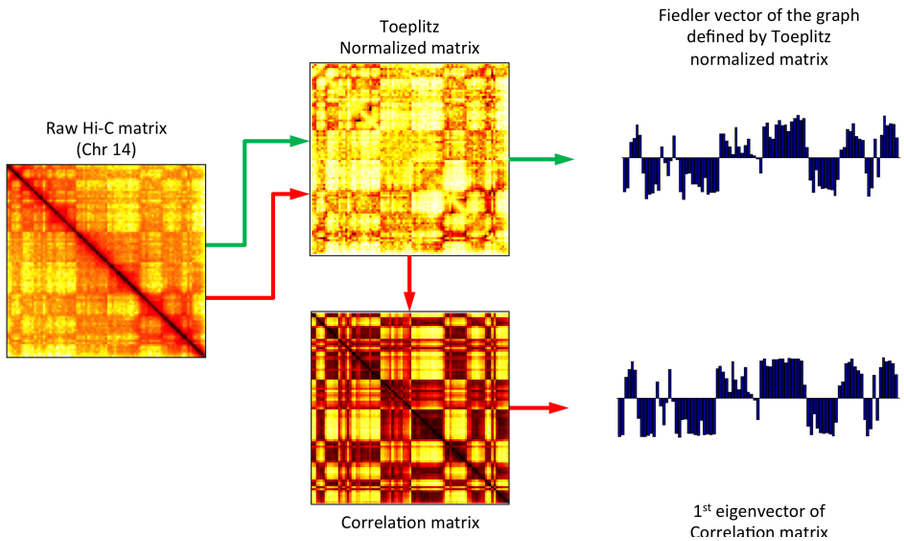


Fig. S1B. This figure shows how our method of chromosome partitioning compares to previously published methods from (Lieberman-Aiden et al. 2009). Green arrows represent our method, red arrows represent Lieberman-Aiden et al method. In both methods the raw Hi-C Matrix is Toeplitz normalized to highlight long distance Hi-C contacts. (Lieberman-Aiden et al. 2009) then creates a correlation matrix based on the Toeplitz normalized matrix. The first eigen vectors of this matrix acts to partition the chromosome into two compartments. Our method partitions the chromosome directly from the Toeplitz normalized matrix by calculating the fiedler vectors of this matrix. The end results from each method are comparable, but Fiedler vector partitioning has clear physical interpretation and is more flexibility for extra analysis (such as TAD identification and 3D reconstruction).

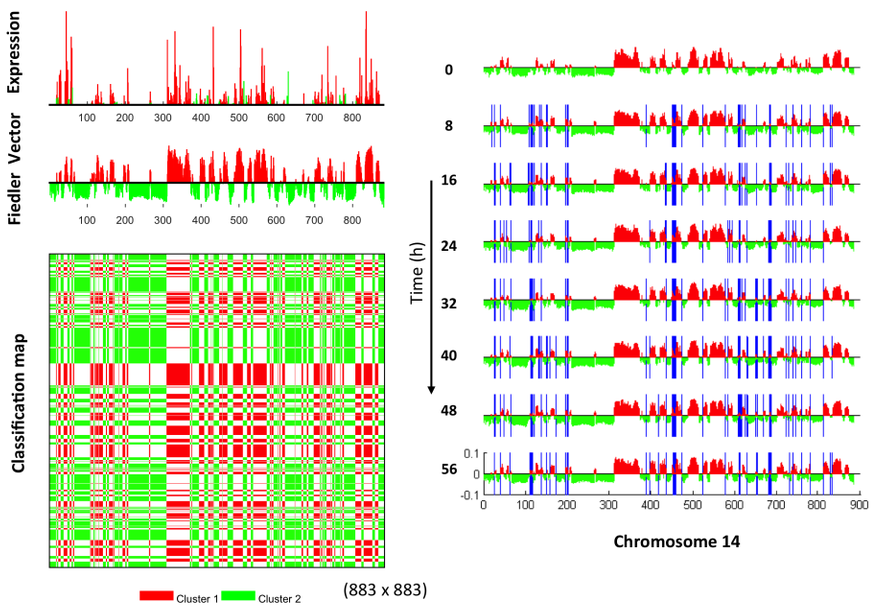


Fig. S1C. Compartment identification with Fiedler vector for Hi-C at 100kb resolution for chromosome 14. Left : Comparison of gene expression levels (represented by average RNA-seq counts over time, over 100kB binned region; top) and the clustering matrix results on chromosome 14 at 100kb resolution (bottom). The Fiedler vector divides the bins by the sign of its value(+/-). Clustering results are projected onto gene expression and annotated by colors. It can be observed that clustering from Hi-C divides the bins into transcriptionally active and inactive regions. The Classification Map shows the long-range connection of clusters. Right part: Fiedler vectors of chromosome 14 Hi-C over time. Bins with different clustering results compared with time T = 0h are denoted by blue vertical lines. Note that this large scale structure is stable over time, and slight changes only appear at cluster boundaries.

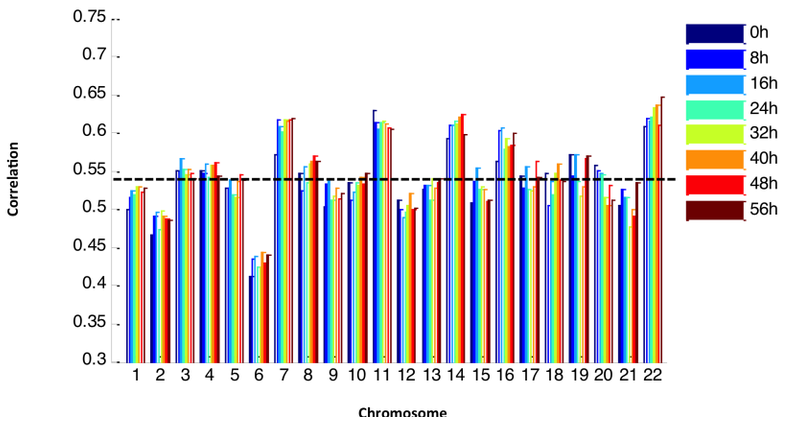


Fig. S1D. Correlation coefficients between thresholded Fiedler vectors and RNA-seq counts for chromosomes 1-22 over time, and the average coefficient across chromosomes is 0.5421 (dashed black line), while the average coefficient from the principle component analysis is 0.5171.

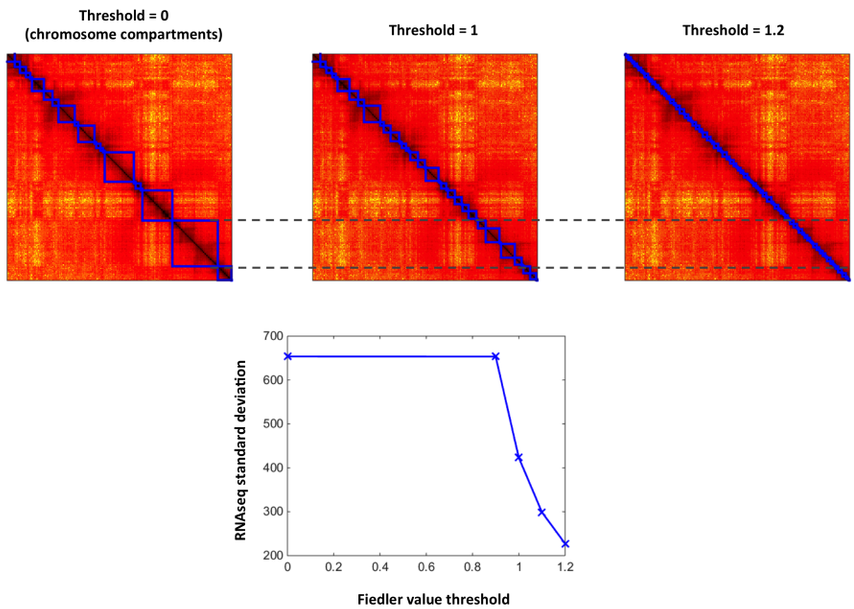


Fig. S1E. Upper panel: Hierarchical identification of TADs with graph Laplacian, illustrated by the region of bins 229 ? 491 Hi-C matrix of chromosome 4, at 100kb resolution. With the increase of Fiedler value threshold, we obtained finer TADs that are inherited from the previous level. Lower panel: standard deviation of RNAseq within TADs with the increase of Fiedler number threshold (decrease of TAD sizes).

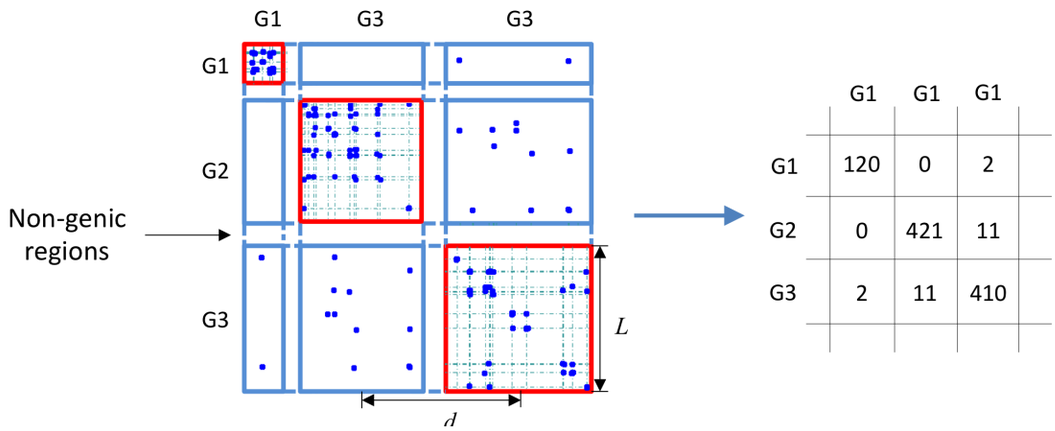


Fig. S1F. Individual gene fragment contact and a gene-resolution contact map. The left dot plots shows the fragment contact maps of three genes: G1: RBM23(Chr 14:23369853-23388396), G2: SLC7A8(Chr 14:23594503-23652869), and G3: THTPA(Chr 14:23980968-24048009). The red squares along the diagonal enclose all interaction fragments of each gene with defined genomic locations, dashed green lines indicate Hind-III cutting sites. The off-diagonal rectangles display the contact between gene pairs from Hi-C data. The distance d between two genes is defined as the linear distance between their centers. The gene size is denoted by L. The right plot shows gene-resolution contact matrix generated from the left contact map. The values are obtained by summing up the number of fragment reads in the associated regions.

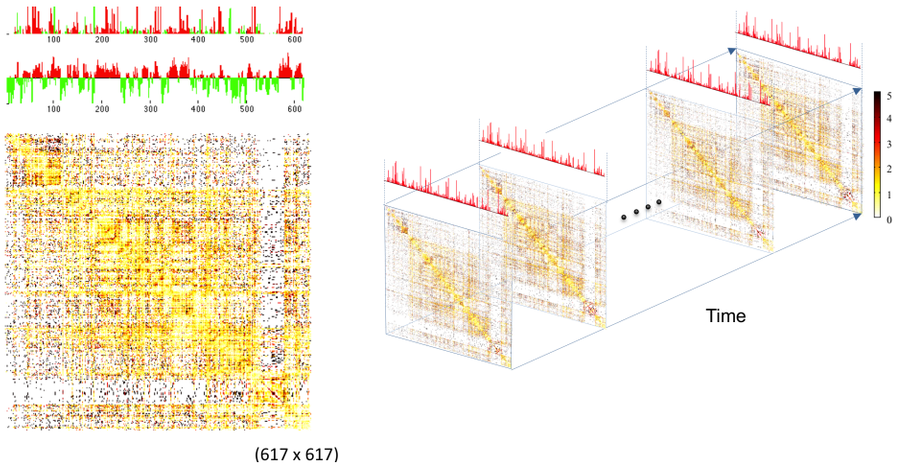


Fig. S1G. The left figure shows the normalized gene-resolution contact matrix averaged over time for all genes on chromosome 14. The bins (genes) are normalized by gene length and its expected contact frequency (see supplementary for the normalization method). The red-green bar plots above the gene-resolution contact matrix show the RNA-seq counts for each gene (top) and the Fiedler vector of this normalized gene contact matrix (second bar plot from top). The correlation coefficient between these two vectors is 0.34 less than that in Fig. S1D. Right: Gene-resolution contact maps over time for all the genes on Chr 14. The red bar plots show the expression vector.

