Comparing Pattern Detection Methods in Microarray Time Series of the Segmentation Clock

Mary-Lee Dequéant[1], Sebastian Ahnert[2], Herbert Edelsbrunner[3,4], Thomas M. A. Fink[5,6], Earl F. Glynn[1], Gaye Hattem[1], Andrzej Kudlicki[7], Yuriy Mileyko[3], Jason Morton[8], Arcady R. Mushegian[1], Lior Pachter[8], Maga Rowicka[7], Anne Shiu[8], Bernd Sturmfels[8], Olivier Pourquié[1,9*]

1 Stowers Institute for Medical Research, Kansas City, Missouri, United States of America, **2** Theory of Condensed Matter, Cavendish Laboratory, Cambridge, United Kingdom, **3** Departments of Computer Science and of Mathematics, Duke University, Durham, North Carolina, United States of America, **4** Geomagic, Research Triangle Park, North Carolina, United States of America, **5** Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche (UMR), Paris, France, **6** Systems Biology, Institut Curie, Paris , France, **7** Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **8** Department of Mathematics, University of California Berkeley, Berkeley, California, United States of America, **9** Howard Hughes Medical Institute, Kansas City, Missouri, United States of America

*Corresponding author
Olivier Pourquié, Ph.D.
Howard Hughes Medical Institute and
Stowers Institute for Medical Research
1000 E. 50th Street
Kansas City, Missouri 64110
Phone: 816-926-4442
Fax: 816-926-2095
Email: olp@stowers-institute.org

Running heading: Detecting Patterns in Microarray Time Series

Abbreviations: A, Address reduction; C, Cyclohedron test; EST, expressed sequence tags; FOM, Figure of Merit; L, Lomb-Scargle analysis; MEV, MultiExperiment Viewer; P, Phase consistency; PSM, presomitic mesoderm; S, Stable persistence

Abstract

Background

While genome-wide gene expression profiling data are generated at an increasing rate, the repertoire of approaches for pattern discovery in these data is still limited. Finding subtle patterns of interest in large amounts of data (tens of thousands of profiles) associated with a certain level of noise is still a challenge. A microarray time series was recently generated to study the transcriptional program of the mouse segmentation clock, a biological oscillator associated with the periodic formation of the segments of the body axis. A method related to Fourier analysis, the Lomb-Scargle periodogram, was used to detect periodic profiles in the dataset, leading to the identification of a set of novel cyclic genes associated with the segmentation clock.

Methodology/principal findings

In this work, we apply to the same microarray time series dataset, four distinct mathematical methods: Phase consistency, Address reduction, Cyclohedron test and Stable persistence–each method using a different conceptual framework to identify significant patterns. A common feature of some of these methods is that, unlike Fourier transforms, they are not strongly dependent on the assumption of periodicity of the pattern of interest. Remarkably, these methods identified blindly the expression profiles of known cyclic genes as the most significant patterns in the dataset. Many candidate genes predicted by more than one approach appeared to be true positive cyclic genes (including known cyclic genes and a novel candidate cyclic gene experimentally validated in the mouse embryo) and will be of particular interest for future research.

Significance, and conclusion

Thus, our results validate the use of these novel pattern detection strategies, notably for the detection of periodic profiles. It further suggests that combining several distinct mathematical approaches to analyze microarray datasets comprise a valuable strategy for the identification of genes exhibiting novel, interesting transcriptional patterns.

Introduction

The dynamics of gene expression in a biological system exposed to varying experimental conditions, such as dose response to a drug or a time course, can be analyzed now at the whole genome level by generating series of microarrays. The main difficulty is then to identify, among the thousands of gene expression profiles recorded, hidden trends or patterns in specific gene expression profiles revealing particular properties of the system that may lead to the formation of novel biological hypothesis. However, because some of these patterns tend to be subtle and of partially or completely unknown shape, as well as noisy, determining appropriate methods of pattern detection still remains challenging.

