Identification of Secondary Structure Elements in *miRNA* using Signal Processing Method

Submitted by: Nancy Singh

Roll number: 131505

May 2017

Submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Technology

in

Bioinformatics

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan Dist. 173234, Himachal Pradesh, India

CERTIFICATE

This is to certify that the work titled "**Identification of Secondary Structure In *miRNA* using Signal processing**" submitted by **Nancy Singh** in partial fulfillment for the award of degree of Bachelor of Technology in Bioinformatics of Jaypee University of Information Technology, Waknaghat, India, has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Signature:
Name: Dr. Ragothaman M Yennamalli
Date: 1/5/17
Designation: Assistant Professor
(Grade II)

Signature:
Name: Dr. Sunil Datt Sharma
Date: 01.05.2017
Designation: Assistant Professor
(Senior Grade)

ACKNOWLEDGEMENT

We would like to express our deepest appreciation to all those who provided us the possibility to complete this report. A special gratitude we give to our final year project supervisor, **Dr. RagothamanM. Yennamalli and Dr. Sunil Datt Sharma** whose contribution in stimulating suggestions and encouragement helped us to coordinate our project especially in writing this report. Furthermore, we would also like to acknowledge with much appreciation the crucial role of the staff of Bioinformatics lab, who gave the permission to use all required equipment and the necessary material to complete the task. Last but not least, we have to appreciate the guidance given by other supervisors and the evaluation panels especially in our project presentation that has improved our communication skills.

CONTENTS

LIST OF FIGURES

Figure 16: Spectrum corresponding to Periodicity 2-11 for sequences MI0000182 TO MI0000195

Figure 17: Spectrum corresponding to Periodicity 2-11 for sequences MI0000196 TO MI0000210

Figure 18: Spectrum corresponding to Periodicity 2-11 for sequences MI0000211 TO MI0000225

Figure 19: Spectrum corresponding to Periodicity 2-11 for sequences MI0000227 TO MI0000239

Figure 20: Spectrum corresponding to Periodicity 2-11 for sequences MI0000241 TO MI0000254

Figure 21: Spectrum corresponding to Periodicity 2-11 for sequences MI0000255 TO MI0000281

Figure 22: Spectrum corresponding to Periodicity 2-11 for sequences MI0000282 TO MI0000303

Figure23: Spectrum corresponding to Periodicity 2-11 for sequences MI0000304 TO MI0000315

Figure24: Spectrum corresponding to Periodicity 2-11 for sequences MI0000316 TO MI0000343

Figure25: Spectrum corresponding to Periodicity 2-11 for sequences MI0000354 TO MI0000366

Figure26: Spectrum corresponding to Periodicity 2-11 for sequences MI0000367 TO MI0000380

Figure 27: Spectrum corresponding to Periodicity 2-11 for sequences MI0000383 TO MI0000403

Figure 28: Spectrum corresponding to Periodicity 2-11 for sequences MI0000404 TO MI0000415

Figure 29: Spectrum corresponding to Periodicity 2-11 for sequences MI0000416 TO MI0000428

Figure 30: Spectrum corresponding to Periodicity 2-11 for sequences MI00004429 TO MI0000450

Figure 31: Spectrum corresponding to Periodicity 2-11 for sequences MI0000451 TO MI0000469

Figure 32: Spectrum corresponding to Periodicity 2-11 for sequences MI0000470 TO MI0000483

Figure 33: Spectrum corresponding to Periodicity 2-11 for sequences MI0000484 TO MI0000552

Figure 34: Spectrum corresponding to Periodicity 2-11 for sequences MI0000553 TO MI0000564

Figure 35: Spectrum corresponding to Periodicity 2-11 for sequences MI0000565 TO MI0000578

Figure 36: Spectrum corresponding to Periodicity 2-11 for sequences MI0000579 TO MI0000595

Figure 37: Spectrum corresponding to Periodicity 2-11 for sequences MI0000597 TO MI0000630

Figure 38: Spectrum corresponding to Periodicity 2-11 for sequences MI0000632 TO MI0000657