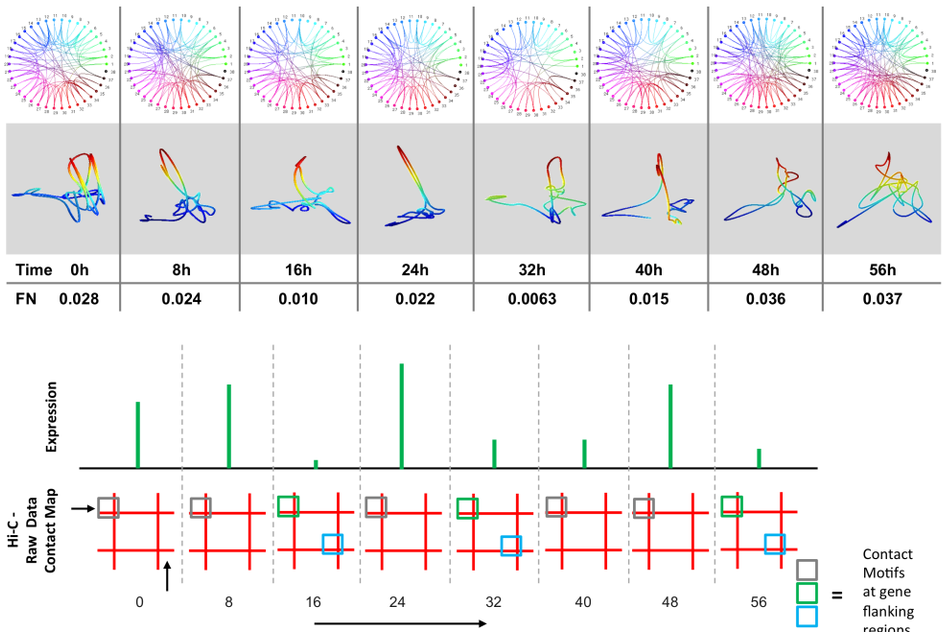


Fig. S1H. Top: The circus plot of interactions, modeled 3D structure, time, and Fiedler number for CLOCK. Bottom: Fragment interaction dynamics of PER2. PER2 is a core circadian gene with a period of 24h (depicted in this figure). The evolution of PER2 expression, measured by RNAseq, plotted along the top of the figure (green). Simplified fragment level Hi-C contact maps of each time points are depicted by red grids. The regulatory regions containing the significant patterns are denoted with grey and blue rectangles within the red grids. Note that the times points with high transcription have similar contact patterns in the regulatory regions.

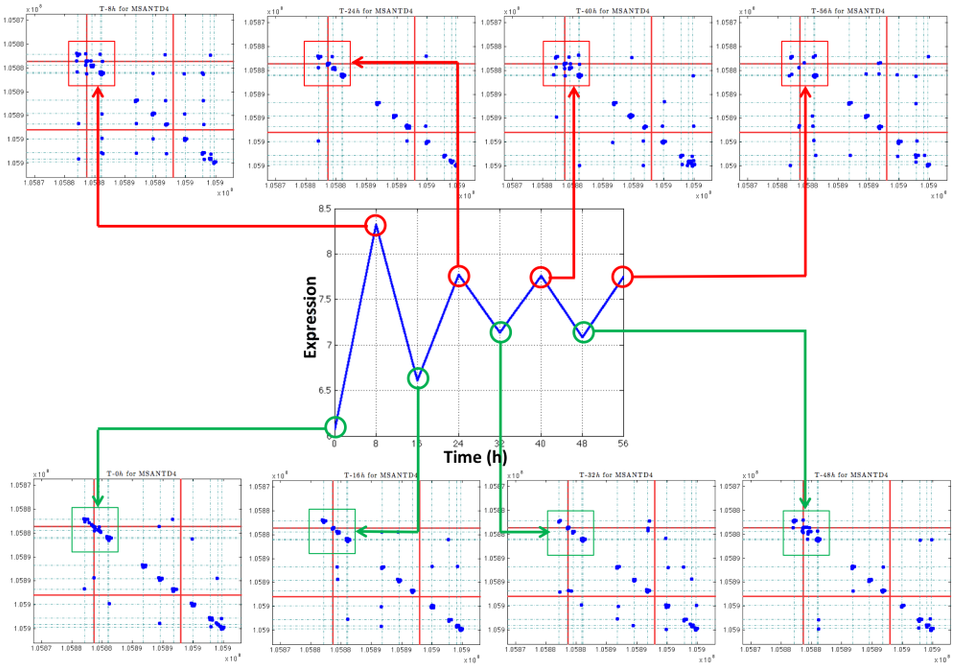


Fig. S1I. MSANTD4 is a gene determined to have a period of 16 hours. Its RNA-seq evolution is plotted in the middle of the figure. Hi-C matrices of each time points are plotted correspondingly. Gene boundaries are denoted by the red vertical and horizontal lines. Hind-III cutting sites are denoted by non-uniform teal grids. The regulation regions with specific changing patterns are denoted in red and green rectangles. Note that the time points with high transcription have similar contact patterns (red rectangles), while those with low transcription have distinctly different pattern (green rectangles).

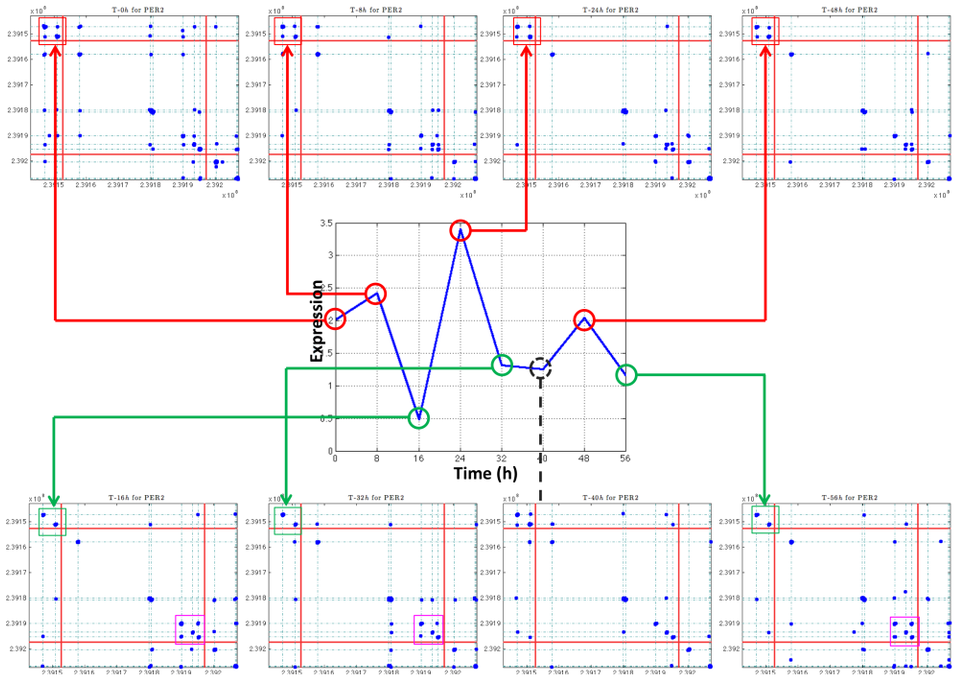


Fig. S1J. PER2 is a core circadian gene with a period of 24 hours. Its RNA-seq evolution is plotted in the middle of the figure. Hi-C matrices of each time points are plotted correspondingly. Gene boundaries are denoted by the red vertical and horizontal lines. Hind-III cutting sites are denoted by non-uniform teal grids. The regulation regions with specific changing patterns are denoted in red and green rectangles. Note that the times points with high transcription have similar contact patterns (red rectangles), while those with low transcription have a distinctly different pattern (green rectangles), except for the outlier time T = 40h. Furthermore, note that the pink rectangles emphasize another replicated contact pattern for the low transcription time points.

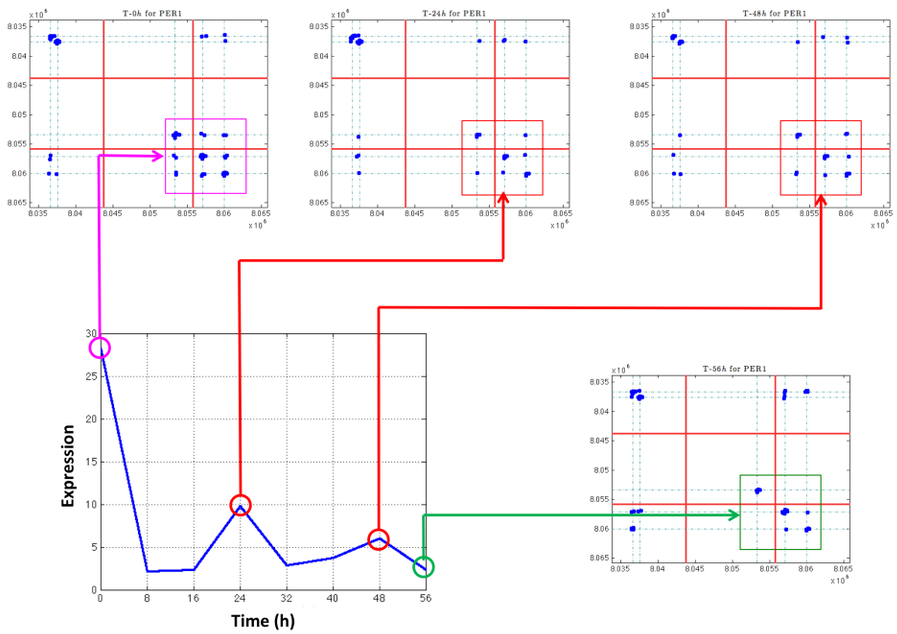


Fig. S1K. PER1 is a gene that has significant response to DEX. Its RNA-seq evolution is plotted in the bottom-left of the figure. Fragment level Hi-C matrices of selected time points are plotted correspondingly. Gene boundaries are denoted by the red vertical and horizontal lines within the Hi-C matrices. Hind-III cutting sites are denoted by non-uniform teal grids. Note that replicated patterns in red rectangles exist at high transcription time points T = 24h and 48h, which differ from the pattern at T = 56. Furthermore, note the more contacted pattern at T = 0h, which could be caused by the DEX. Contact matrices of the other time points are not shown here as they do not show clear particular patterns.

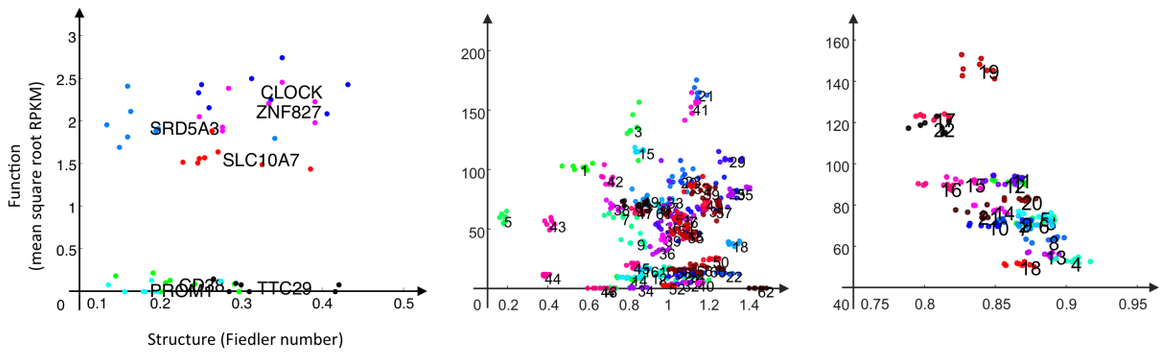


Fig. S1L. Phase plane characterization of genome form-function at different scales. Left: Phase plane for selected genes on chromosome 4. Middle: Phase plane for selected topological domains on chromosome 4. Right: Phase plane for chromosomes 4.

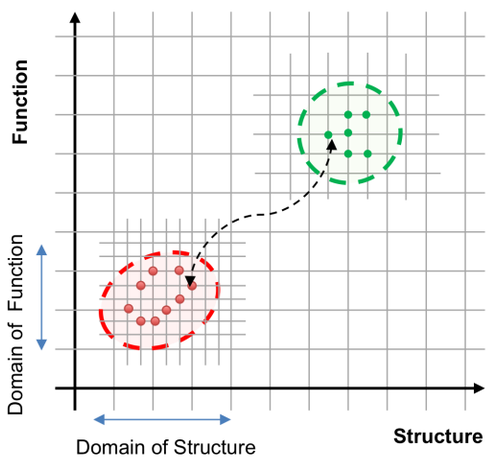


Fig. S1M. A 4DN phase plane illustration shows how genomic regions (at an arbitrary scale) of specific cell types may occupy different regions. The red shaded ellipse depicts a given cell type's phase plane domain of structure and function (DSF), derived from multiple time points of S-F positioning. The green shaded ellipse depicts another cell type's DSF (or the DSF of an abnormal cell). The arrows depict a path showing potential shifts from one DSF to another with changes in cell state (differentiation or reprogramming) or with disease progression.