Microarray time series have been extensively generated to study periodic biological processes, such as the cell cycle [1], circadian regulation [2] [3], plasmodium cycle [4] and vertebrae segmentation [5]. Several approaches have been used to identify genes whose periodic expression underlies the cellular or tissue-level periodic behavior of the system. However, a common feature of these approaches is their strict assumptions about the shape of periodic profiles. For example, popular Fourier-based methods detect

periodicity by decomposing gene expression profiles into a series of sine curves. However, these methods would be less sensitive to many types of periodic profiles that are poorly approximated by sine curves (because of the noise in the experimental measurements or because periodic profiles might have a different shape, such as asymmetric profiles with short peak and long trough), introducing biases to the results. Here, we applied to the same dataset four methods that do not have such prior assumptions about the shape of the profiles of interest.

The segmentation of the vertebrate axis into periodic structures, such as vertebrae, occurs during embryogenesis when the vertebral precursors, the somites, are formed rhythmically from the presomitic mesoderm (PSM). This process is associated with a molecular oscillator that drives periodic gene expression in the PSM with a period corresponding to that of somite formation [6]. During one oscillation cycle, cyclic genes, such as *Lunatic fringe (Lfng)*, are expressed as a wave initiated in the posterior PSM that progressively migrates along the PSM and narrows as it moves anteriorly [7-9]. A microarray time series of PSM samples encompassing one period of the segmentation clock has been generated in the mouse and analyzed using the Lomb-Scargle (L) algorithm, a method related to Fourier analysis [5,10]. This analysis identified a large number of novel cyclic genes that fall into two biologically coherent clusters associated with the Wnt and the Notch/FGF signaling pathways.

In this paper, we applied four different mathematical approaches to the same mouse segmentation data and compared the results to the original study. The four methods are: Phase consistency (P); Address reduction (A); Cyclohedron test (C); and Stable persistence (S). These methods can be divided into two groups. In the first group, the P and S methods search for periodic profiles but in a very different way compared to the L method. The P method optimizes the ratio of the total variation to the sum of the piecewise variations, with the pieces set by the behavior of the known cyclic gene *Lfng,* and the S method is based on a numerical assessment that is provably stable (See Material and Methods). In the second group, the A and C methods attempt to identify significant patterns without assuming the periodic nature of the patterns of interest. Briefly, A and C associate significance inversely with the likelihood of certain groups of patterns–the two methods differing from each other mainly in how they partition the set of all possible patterns into groups.

All methods performed similarly well in identifying a list of benchmark cyclic genes among their top ranked candidates, validating their use for the detection of periodic profiles. Since some of the methods do not search specifically for periodic profiles, it further suggests that the cyclic gene profiles are the most significant patterns in the dataset. In addition to the genes in the benchmark list, each method identified a number of novel candidate cyclic genes of the Wnt pathway. We show that the Wnt-target and Wnt-modulator cysteine rich protein 61 (*Cyr61*), identified by three of the methods, represents a novel bona fide cyclic gene of the mouse segmentation clock. Thus, our data suggest that combining several distinct mathematical approaches to analyze microarray datasets is a valuable strategy for the identification of genes exhibiting interesting transcriptional patterns.

Results

In this study, we used the microarray dataset generated as described [5]. Briefly, in order to identify cyclic genes associated with the mouse segmentation clock, a time series was generated by collecting the PSM tissue that undergoes the clock oscillations from 17 embryos (17 time points). This microarray dataset was analyzed [5] using L analysis that focuses on the genes whose expression patterns display the best fit to a sine curve [10]. A set of 27 cyclic probe sets out of the top 40 probe sets, identified by a period between 80 to 130 minutes and ranked by their statistical significance, was subsequently validated by *in situ* hybridization [5]. This 27 probe set list (Table S1) contains seven cyclic genes whose cyclic expression has been discovered independently from the microarray study by *in situ* hybridization (Table S1) [7-9,11-16].

We used the four methods (P, A, C and S) to rank the 7,549 probe sets of the dataset in order of the significance of their expression profile (as defined by each method-See Materials and Methods) and we restricted the analysis to the top 300 ranked probe sets from each list (Table S2-S6). First, we compared the rank of the benchmark cyclic genes that corresponded to seven probe sets and independently identified from non-microarray experimental methods. The L, P, A and S methods identified at least five out of the seven benchmark genes in the top ranked 100 probe sets, whereas C identified three of the seven known genes (Figure 1A). However, if we concentrate on the top 10 probe sets in each list (a more practical number for potential detailed experimental investigation), all of the methods perform similarly by detecting from one to two of these known genes. To measure the performance to a higher resolution, we repeated this analysis using a larger validation set of 27 probe sets. In addition to the seven probe sets above, this set is composed of cyclic genes identified by the L analysis and then experimentally validated [5]. Since this benchmark set is based on the L method, this method was removed from the analysis comparison (Figure 1B). As indicated in Figure 1B, method S performs best by identifying 90% of the benchmark genes, followed by methods A and P (approximately 75% and 63%, respectively) and C (approximately 37%). Again, when we restrict our attention to the top 10 probe sets of each list, the methods perform similarly by ranking among them from four to five benchmark genes. In conclusion, like the L method, the P, A, C and S methods identify known cyclic genes from among their top ranked probe sets.