Figure 39: Spectrum corresponding to Periodicity 2-11 for sequences MI0000658 TO MI0000670

Figure 40: Spectrum corresponding to Periodicity 2-11 for sequences MI0000674 TO MI0000689

Figure 41: Spectrum corresponding to Periodicity 2-11 for sequences MI0000690 TO MI0000702

Figure 42: Spectrum corresponding to Periodicity 2-11 for sequences MI0000703 TO MI0000717

Figure 43: Spectrum corresponding to Periodicity 2-11 for sequences MI0000718 TO MI0000731

Figure 44: Spectrum corresponding to Periodicity 2-11 for sequences MI0000732 TO MI0000745

Figure 45: Spectrum corresponding to Periodicity 2-11 for sequences MI0000746 TO MI0000764

Figure 46: Spectrum corresponding to Periodicity 2-11 for sequences MI0000765 TO MI0000786

Figure 47: Spectrum corresponding to Periodicity 2-11 for sequences MI0000787 TO MI0000803

Figure 48: Spectrum corresponding to Periodicity 2-11 for sequences MI0000804 TO MI0000817

Figure 49: Spectrum corresponding to Periodicity 2-11 for sequences MI0000819 TO MI0000985

Figure 50: Spectrum corresponding to Periodicity 2-11 for sequences MI0000987 TO MI0001006

Figure 51: Spectrum corresponding to Periodicity 2-11 for sequences MI0001013 TO MI0001033

Figure 52: Spectrum corresponding to Periodicity 2-11 for sequences MI0001034 TO MI0001072

Figure 53: Spectrum corresponding to Periodicity 2-11 for sequences MI0001080 TO MI0001102

Figure 54: Spectrum corresponding to Periodicity 2-11 for sequences MI0001104 TO MI0001124

Figure 55: Spectrum corresponding to Periodicity 2-11 for sequences MI0001125 TO MI0001142

Figure 56: Spectrum corresponding to Periodicity 2-11 for sequences MI0001143 TO MI0001162

Figure 57: Spectrum corresponding to Periodicity 2-11 for sequences MI0001163 TO MI0001204

Figure 58: Spectrum corresponding to Periodicity 2-11 for sequences MI0001205 TO MI0001361

Figure 59: Spectrum corresponding to Periodicity 2-11 for sequences MI0001363 TO MI0001387

Figure 60: Spectrum corresponding to Periodicity 2-11 for sequences MI0001445 TO MI0001487

Figure 61: Spectrum corresponding to Periodicity 2-11 for sequences MI0001489 TO MI0001558

Figure 62: Spectrum corresponding to Periodicity 2-11 for sequences MI0001585 TO MI0001726

Figure 63: Spectrum corresponding to Periodicity 2-11 for sequences MI0001734 TO MI0001815

Figure 64: Spectrum corresponding to Periodicity 2-11 for sequences MI0001817 TO MI0001862

Figure 65: Spectrum corresponding to Periodicity 2-11 for sequences MI0001870 TO MI0001896

Figure 66: Spectrum corresponding to Periodicity 2-11 for sequences MI0001897 TO MI0001910

Figure 67: Spectrum corresponding to Periodicity 2-11 for sequences MI0001912 TO MI0001940

Figure 68: Spectrum corresponding to Periodicity 2-11 for sequences MI0001941 TO MI0001966

Figure 69: Spectrum corresponding to Periodicity 2-11 for sequences MI0001967 TO MI0001987

Figure 70: Spectrum corresponding to Periodicity 2-11 for sequences MI0001992 TO MI0002012

Figure 71: Spectrum corresponding to Periodicity 2-11 for sequences MI0002015 TO MI0002035

Figure 72: Spectrum corresponding to Periodicity 2-11 for sequences MI0002036 TO MI0002072

Figure 73: Spectrum corresponding to Periodicity 2-11 for sequences MI0002073 TO MI0002097

Figure 74: Spectrum corresponding to Periodicity 2-11 for sequences MI0002113 TO MI0002181