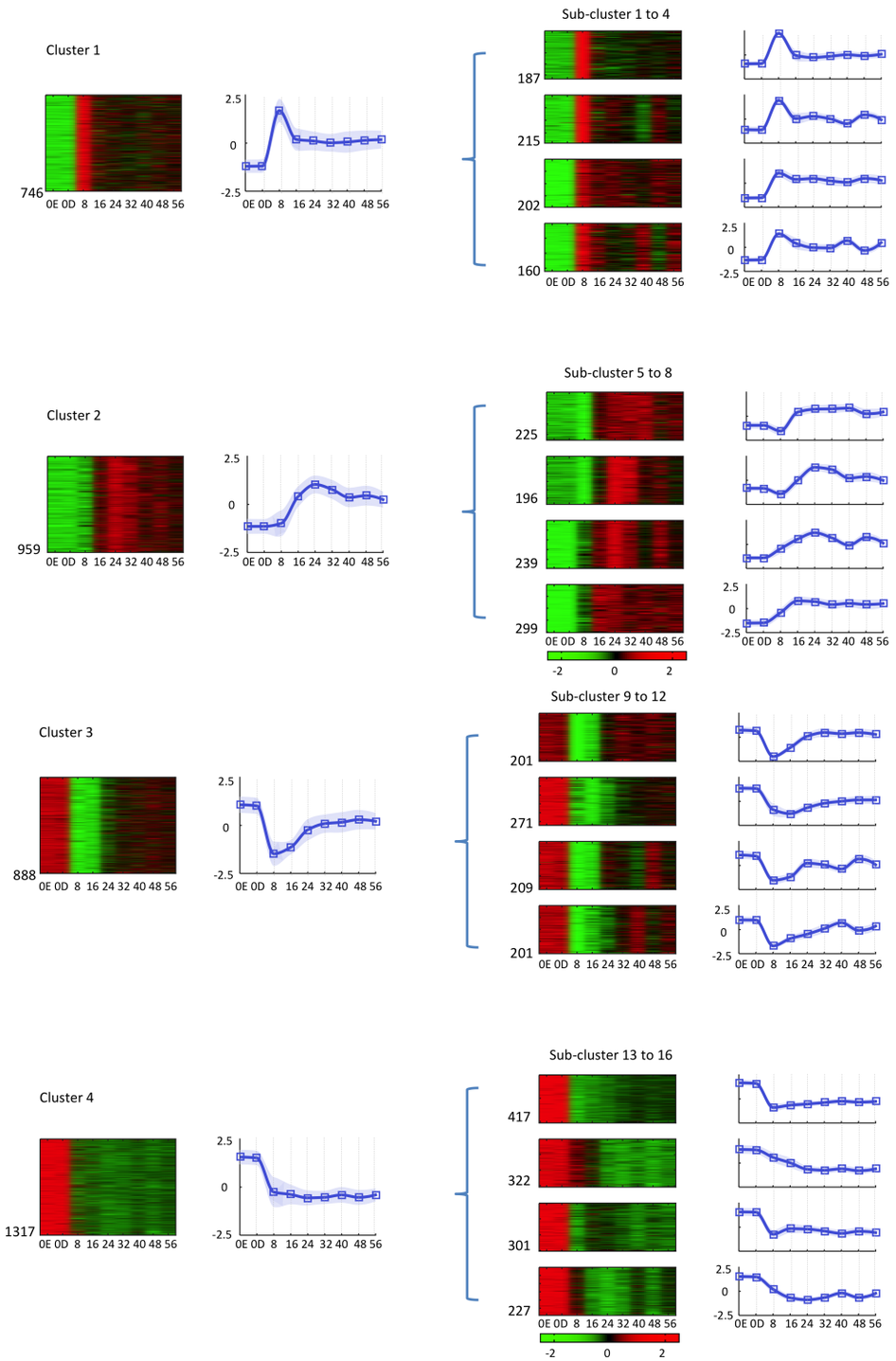


Fig. S2A. Expression patterns of the 7786 genes identified by Spectral clustering. We identified 7786 genes with high temporal dynamics out of the total 23615 genes picked in RNAseq, then cluster these genes into 8 clusters based on expression profiles. The heat maps of the 8 clusters are illustrated in the 1st column. The average normalized values of each cluster are illustrated with line plots in the 2nd column (blue line), with the shaded area denoting the standard deviations, and boxes indicate sampling time points. The time points of sampling are horizontally shown at the bottom on each panel, D: time zero samples treated with Dexamethasone for 1 hour without exposure to serum (contrast to E) , E: indicates base line time zero control without exposure to Dexamethasone & serum , and numbers indicate hours after serum stimulation from time zero. The 8 clusters are then clustered into 4 sub-clusters to investigate finer expression patterns?with heat maps shown in the 3rd column, and the average normalized values of each cluster are illustrated in the 4th column. Color: green represents expression levels below the mean, red above the mean, dark: mean level, the scale is shown in a line plot next to each heat map.

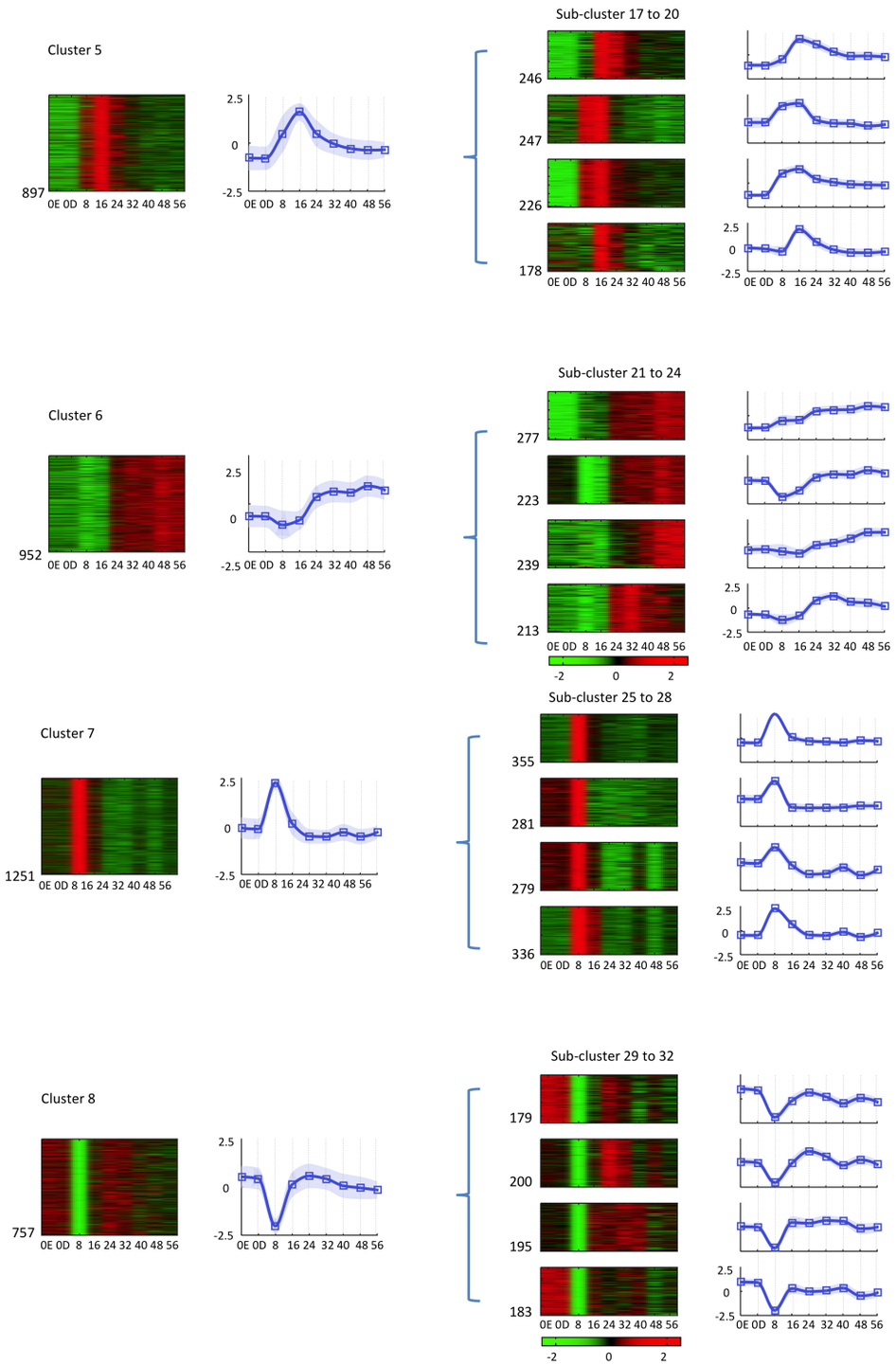


Fig. S2A. (Cont'd). Expression patterns of the 7786 genes identified by Spectral clustering.

Gene ontology enrichment analysis of the 32 clusters

Cluster	Main GO term significant enrichment (FDR < 0.05)
1	Protein folding, RNA Processing
2	RNA processing, ribosome biogenesis
3	Sensory perception of mechanical stimulus
4	Post-translational protein modification
5	Biopolymer metabolic process, amino acid activation
6	Cell cycle, G2/M transition, cell division
7	Translation, macromolecule biosynthetic process
8	Cytoskeleton organization and biogenesis
9	Response to virus
10	No significant enrichment
11	No significant enrichment
12	Post-translational protein modification
13	Developmental process, anatomical structure development
14	Wound healing, regulation of signal transduction
15	Organ morphogenesis, system development
16	Regulation of transcription (DNA-dependent)
17	DNA replication, DNA repair, cell cycle process
18	Macromolecule localization, protein transport/localization
19	Activation of MAPKK activity
20	No significant enrichment
21	Cell adhesion
22	No significant enrichment
23	No significant enrichment
24	Cell division, mitosis
25	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process, transcription
26	Signal transduction, cell communication
27	Biopolymer metabolic process, post-translational protein modification
28	Macromolecule metabolic process, post-translational protein modification
29	No significant enrichment
30	DNA metabolic process
32	Cell cycle, cell cycle checkpoint, DNA repair

Fig. S2B. Summary of gene ontology (GO) analysis for the 32 sub-clusters. Genes from each sub-cluster are subjected to GO analysis. Main significant GO terms with false discovery rate ¡ 0.05 are summarized in this figure.

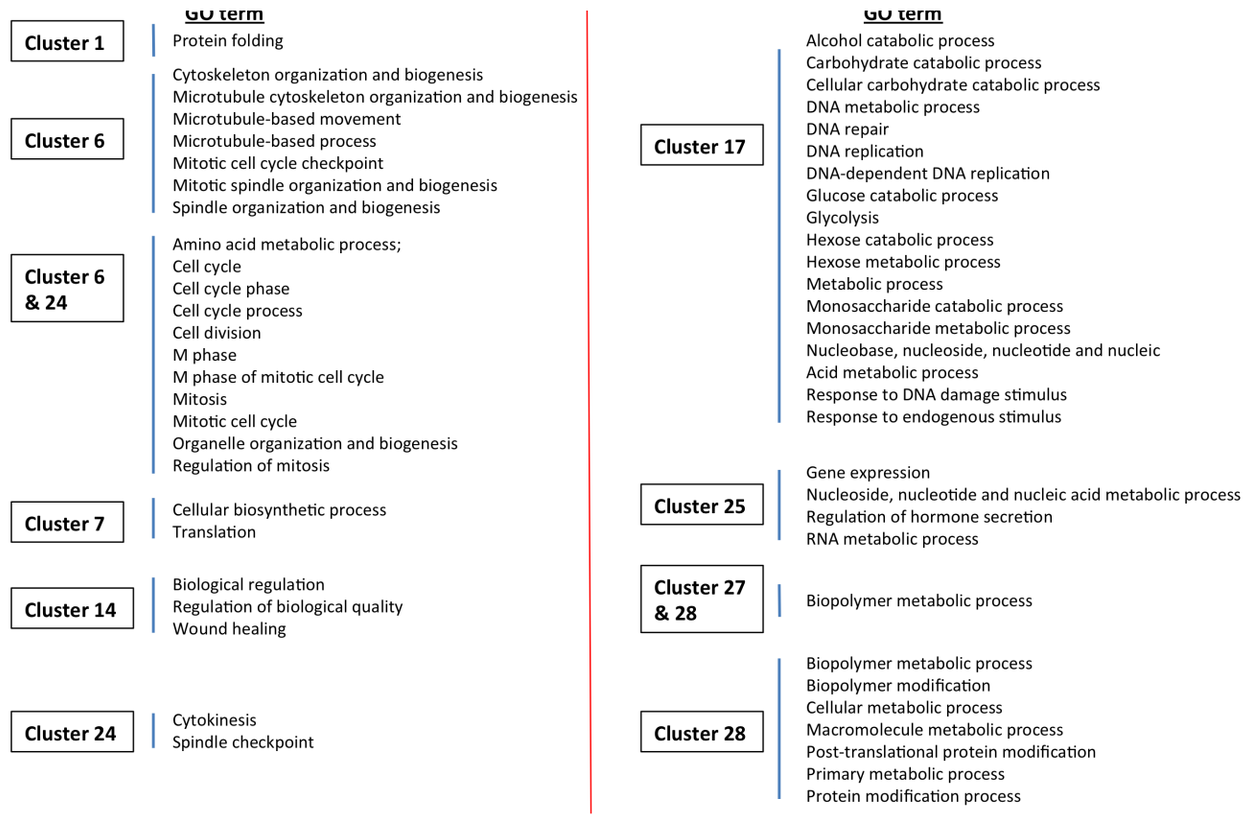


Fig. S2C. Summary of significant biological process terms from gene ontology (GO) analysis (Bonferroni $P < 0.05$) for the most significant clusters.