We next compared the intersection of the top 300 ranked probe sets from the four new methods and method L. This is represented in Figure 2A as a 5-set Venn diagram in which each color corresponds to a different method. A different representation of the same diagram is shown in Figure 2B in the form of the lattice of the subsets of a five element set (Hasse diagram). The total number of distinct probe set profiles in all of the five sets (the union) is 884; the total number in each of the five sets (the intersection) is 21. Remarkably, the common overlap contains eight true positive cyclic genes (Table 1). Note that the L, P and C methods identified a large number of unique genes (104, 160 and 154, respectively) compared to method A (67) and method S (47).

We then studied the possible connections among the top 300 probe sets, identified by each method, to the segmentation clock process. Most of the validated cyclic genes (77%) in the top 40 probe sets of the L list are associated with the Notch, FGF and Wnt signaling pathways [5]. These genes essentially belong to two clusters: the first contains genes associated with the Notch- and FGF-signaling pathways; the second contains genes associated with the Wnt pathway and oscillating in opposite phase with the genes of the

first cluster. Bootstrap analysis provides particularly strong statistical support for the Wnt cluster (data not shown). In accordance with the principle of "guilt by association," the so called L-Wnt cluster is strikingly homogenous because 90% of the genes contained in this cluster are indeed linked to Wnt signaling.

Therefore, we tested the predictions of each of the four novel methods by investigating the link of their candidates to Wnt signaling. We independently clustered the top 300 gene expression profiles identified by the P, A, C, S and L methods (see Material and Methods). In each of the five analyses, we identified a cluster of periodic profiles containing at least the known cyclic genes (*Axin2, dickkopf homolog 1* [*Dkk1*]*, c-Myc* and *dapper homolog 1* [*Dact1*]) of the Wnt pathway. We refer to the cluster that contains these four genes as the Wnt cluster of the method (Table 2). We find that the S method identifies the eight known members of the Wnt cluster previously identified by the L method and validated as described [5], while the P and A methods identify six out of eight, and the C method identifies four out of eight of the known Wnt cyclic genes. We further analyzed the novel candidate cyclic gene contained in each of the Wnt clusters through a literature search to investigate their potential link to Wnt signaling. This allowed us to identify six novel candidate genes in the L-Wnt cluster showing a link to the Wnt pathway (Table 2). Interestingly, the P method performs similarly to the L method by predicting six new candidates, while the A, C and S methods predict two, four and three additional members, respectively (Table 2). Interestingly, most of the Wnt cyclic gene candidates predicted by the L method are also predicted by at least one of the other methods (Table 2), lending added credibility to these candidates. In contrast, the P and C methods strikingly predict the largest numbers of unique Wnt members (that are not predicted by any other method). Only one candidate Wnt cyclic gene, *Cyr61* (Figure 3A), was identified by three independent methods (including the L method). We investigated experimentally by *in situ* hybridization the expression pattern of this gene in the PSM of E9.0 mouse embryos (Figure 3B, C) and indeed observed a dynamic pattern of expression in the posterior PSM, validating it as a novel cyclic gene.

Discussion

In this work, we apply four distinct mathematical methods of pattern detection in high-dimensional gene expression data and we compare their predictions to those from the L analysis on the same microarray dataset in three ways.

(1) We first examined the location of a set of benchmark cyclic genes in the top 300 probe set list of each method; the probe sets being ordered by significance of the pattern of their expression profile. Remarkably, the new methods accurately identified a large percentage of benchmark cyclic genes with a very high ranking. In addition, they performed as well as the L method in detecting known cyclic genes (that were discovered independently from the microarray study), validating the use of these methods in detecting periodic profiles. Some of the methods were designed to detect any significant patterns independently from the nature of the pattern–whether or not it is periodic. Remarkably, these methods identified cyclic genes from among their top ranked candidates, suggesting that periodic patterns are predominant in this dataset. Due to technical issues, the enrichment in periodic profiles in the dataset is very likely indeed, because the time series is constituted from samples from five consecutive cycles that

were ordered based on their phase (revealed by the pattern of expression of *Lfng* by *in situ* hybridization) to reconstitute one oscillation [5] (. One of the consequences of this strategy is that the collapsed dataset generated by this procedure preserves periodic patterns associated with the segmentation clock, while it affects most other patterns, such as a linear increase with time.

(2) We then calculated the overlap between the sets of the top 300 ranked probe sets determined by each method. The intersection contained notably eight true positive cyclic genes. Applying different methods allows us to confirm the findings by the methods, but in addition, it leads to the identification of complementary types of patterns by identifying profiles that could be missed by some other methods.

(3) We independently clustered the top 300 ranked outputs generated by each method and tested the predicted candidate genes clustering with the known cyclic genes of the Wnt pathway. Many of the candidates appeared to be biologically consistent with previous data, since they were reported in the literature to be associated with the Wnt pathway. These are attractive potential new cyclic genes involved in the mouse segmentation clock. Furthermore, one of the candidate genes, *Cyr61*, which is a canonical Wnt target [17] and codes for a modulator of the Wnt signaling pathway [18], was identified by three of the methods and was experimentally validated in mouse PSM, extending the list of known cyclic genes associated with the Wnt pathway. Because it is technically difficult (time and labor intensive) to validate large numbers of candidates, it is usually easier to work with a small gene list resulting from the use of stringent criteria aimed at minimizing the percentage of false positives. This strategy, however, tends to exclude true positives. The other alternative is then to examine a larger list but with a higher rate of false positives. In this case, different strategies can be adopted to filter out false positives: clustering the profiles and checking the biological coherence (for example, co-regulation by the same signaling pathway) but also cross-validating the results from different analyses and concentrating on the candidates identified by more than one method. This proved to be a valid strategy, leading to the discovery of the new cyclic gene *Cyr61*.

An interesting feature common to the A, C and S methods is that they work on the ranked data as opposed to the raw amplitudes of the signal. In other words, the signal intensities that describe over time the expression of a gene are sorted by magnitude, and each signal intensity is then replaced by the integer rank within this sorted order. Thus, each gene expression profile is represented by a permutation, which is invariant over one-to-one transformations of the data (such as log or taking the square) and is much more mathematically tractable. These methods offer a particular advantage for the analysis of this segmentation clock time series where only the ordering of the time points could be estimated, but not the time interval in between. These methods, by taking the rank permutation, do not require such precise timing information. Since these methods perform similar to the L and P methods, which use the raw signal intensities, this suggests that moving to ranked data despite loosing some information (like the amplitude of the signal), might be advantageous in certain cases. Finally, our study suggests that these four new methods of pattern detection are valid in the detection of periodic profiles.

Furthermore, since these methods have no strict prior assumptions on the shape of the profiles of interest, they are promising exploratory tools to discover novel, interesting transcriptional patterns in large-scale expression analysis.

Materials and Methods

Description of the starting microarray dataset. Microarray data, available at ArrayExpress at www.ebi.ec.uk/arrayexpress/ under accession number E-TABM-163, were normalized as described [5] and filtered based on detection call (by removing the probe sets called "absent" throughout the experiment), signal intensity (by removing genes with low expression level <50) and amplitude (by eliminating peak-to-trough variation < 1.65). After these filters, the dataset consisted of 7,549 probe sets.

Cluster analysis. The expression profiles were clustered using K-means based on the Pearson correlation distance in MultiExperiment Viewer (MEV) software. The optimal number of clusters was determined using the Figure of Merit (FOM) function [19] in the MEV package that provides a measure of the fit of the expression patterns for the clusters produced by K-means.

PubMed search. The PubMed database was searched for each of the 142 genes in the Wnt clusters to identify articles indicating a link between these genes and the Wnt pathway. The results indicated some articles containing the two search terms: the "gene name" and "Wnt." Manual curation of these results was necessary to verify the biological connection between each gene and the Wnt pathway.