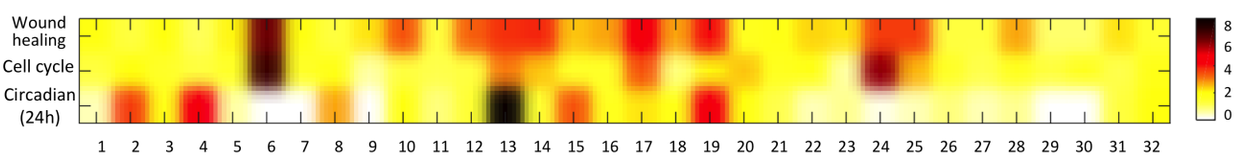


Fig. S2D. Distribution of the genes (in percentage) from three biological modules: wound healing, cell cycle, and circadian clock over the 32 sub-clusters. Note that cell cycle genes most concentrate on clusters 6, 17, and 21, consistent to the biology functions identified for these clusters. This shows the effectiveness of our clustering.

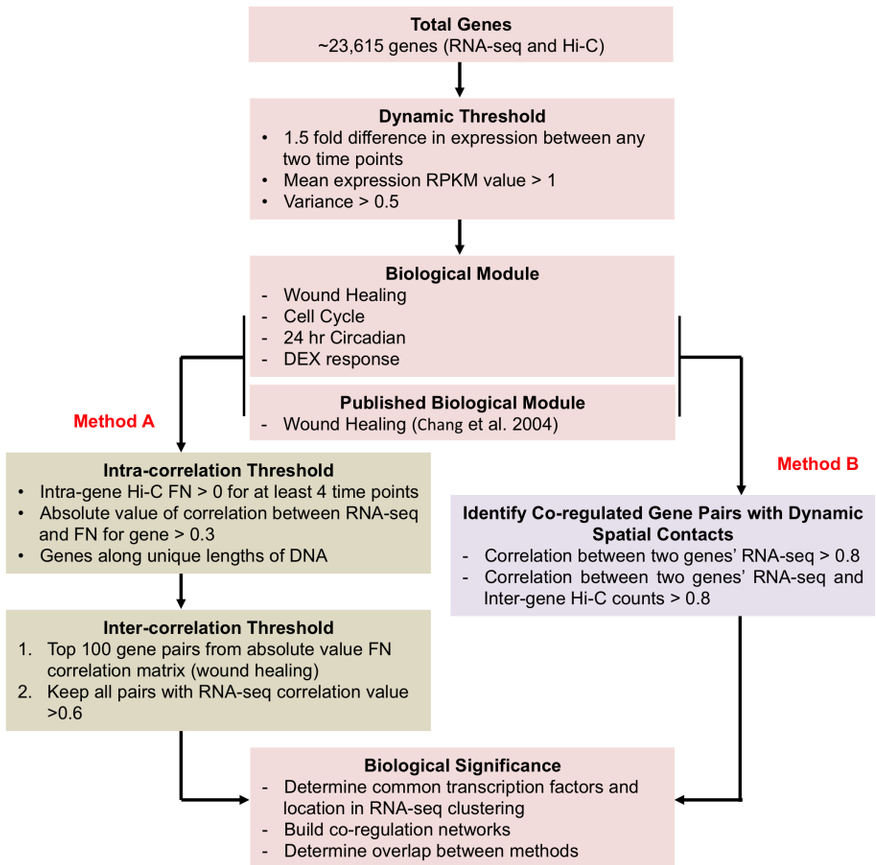


Fig. S3A. Algorithm for extraction of Dynamic Form Function correlated gene pairs.

Chromosome	# of intra-correlated genes	# of inter-correlated gene pairs analyzed	av. # of shared motifs	av. # of shared motifs (random)	Statistically significant?	P-value
1	259	50	6.86 ± 2.52	4.54 ± 2.07	1	2.37e-15
2	202	44	6.98 ± 2.04	4.69 ± 2.12	1	8.88e-13
3	165	52	8.02 ± 1.59	4.72 ± 2.20	1	3.80e-27
4	108	55	7.35 ± 1.39	4.59 ± 2.32	1	1.72e-18
5	123	50	6.68 ± 2.37	4.28 ± 2.26	1	9.55e-14
6	121	35	5.80 ± 2.23	3.96 ± 2.10	1	2.29e-7
7	129	50	5.70 ± 2.34	4.57 ± 2.09	1	1.46e-4
8	100	53	6.53 ± 2.19	4.49 ± 2.12	1	3.80e-12
9	117	39	7.92 ± 2.04	4.61 ± 2.14	1	7.94e-22
10	122	48	6.75 ± 1.82	4.32 ± 2.17	1	1.20e-14
11	154	54	5.63 ± 2.18	4.30 ± 2.01	1	1.19e-6
12	153	40	6.80 ± 2.17	4.48 ± 2.01	1	3.07e-13
13	42	30	6.03 ± 2.51	4.87 ± 2.45	1	0.01
14	80	53	5.19 ± 2.42	4.18 ± 1.84	1	7.55e-5
15	91	39	6.15 ± 1.86	4.70 ± 2.08	1	1.33e-5
16	97	52	4.85 ± 2.36	4.36 ± 2.10	0	0.10
17	172	38	5.37 ± 2.43	4.55 ± 2.01	1	0.01
18	50	38	5.89 ± 2.12	4.71 ± 2.07	1	5.54e-4
19	119	38	4.39 ± 2.33	4.04 ± 2.24	0	0.34
20	69	48	5.65 ± 1.80	5.04 ± 1.75	1	0.02
21	29	44	5.23 ± 1.96	4.59 ± 2.06	0	0.05
22	72	36	4.19 ± 1.53	3.98 ± 1.71	0	0.46

Fig. S3B. Summary of Statistically Significant Gene Pairs: From the set of "Dynamic Intra-Correlated Genes", gene pairs that were found to be inter-correlated in form and function were extracted from each chromosome. These inter-correlated gene pairs were found to share more motif sequences than randomly expected in 16 out 22 chromosomes. This give evidence to the idea that these gene pairs are inter-correlated due to being involved with common transcription factories. Average number of shared motifs was calculated by averaging the number of shared motifs between all possible gene pairs (taken from the "Dynamic Intra-Correlated Genes" set) on each chromosome.

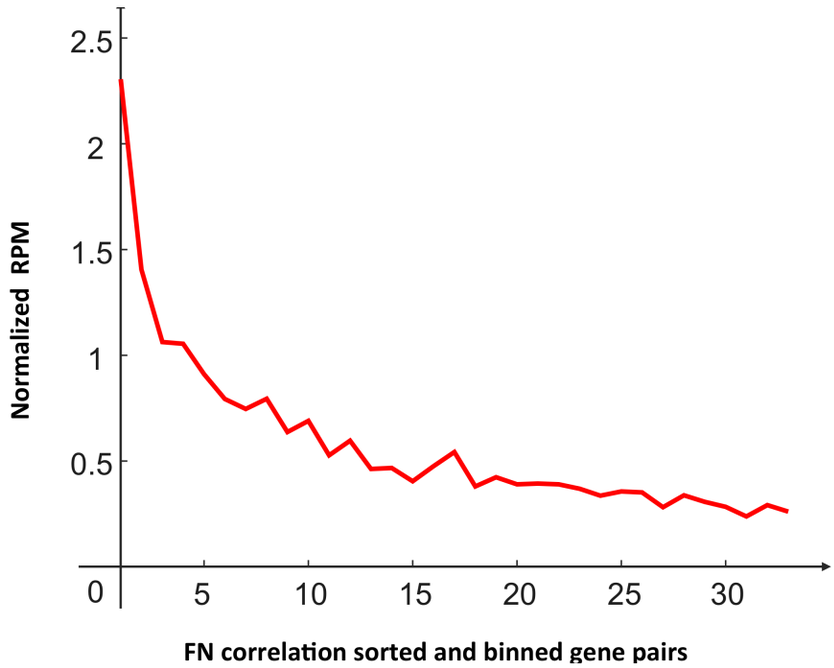


Fig. S3C. . This graph shows that FN correlation corresponds to more Hi-C contact between genes. The 2574 Dynamic intra-correlated genes were analyzed by computing the FN correlations between all possible pairs, resulting in 3,311,451 pair-wise interactions (2574 choose 2). The corresponding gene pair Hi-C contact over time was also extracted, and averaged over the 8 time points. The gene pairs are ordered in terms of FN correlation value (absolute value) and binned in groups of 10,000 gene pairs along the x axis. The y-axis is the corresponding average Hi-C interaction of all the gene pairs in each bin (Normalized RPM). A clear trend emerges, showing that if two genes have high FN correlations over time, these genes are more likely to interact. Normalized RPM is calculated by summation of counts between pairs, divided by total counts of experimentation, multiplied by a hundred million.

Biological module	Total genes	Node dynamic genes	Edge dynamic pairs	av. # of shared motifs	av. # of shared motifs (random)	Statistically significant?	P-value
24 Circadian	1162	211	52	7.38 ± 1.84	4.75 ± 2.16	1	1.68e-18
Cell Cycle	634	136	49	5.06 ± 2.40	4.30 ± 1.96	1	0.01
WH Stanford	939	272	33	6.21 ± 2.38	4.33 ± 2.09	1	2.33e-07
DEX	695	263	46	5.94 ± 2.27	4.96 ± 2.01	1	6.58e-04

Fig. S3D. Summary of Statistically Significant Gene Pairs: From the set of dynamic gene pairs that were found to be inter-correlated in form and function were extracted from each biological module. These inter-correlated gene pairs were found to share more motif sequences than randomly expected for all biological modules. This give evidence to the idea that these gene pairs are inter-correlated due to being involved with common transcription factories. Average number of shared motifs was calculated by averaging the number of shared motifs between all possible gene pairs of a biological module.

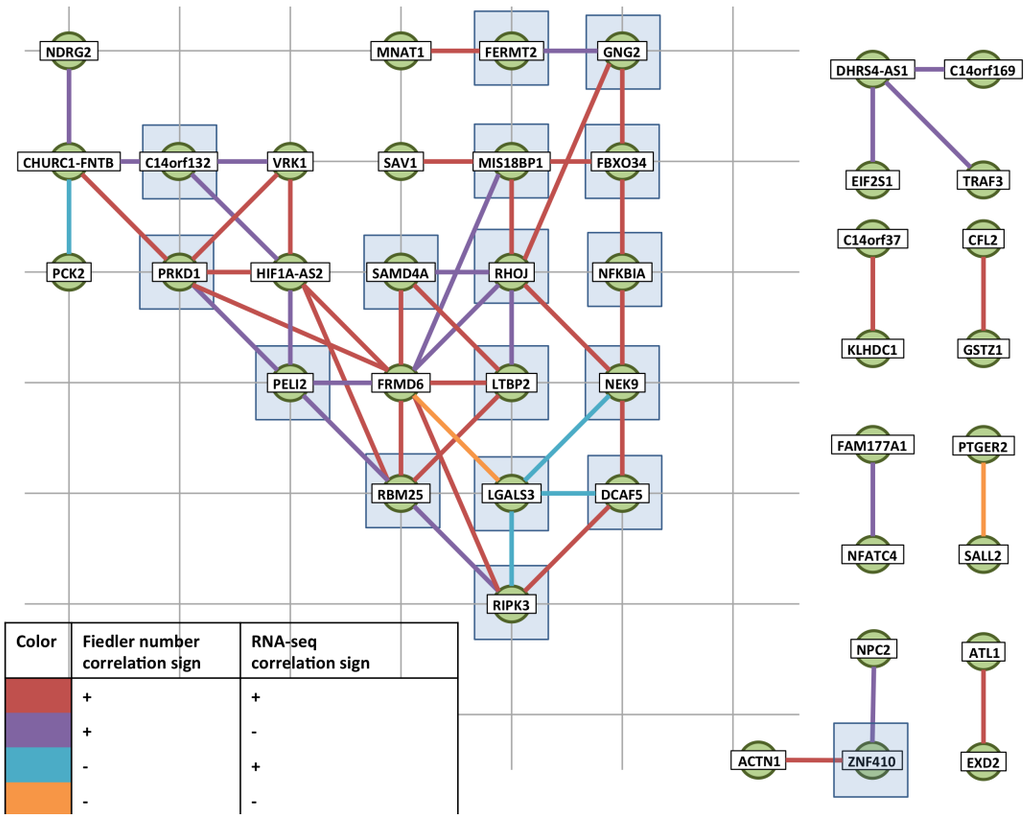


Fig. S3E. Networks of Dynamic intra- and inter-correlated form-function gene pairs on chromosome 14. Green nodes represent genes, thick edges between pairs of nodes represent a correlation. The colors of edges show how the two genes are correlated. Genes that have transcription factors in common with all edges are denoted by shaded blue squares. Transcription factors associated with gene pairs are shown in Table S10.