The search of the PubMed database was automated using the MedlineR library [20] for the R statistical language. This library utilizes the Entrez Programming Utilities interface to connect to the PubMed database. The number of matches for each pair of search terms was returned, as well as a link to the abstracts for each match. Fifteen genes were identified as novel Wnt cyclic gene candidates.

Experimental validation. The candidate cyclic gene *Cyr61* was experimentally validated by whole mount *in situ* hybridization that was performed as described [21] on 9.0 dpc mouse embryos using expressed sequence tag (ESTs) from Image clone 5716887 as a probe for *Cyr61*.

See attached PDF for Pattern Detection Methods

Acknowledgments
We thank Z.Otwinowski for helpful discussions and S. Esteban for artwork.

Author contributions. The following authors conceived and designed the algorithm indicated in parenthesis: SA and TMAF (A), YM and HE (S), AK and MR (P), JM, AS, LP and BS (C). The following authors implemented the algorithms: SA (A), YM (S), AK and MR (P), JM and AS (C), EFG (L). GH implemented the PubMed search algorithm. MLD prepared the microarray data (filtration), made the comparison analysis of the

different methods and performed experimental validation. MLD, ARM, AK, MR, HE, TMAF,and OP wrote the paper ?????

Funding. This research was partially supported by DARPA grant HR 0011-05-1-0057. HE and YM mathematical work supported by DARPA grant HR0011-05-1-0007. AS research was supported by a Lucent Technologies Bell Labs Graduate Research Fellowship; AK and MR research was supported by NIH grant GM U54 GM74942; and SA research was supported by Association pur le Recherche sur le Cancer (ARC), France. OP, AM, MLD, EG and GH research was supported by the Stowers Institute for Medical Research. OP is a Howard Hughes Medical Institute Investigator.

Competing interests. TMAF and SA have filed U.S. patent 20070086635, Method of identifying pattern in a series of data.

References

1. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9: 3273-3297.
2. McDonald MJ, Rosbash M (2001) Microarray analysis and organization of circadian gene expression in Drosophila. Cell 107: 567-578.
3. Claridge-Chang A, Wijnen H, Naef F, Boothroyd C, Rajewsky N, et al. (2001) Circadian regulation of gene expression systems in the Drosophila head. Neuron 32: 657-671.
4. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, et al. (2003) The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. PLoS Biol 1: E5.
5. Dequeant ML, Glynn E, Gaudenz K, Wahl M, Chen J, et al. (2006) A complex oscillating network of signaling genes underlies the mouse segmentation clock. Science 314: 1595-1598.
6. Palmeirim I, Henrique D, Ish-Horowicz D, Pourquié O (1997) Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesiss. Cell 91: 639-648.
7. McGrew MJ, Dale JK, Fraboulet S, Pourquie O (1998) The lunatic fringe gene is a target of the molecular clock linked to somite segmentation in avian embryos. Curr Biol 8: 979-982.
8. Aulehla A, Johnson RL (1999) Dynamic expression of lunatic fringe suggests a link between notch signaling and an autonomous cellular oscillator driving somite segmentation. Dev Biol 207: 49-61.
9. Forsberg H, Crozet F, Brown NA (1998) Waves of mouse Lunatic fringe expression, in four-hour cycles at two-hour intervals, precede somite boundary formation. CurrBiol 8: 1027-1030.
10. Glynn EF, Chen J, Mushegian AR (2006) Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. Bioinformatics 22: 310-316.