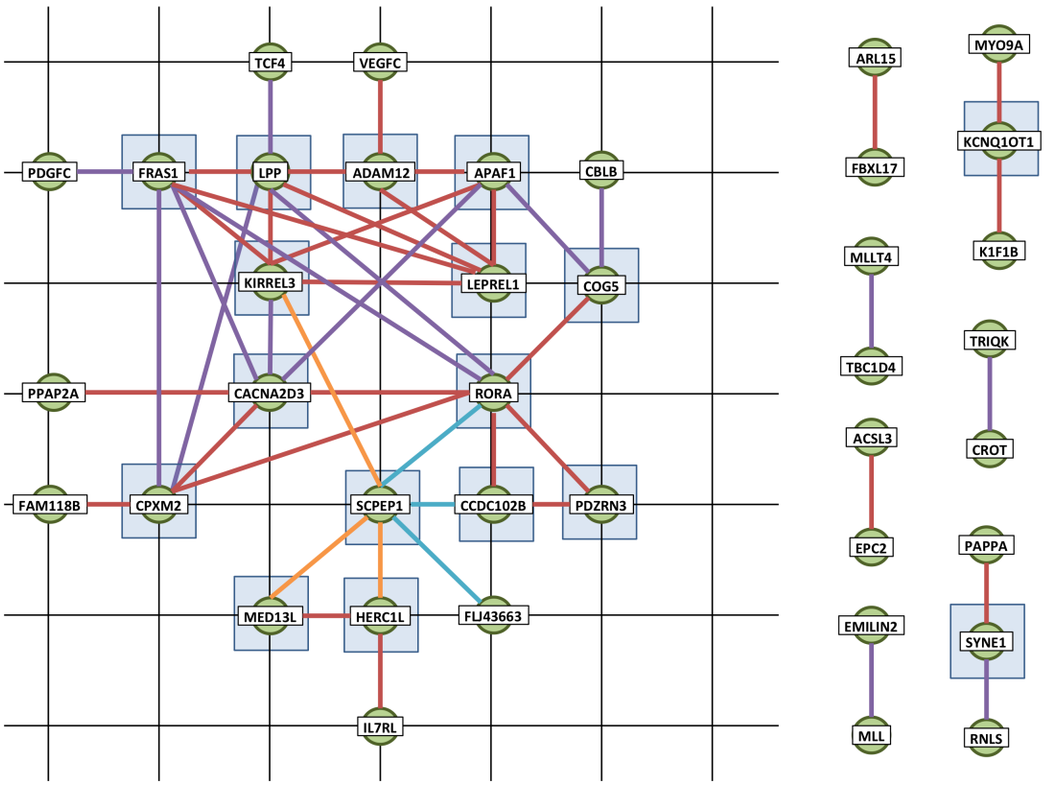


Fig. S3F. Networks of Dynamic intra- and inter-correlated form-function gene pairs of cell cycle gene set. Green nodes represent genes, thick edges between pairs of nodes represent a correlation. The colors of edges show how the two genes are correlated. Genes that have transcription factors in common with all edges are denoted by shaded blue squares. Transcription factors associated with gene pairs are shown in Table S10.

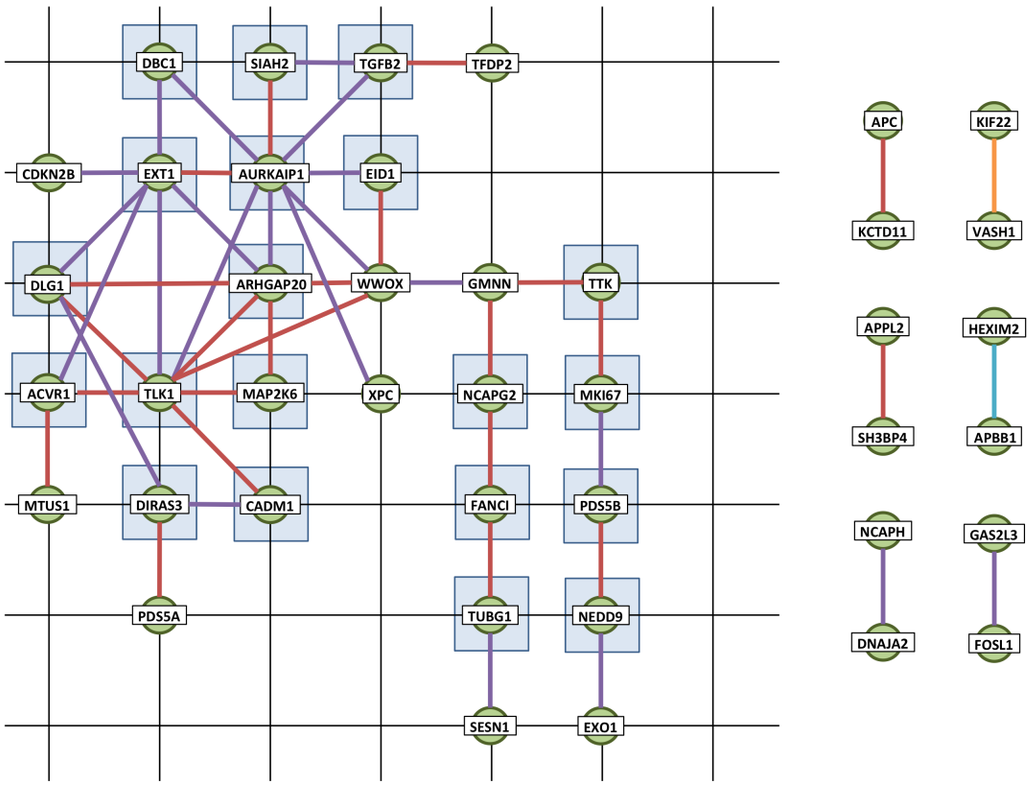


Fig. S3G. Networks of Dynamic intra- and inter-correlated form-function gene pairs of circadian 24h gene set. Green nodes represent genes, thick edges between pairs of nodes represent a correlation. The colors of edges show how the two genes are correlated. Genes that have transcription factors in common with all edges are denoted by shaded blue squares. Transcription factors associated with gene pairs are shown in Table S10.

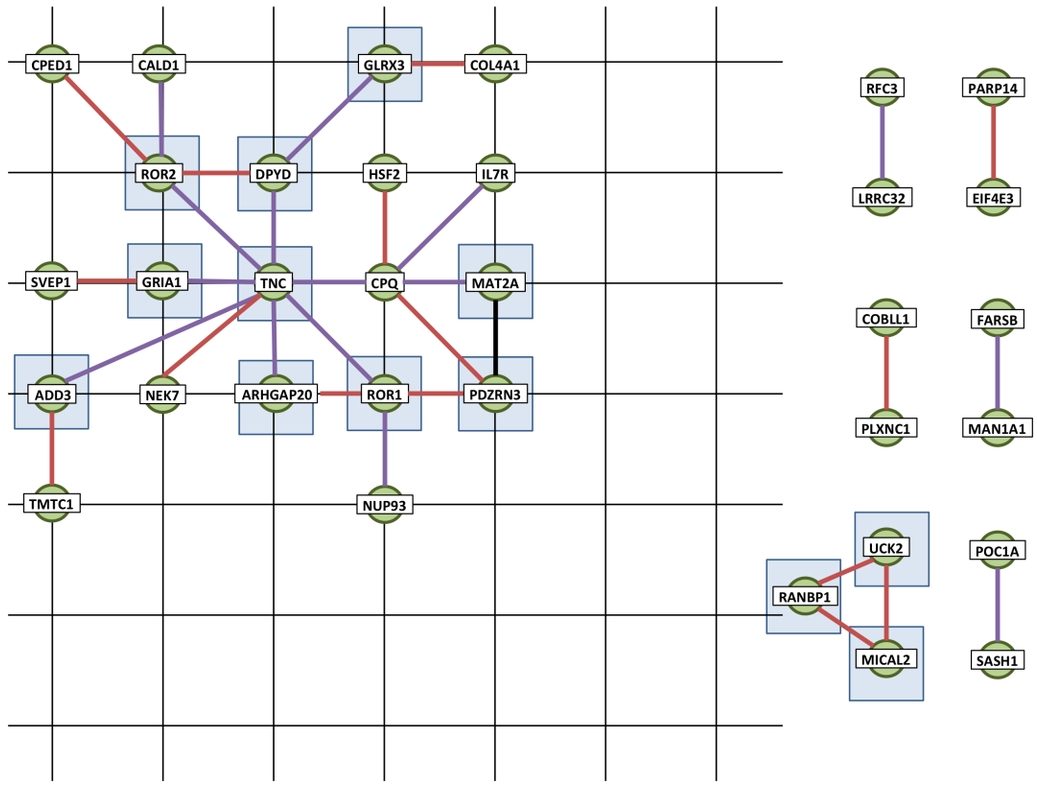


Fig. S3H. Networks of Dynamic intra- and inter-correlated form-function gene pairs of would healing gene set. Green nodes represent genes, thick edges between pairs of nodes represent a correlation. The colors of edges show how the two genes are correlated. Genes that have transcription factors in common with all edges are denoted by shaded blue squares. Transcription factors associated with gene pairs are shown in Table S10.

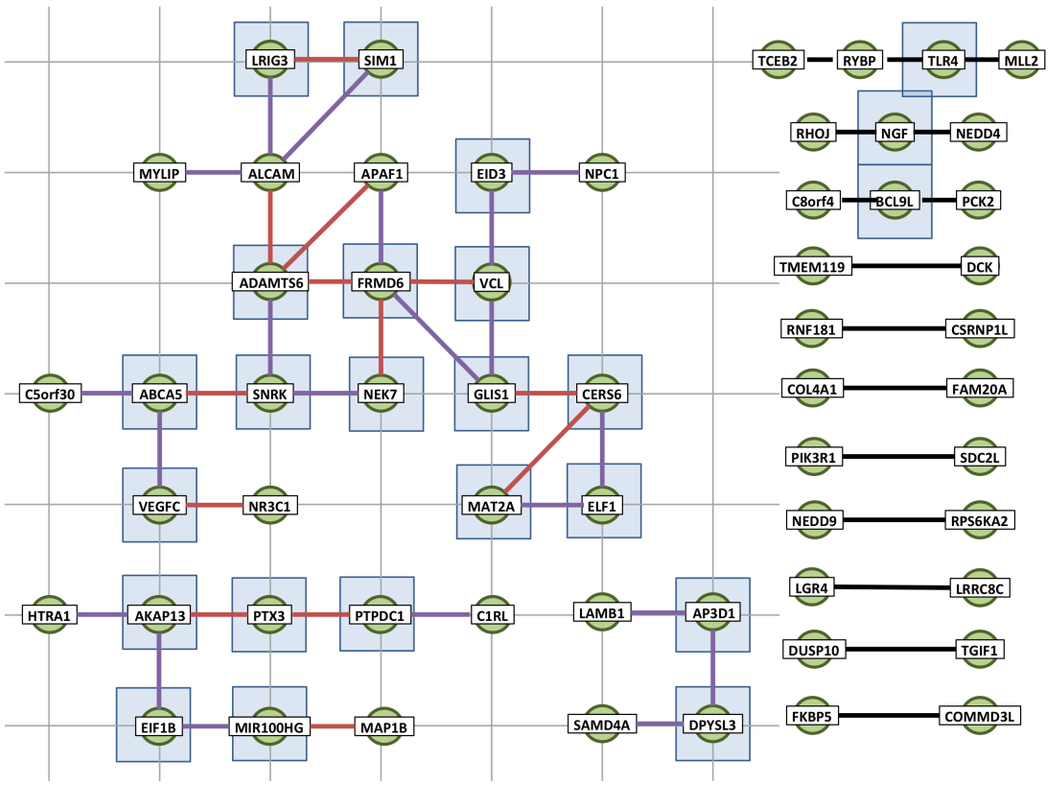
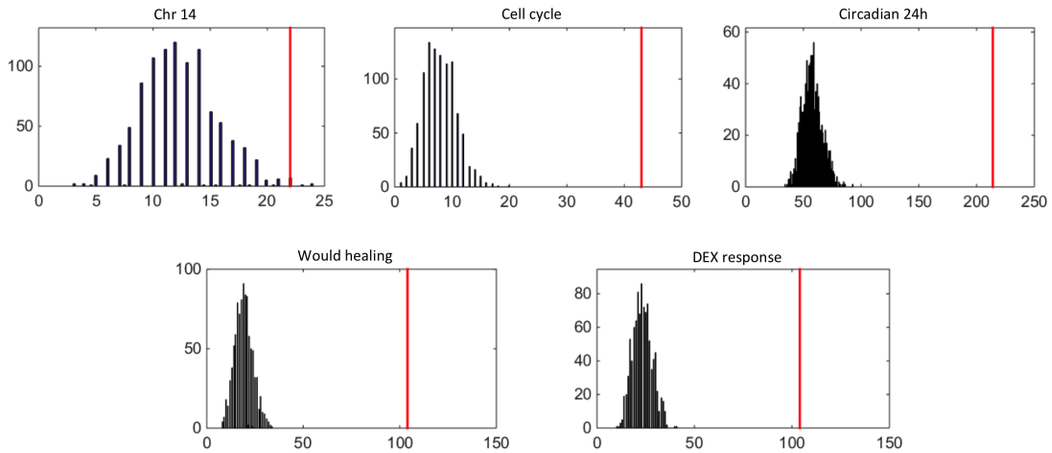


Fig. S3I. Networks of Dynamic intra- and inter-correlated form-function gene pairs of DEX response gene set. Green nodes represent genes, thick edges between pairs of nodes represent a correlation. The colors of edges show how the two genes are correlated. Genes that have transcription factors in common with all edges are denoted by shaded blue squares. Transcription factors associated with gene pairs are shown in Table S10.



Gene set	Chr 14	Cell cycle	Circadian 24h	Wound healing	DEX response
Number of pairs	22	43	214	104	104
p value	0.003	<0.001	<0.001	<0.001	<0.001
Significance (0.01)	YES	YES	YES	YES	YES

Fig. S3J. Significance of identified pair numbers via permutation test. The empirical null was generated by randomly shuffling 1000 times the contacts of gene pairs.

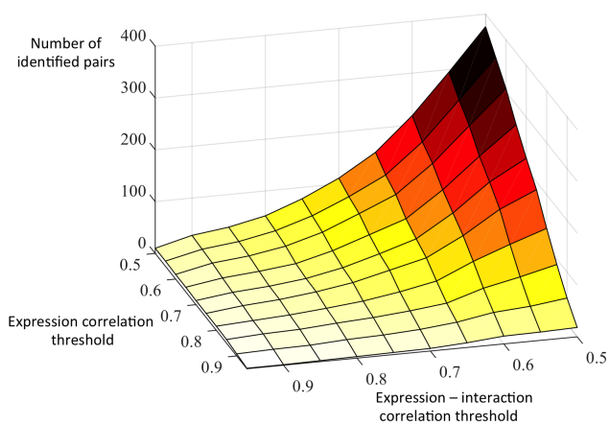


Fig. S3K. The number of identified gene pairs for chromosome 14 as a function of expression correlation threshold and expression-interaction correlation threshold.

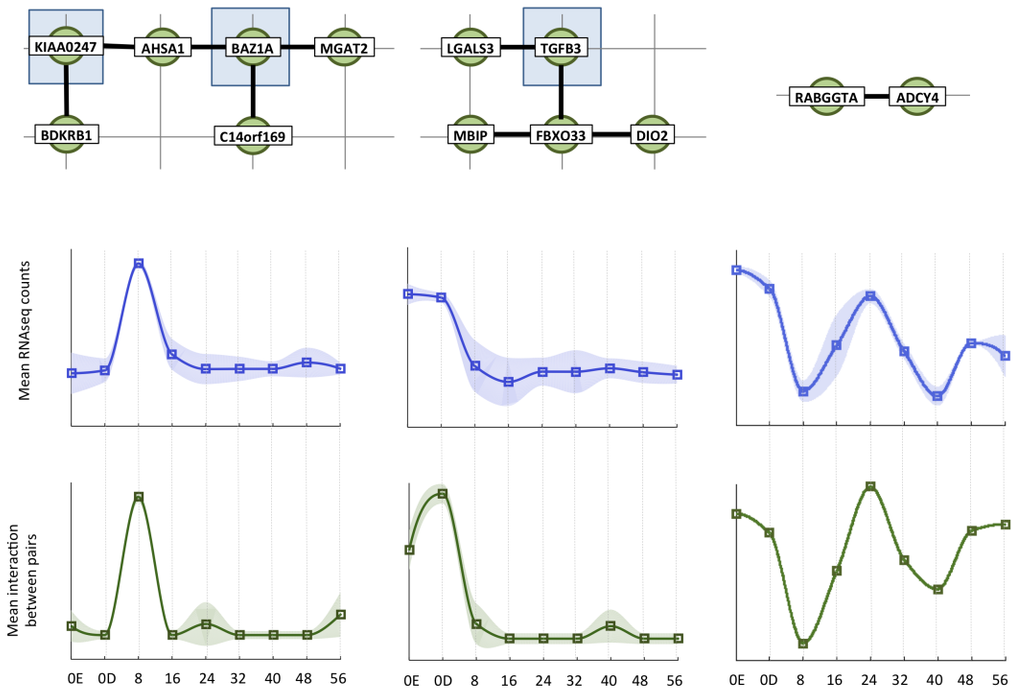


Fig. S3L. . Top panel shows three co-regulation networks of chromosome 14. Two genes of each connected pair have expression correlation larger than 0.8 and expression-inter-contact correlation larger than 0.8. Middle panel: line plots show the average expression values of the genes corresponding to each network lined in columns over time, shaded area indicates expression variance. Bottom panel: line plots show the average inter-contact of gene pairs in corresponding networks lined in columns over time. These figures illustrate how the expression responds to the spatial interaction of genes, implying co-transcription or the existence of transcription factories. Gene information and shared transcription factors are reported in Table S11.

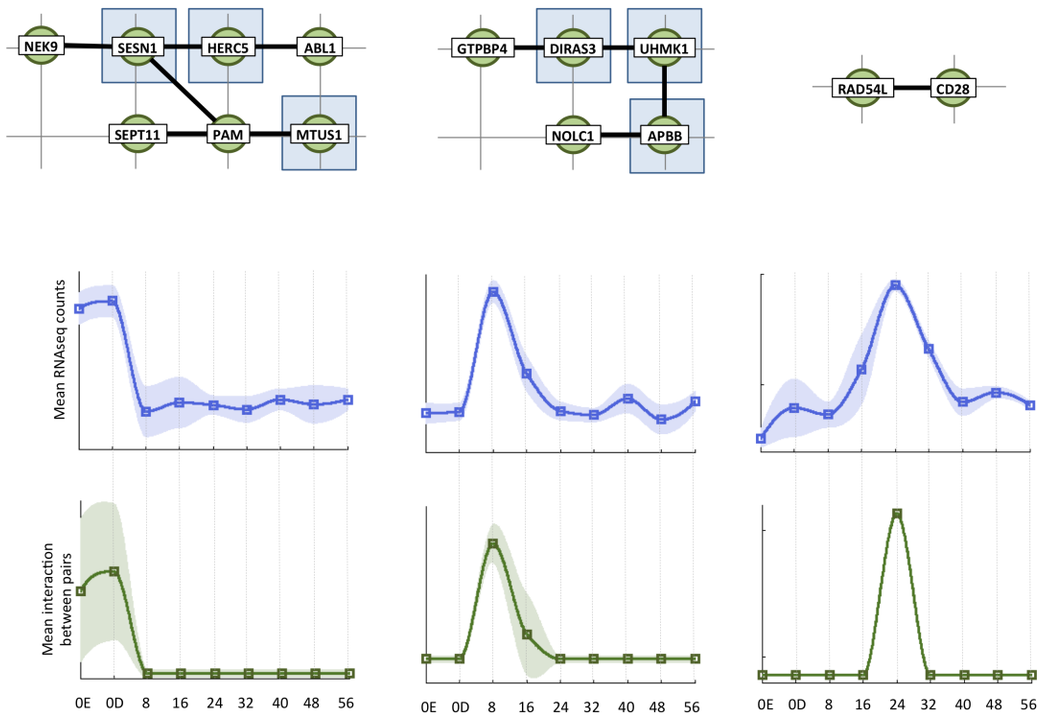


Fig. S3M. Top panel shows three co-regulation networks of the cell cycle set. Two genes of each connected pair have expression correlation larger than 0.8 and expression-inter-contact correlation larger than 0.8. Middle panel: line plots show the average expression values of the genes corresponding to each network lined in columns over time, shaded area indicates expression variance. Bottom panel: line plots show the average inter-contact of gene pairs in corresponding networks lined in columns over time. These figures illustrate how the expression responds to the spatial interaction of genes, implying co-transcription or the existence of transcription factories. Gene information and shared transcription factors are reported in Table S11.

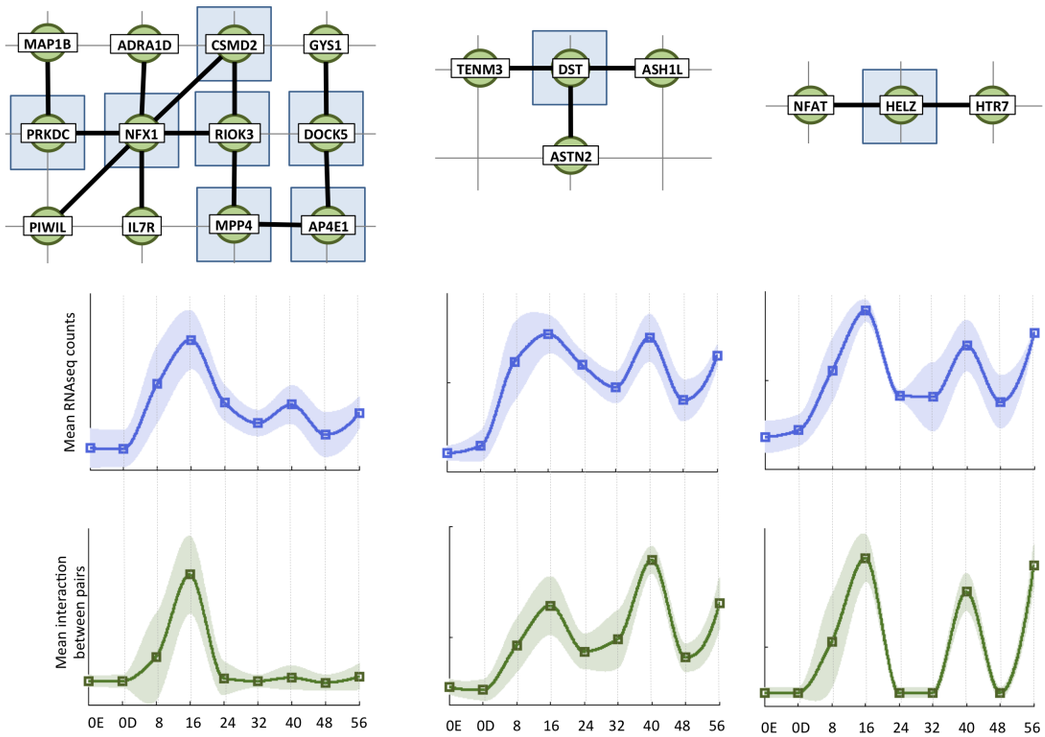


Fig. S3N. Top panel shows three co-regulation networks of the circadian 24h gene set. Two genes of each connected pair have expression correlation larger than 0.8 and expression-inter-contact correlation larger than 0.8. Middle panel: line plots show the average expression values of the genes corresponding to each network lined in columns over time, shaded area indicates expression variance. Bottom panel: line plots show the average inter-contact of gene pairs in corresponding networks lined in columns over time. These figures illustrate how the expression responds to the spatial interaction of genes, implying co-transcription or the existence of transcription factories. Gene information and shared transcription factors are reported in Table S11.

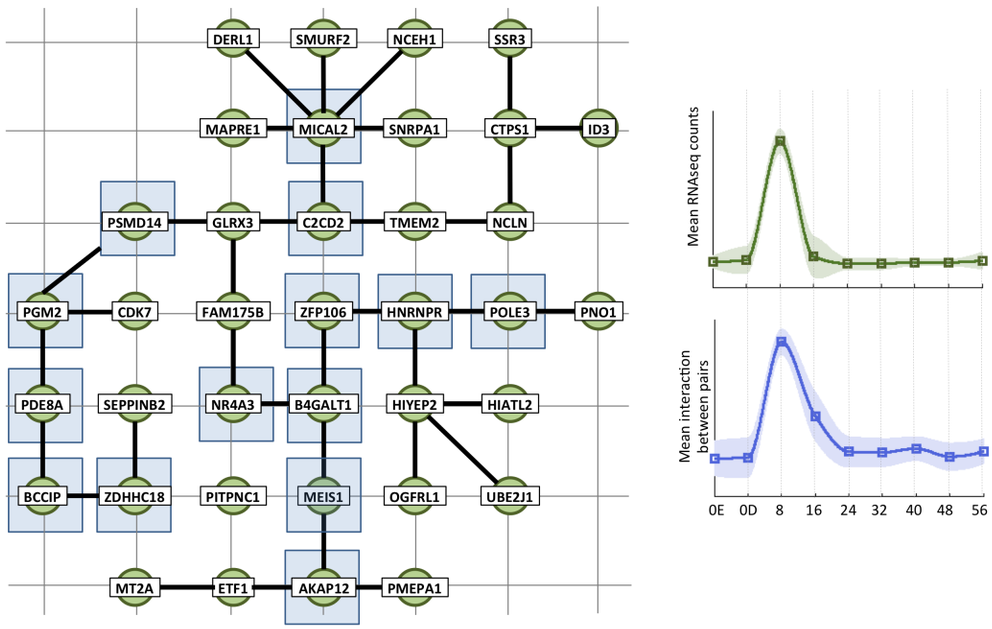


Fig. S3O. Left panel shows a co-regulation network of the wound healing gene set. Two genes of each connected pair have expression correlation larger than 0.8 and expression-inter-contact correlation larger than 0.8. Middle panel: line plots show the average expression values of the genes corresponding to each network lined in columns over time, shaded area indicates expression variance. Bottom panel: line plots show the average inter-contact of gene pairs in corresponding networks lined in columns over time. These figures illustrate how the expression responds to the spatial interaction of genes, implying co-transcription or the existence of transcription factories. Gene information and shared transcription factors are reported in Table S11.