11. Aulehla A, Wehrle C, Brand-Saberi B, Kemler R, Gossler A, et al. (2003) Wnt3a plays a major role in the segmentation clock controlling somitogenesis. Dev Cell 4: 395-406.
12. Suriben R, Fisher DA, Cheyette BN (2006) Dact1 presomitic mesoderm expression oscillates in phase with Axin2 in the somitogenesis clock of mice. Dev Dyn 235: 3177-3183.
13. Jouve C, Palmeirim I, Henrique D, Beckers J, Gossler A, et al. (2000) Notch signalling is required for cyclic expression of the hairy-like gene HES1 in the presomitic mesoderm. Development 127: 1421-1429.
14. Ishikawa A, Kitajima S, Takahashi Y, Kokubo H, Kanno J, et al. (2004) Mouse Nkd1, a Wnt antagonist, exhibits oscillatory gene expression in the PSM under the control of Notch signaling. Mech Dev 121: 1443-1453.
15. Leimeister C, Dale K, Fischer A, Klamt B, Hrabe de Angelis M, et al. (2000) Oscillating expression of c-hey2 in the presomitic mesoderm suggests that the segmentation clock may use combinatorial signaling through multiple interacting bHLH factors. Dev Biol 227: 91-103.
16. Dunwoodie SL, Clements M, Sparrow DB, Sa X, Conlon RA, et al. (2002) Axial skeletal defects caused by mutation in the spondylocostal dysplasia/pudgy gene Dll3 are associated with disruption of the segmentation clock within the presomitic mesoderm. Development 129: 1795-1806.
17. Si W, Kang Q, Luu HH, Park JK, Luo Q, et al. (2006) CCN1/Cyr61 is regulated by the canonical Wnt signal and plays an important role in Wnt3A-induced osteoblast differentiation of mesenchymal stem cells. Mol Cell Biol 26: 2955-2964.
18. Latinkic BV, Mercurio S, Bennett B, Hirst EM, Xu Q, et al. (2003) Xenopus Cyr61 regulates gastrulation movements and modulates Wnt signalling. Development 130: 2429-2441.
19. Yeung KY, Haynor DR, Ruzzo WL (2001) Validating clustering for gene expression data. Bioinformatics 17: 309-318.
20. Lin SM, McConnell P, Johnson KF, Shoemaker J (2004) MedlineR: an open source library in R for Medline literature data mining. Bioinformatics 20: 3659-3661.
21. Henrique D, Adam J, Myat A, Chitnis A, Lewis J, et al. (1995) Expression of a Delta homologue in prospective neurons in the chick. Nature 375: 787-790.

Figure Legends

Figure 1. Identification of Benchmark Cyclic Genes in the Top 300 Probe Sets Lists of the Five Methods

(A) Benchmark genes are composed of cyclic genes identified independently from the Lomb-Scargle (L) analysis (seven probe sets).

(B) Benchmark genes also include cyclic genes identified by the L analysis and experimentally validated (27 probe sets).

L, Lomb-Scargle analysis; P, Phase consistency; A, Address reduction; C, Cyclohedron test; S, Stable persistence

Figure 2. Comparison of the Intersection of the Top 300 Ranked Probe Sets from the Five Methods
(A) Venn diagram.
(B) Haase diagram shows the pairwise intersection of two lists, the triple intersection of three lists, and so on. The total number of distinct probe sets in all of the five top 300 lists (the union) is 884; the total number in each of the five sets (the intersection) is 21.
L, Lomb-Scargle analysis; P, Phase consistency; A, Address reduction; C, Cyclohedron test; S, Stable persistence

Figure 3. Identification of *Cyr61* as a Novel Wnt-cyclic Gene
(A) Gene expression profiles (in $\log_2$ ratio) of *Cyr61* (represented by two probe sets) and benchmark Wnt-cyclic gene *Axin2*.
(B) and (C) Experimental validation by *in situ* hybridization. Lateral views of the tails of 9.0 dpc mouse embryos hybridized with the *Cyr61* probe. Two representative images illustrating the dynamic expression of the gene in the presomitic mesoderm (PSM) are shown. The domain of expression is in the posterior PSM (B) and the anterior PSM (C).

Supporting Information

Microarray data available at ArrayExpress at www.ebi.ec.uk/arrayexpress/ under accession number E-TABM-163

Table S1. List of Benchmark Cyclic Genes

The benchmark cyclic genes in bold were identified independently from the microarray analysis.
L, Lomb-Scargle analysis; P, Phase consistency; A, Address reduction; C, Cyclohedron test; S, Stable persistence

Table S2. List of the Top 300 Probe Sets of the Lomb-Scargle (L) Analysis

Table S3. List of the Top 300 Probe Sets of the Phase consistency (P) Analysis

Table S4. List of the Top 300 Probe Sets of the AddressReduction (A) Analysis

Table S5. List of the Top 300 Probe Sets of the Cyclohedron Test C Analysis

Table S6. List of the top 300 Probe Sets of the Stable Persistence (S) Analysis