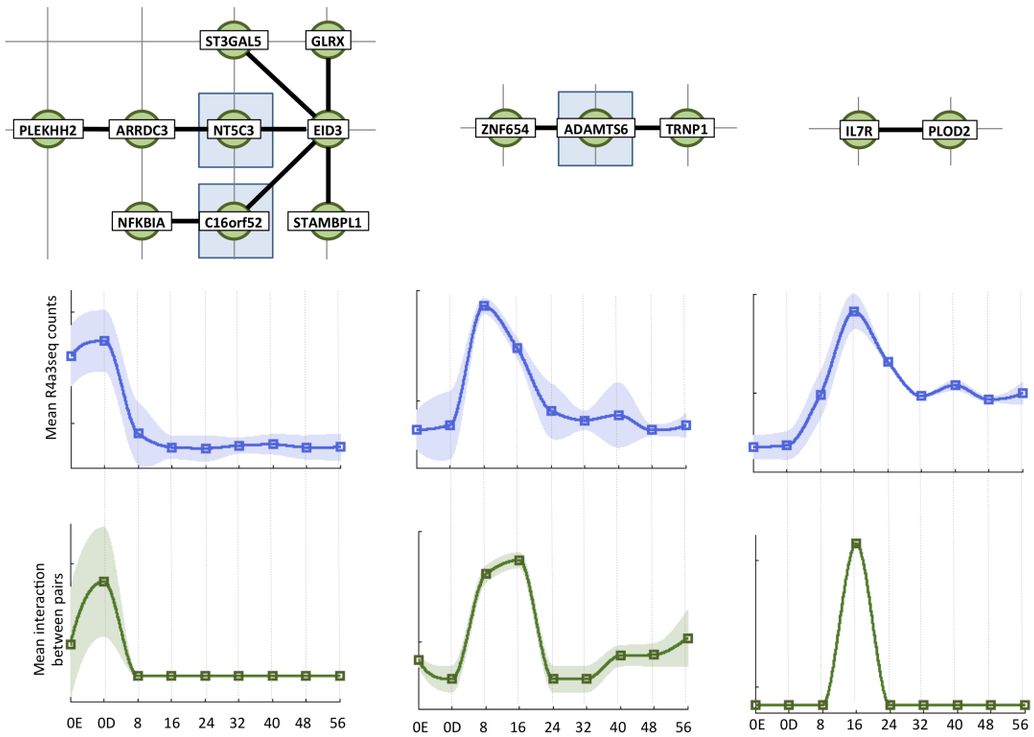


Fig. S3P. Top panel shows three co-regulation networks of the DEX response gene set. Two genes of each connected pair have expression correlation larger than 0.8 and expression-inter-contact correlation larger than 0.8. Middle panel: line plots show the average expression values of the genes corresponding to each network lined in columns over time, shaded area indicates expression variance. Bottom panel: line plots show the average inter-contact of gene pairs in corresponding networks lined in columns over time. These figures illustrate how the expression responds to the spatial interaction of genes, implying co-transcription or the existence of transcription factories. Gene information and shared transcription factors are reported in Table S11.

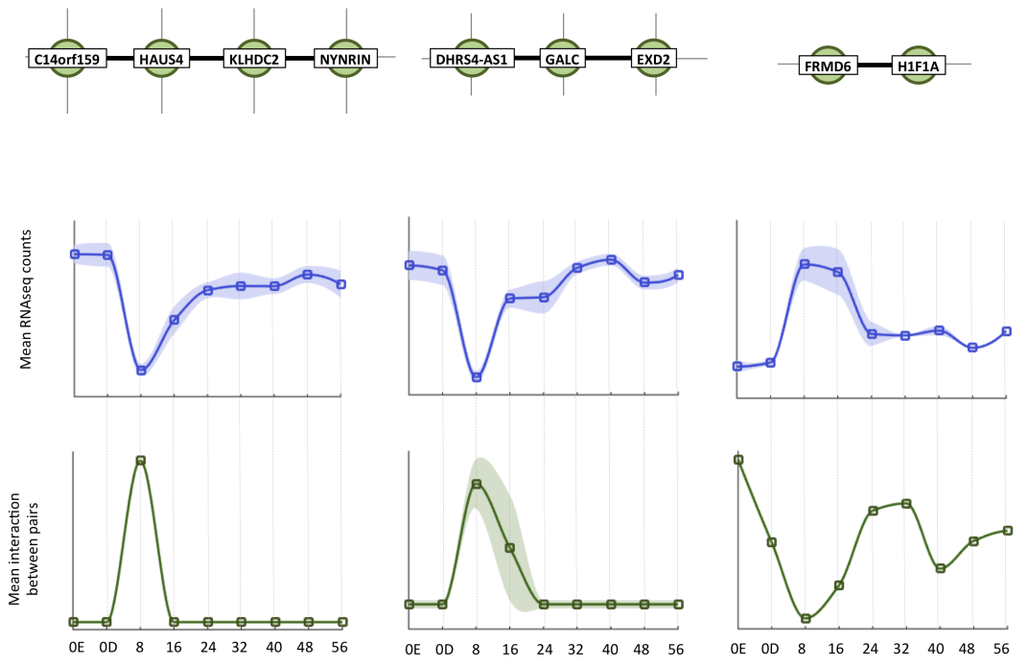


Fig. S3Q. Top panel shows three networks of chromosome 14. The pairs have anti-correlation with RNAseq and interactions. Two genes of each connected pair have expression correlation larger than 0.8 and expression inter-contact correlation larger than -0.8. Middle panel: line plots show the average expression values of the genes corresponding to each network lined in columns over time, shaded area indicates expression variance. Bottom panel: line plots show the average inter-contact of gene pairs in corresponding networks lined in columns over time. These figures illustrate how the expression responds to the spatial interaction of genes, implying co-transcription or the existence transcription factories.

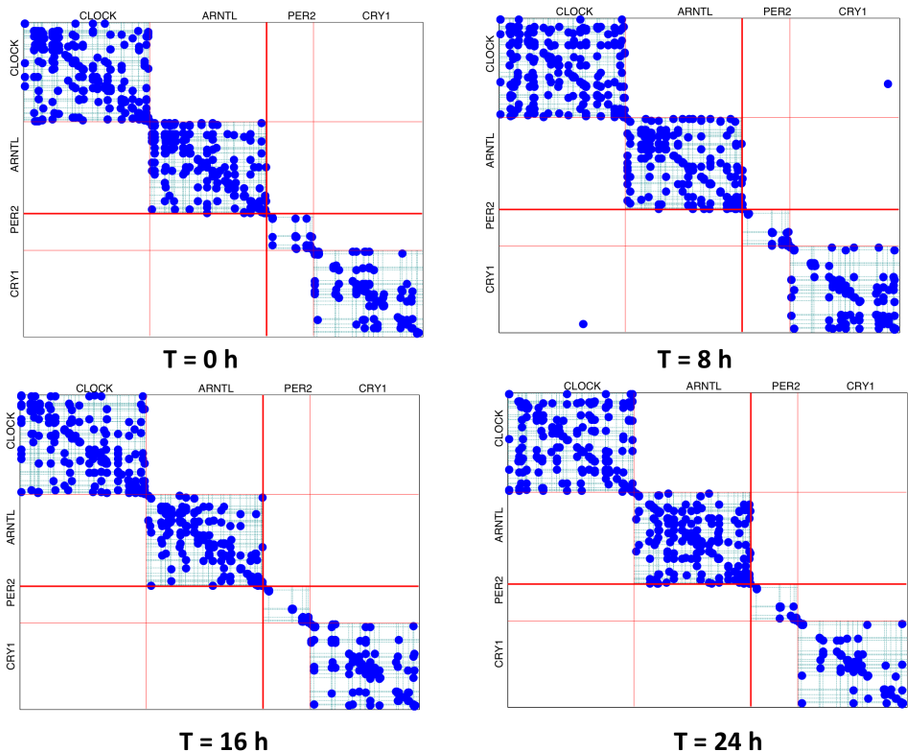


Fig. S4A. . Contact maps of four core circadian genes: CLOCK, ARNTL, PER2 and CRY1 over time T = 0h to 24h. It can be seen there are few inter-gene contacts.

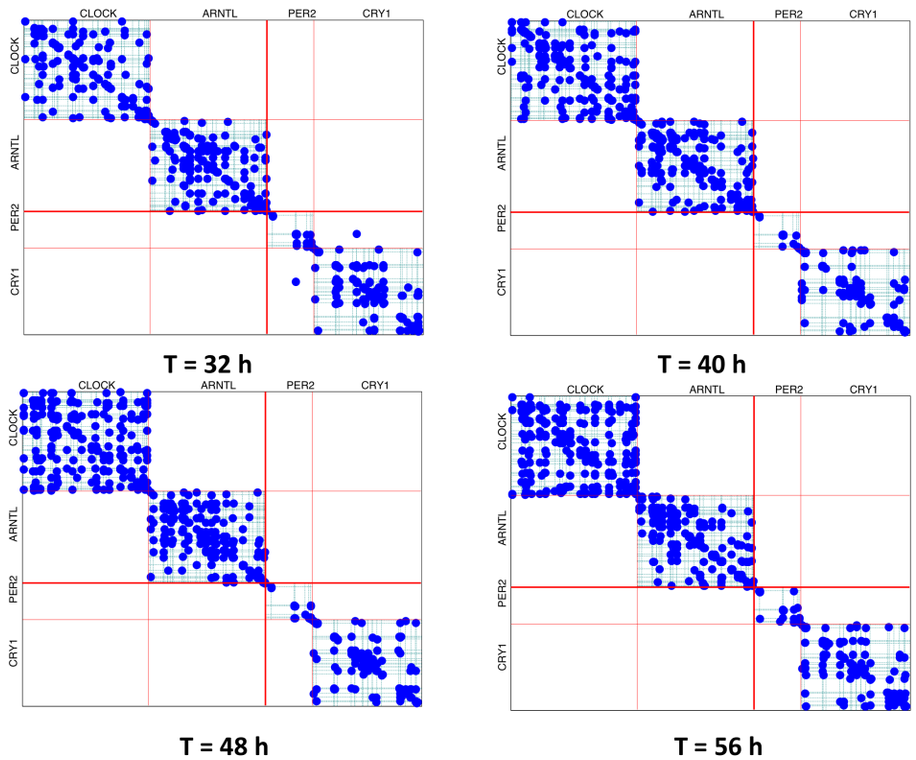


Fig. S4B. Contact maps of four core circadian genes: CLOCK, ARNTL, PER2 and CRY1 over time T = 32h to 56h. It can be seen there are few inter-gene contacts.

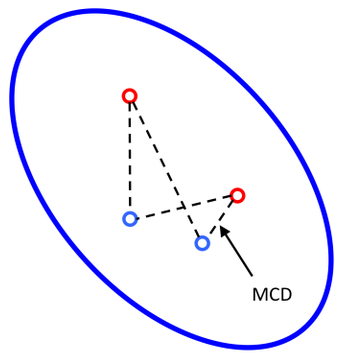


Fig. S4C. Illustration of the value taken for mean closest distance (MCD) calculation between two different genes.

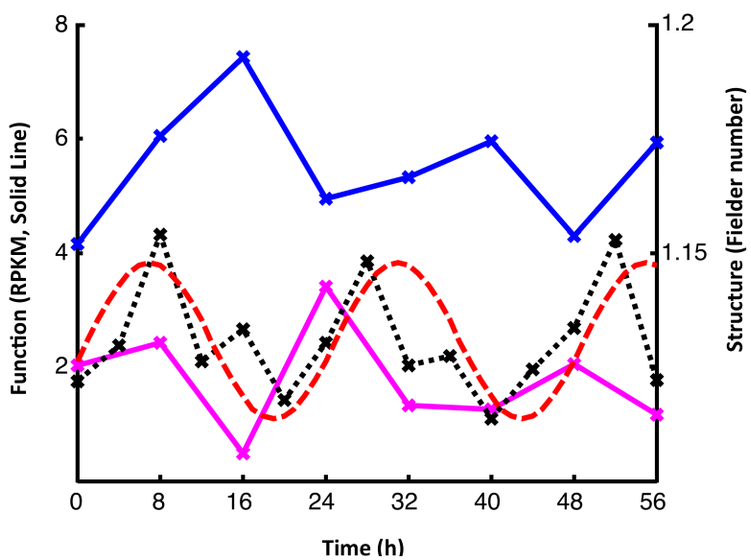


Fig. S4D. Fiedler Number between CLOCK and PER2 distance matrix, and RNA-seq expression of $CLOCK$ and $PER2$ over time. Dashed lines depict Fiedler number real data (black) and sine fit data (red) and correspond to the right y-axis scale (um). Solid lines depict RNA-seq expression of $CLOCK$ (blue) and $PER2$ (magenta) and correspond to the left y-axis scale (RPKM, reads/Kb/million reads).

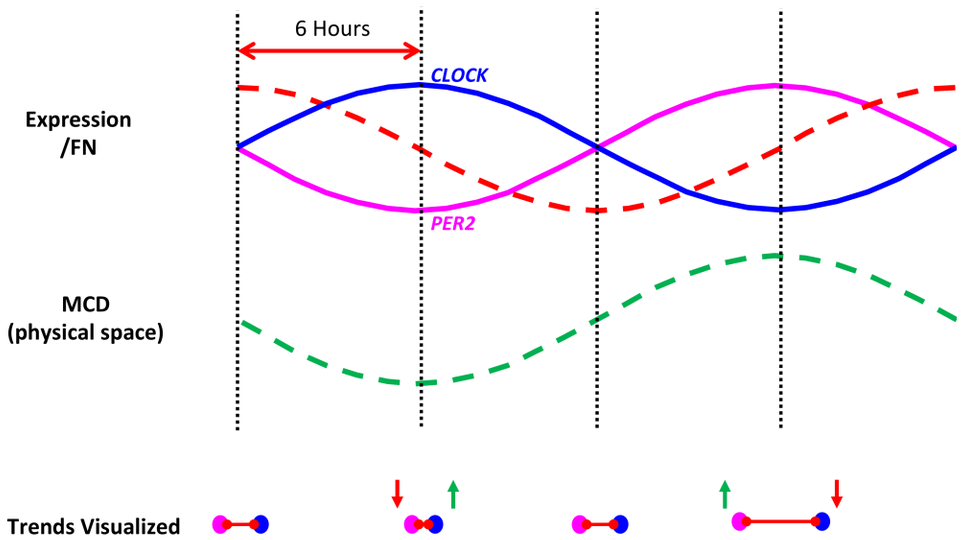


Fig. S4E. CLOCK-PER2 summary. RNA-seq, MCD, and Fiedler number trends are simplified to fit within 4 6-hours time points, representing 1 full circadian cycle. $PER2$ and $CLOCK$ RNA-seq expression trends (pink and blue solid lines, respectively, top). Dashed red line depicts the CLOCK-PER2 Fiedler Number (FN) trend. Dashed green line shows the CLOCK-PER2 MCD trend. Pink and blue circles depict physical space between PER2 and CLOCK, as simplified from MCD analysis.

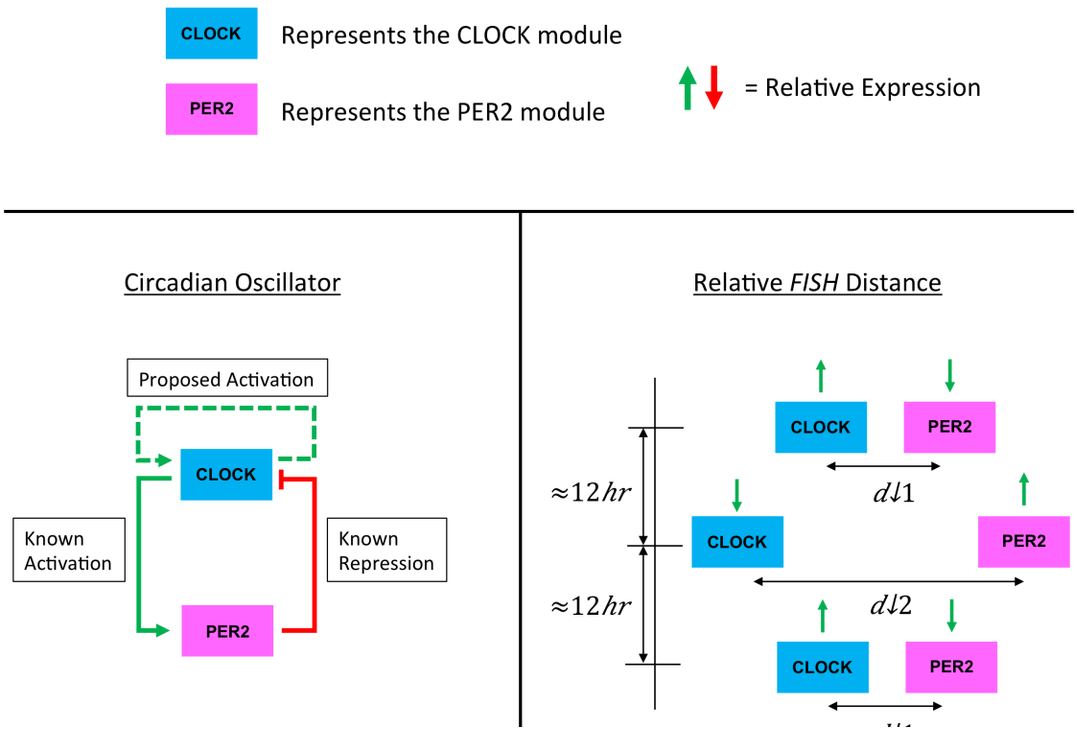


Fig. S4F. Left: A proposed feedback circuit for $CLOCK$ and $PER2$ expression, where $CLOCK$ has a mechanism in which it can reactivate itself (SOM). Right: Relative expression of $CLOCK$ and $PER2$ at given relative Euclidian distances.

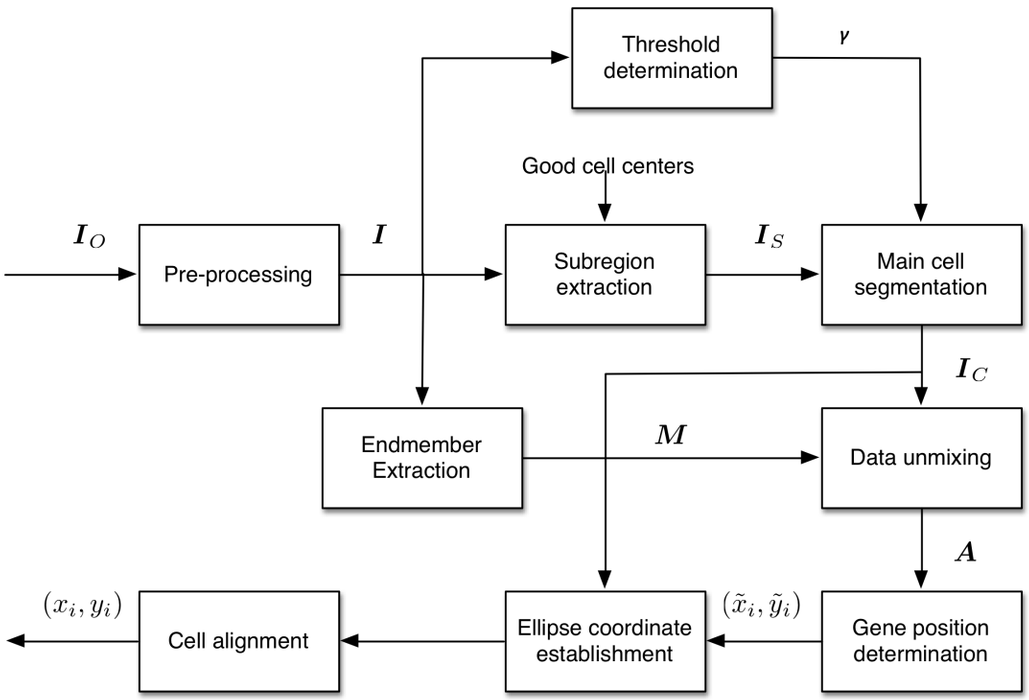


Fig. S4G. Procedure for analyzing the FISH images and extracting gene locations.

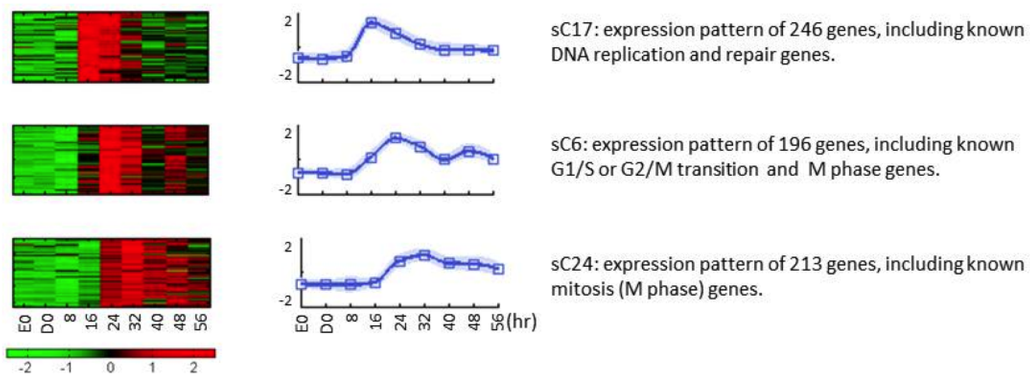


Fig. S5A. Expression patterns of cell cycle phase genes. Normalized gene expression values (over time mean subtracted & variance normalized) are used for spectral clustering. The heat maps of 3 sub-clusters (sC17, sC6, and sC24) are illustrated in the 1st column, and the average of normalized gene expression values (patterns) of each cluster are shown by the curves to the right of the heat maps, with the shaded area denoting the standard deviations, and boxes indicating sampling time points. The time points of sampling are horizontally shown at the bottom on each panel in hours (hr). "D0" is the time zero sample treated with Dexamethasone for 1 hour without exposure to serum (contrast to E); "E0" indicates base line time zero control without exposure to Dexamethasone and serum, and numbers indicate sampling time points in 8-hour intervals after serum stimulation from time zero. Color: green represents expression levels below the mean, red above the mean, dark: mean level.

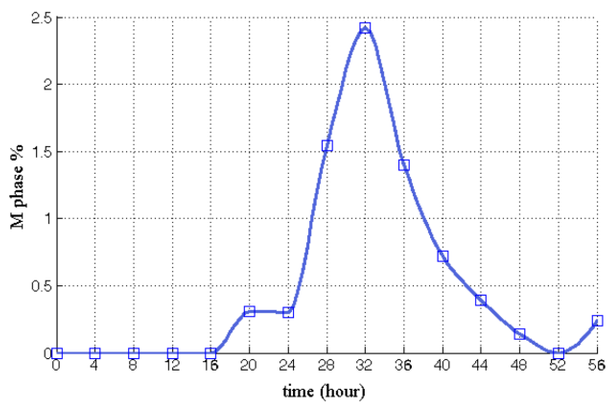


Fig. S5B. M phase cell counts. The number of M phase cells (%) counted in slide cultures are plotted against time in hours (4-hour interval). No M phase cells are seen before hour 20 after serum stimulation.

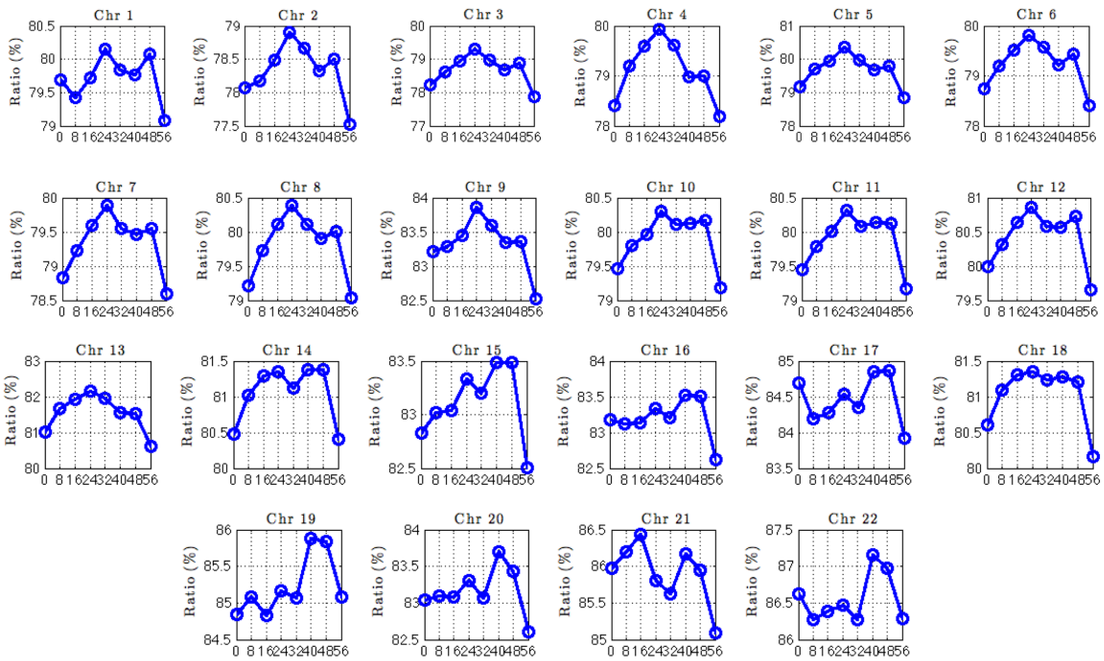


Fig. S5C. Dynamics of contact structures over time in different scales. The ratio between the sum of the diagonal band counts and that of the entire collection of intra-chromosome Hi-C matrices for chromosomes 1-22. Peaks can mainly be found at time points 24h and 48h. This possibly implies the contribution from diagonal dominant matrices introduced by the M-phase of cell cycle