Figure 75: Spectrum corresponding to Periodicity 2-11 for sequences MI0002188 TO MI0002251

Figure 76: Spectrum corresponding to Periodicity 2-11 for sequences MI0002264 TO MI0002352

Figure 77: Spectrum corresponding to Periodicity 2-11 for sequences MI0002356 TO MI0002566

Figure 78: Spectrum corresponding to Periodicity 2-11 for sequences MI0002579 TO MI0003126

Figure 79: Spectrum corresponding to Periodicity 2-11 for sequences MI0003130 TO MI0003189

Figure 80: Spectrum corresponding to Periodicity 2-11 for sequences MI0003190 TO MI0003206

Figure 81: Spectrum corresponding to Periodicity 2-11 for sequences MI0003364 TO MI0003512

Figure 82: Spectrum corresponding to Periodicity 2-11 for sequences MI0003514 TO MI0003535

Figure 83: Spectrum corresponding to Periodicity 2-11 for sequences MI0003536 TO MI0003584

Figure 84: Spectrum corresponding to Periodicity 2-11 for sequences MI0003586 TO MI0003642

Figure 85: Spectrum corresponding to Periodicity 2-11 for sequences MI0003643 TO MI0003710

Figure 86: Spectrum corresponding to Periodicity 2-11 for sequences MI0003714 TO MI0003836

Figure 87: Spectrum corresponding to Periodicity 2-11 for sequences MI0003844 TO MI0004155

Figure 88: Spectrum corresponding to Periodicity 2-11 for sequences MI0004171 TO MI0004485

Figure 89: Spectrum corresponding to Periodicity 2-11 for sequences MI0004486 TO MI0004638

Figure 90: Spectrum corresponding to Periodicity 2-11 for sequences MI0004665 TO MI0004702

Figure 91: Spectrum corresponding to Periodicity 2-11 for sequences MI0004705 TO MI0004727

Figure 92: Spectrum corresponding to Periodicity 2-11 for sequences MI0004728 TO MI0004965

Figure 93: Spectrum corresponding to Periodicity 2-11 for sequences MI0004968 TO MI0005003

Figure 94: Spectrum corresponding to Periodicity 2-11 for sequences MI0005004 TO MI0005066

Figure 95: Spectrum corresponding to Periodicity 2-11 for sequences MI0005085 TO MI0005192

Figure 96: Spectrum corresponding to Periodicity 2-11 for sequences MI0005195 TO MI0005277

Figure 97: Spectrum corresponding to Periodicity 2-11 for sequences MI0005294 TO MI0005402

Figure 98: Spectrum corresponding to Periodicity 2-11 for sequences MI0005406 TO MI0005476

Figure 99: Spectrum corresponding to Periodicity 2-11 for sequences MI0005477 TO MI0005503

Figure 100: Spectrum corresponding to Periodicity 2-11 for sequences MI0005505 TO MI0005548

Figure 101: Spectrum corresponding to Periodicity 2-11 for sequences MI0005549 TO MI0005565

Figure 102: Spectrum corresponding to Periodicity 2-11 for sequences MI0005566 TO MI0005675

Figure 103: Spectrum corresponding to Periodicity 2-11 for sequences MI0005688 TO MI0005811

Figure 104: Spectrum corresponding to Periodicity 2-11 for sequences MI0005812 TO MI0005823

Figure 105: Spectrum corresponding to Periodicity 2-11 for sequences MI0005824 TO MI0005835

Figure 106: Spectrum corresponding to Periodicity 2-11 for sequences MI0005836 TO MI0005847

Figure 107: Spectrum corresponding to Periodicity 2-11 for sequences MI0005848 TO MI0005861

Figure 108: Spectrum corresponding to Periodicity 2-11 for sequences MI0005862 TO MI0005903

Figure 109: Spectrum corresponding to Periodicity 2-11 for sequences MI0005906 TO MI0005941

Figure 110: Spectrum corresponding to Periodicity 2-11 for sequences MI0005942 TO MI0005961

Figure 111: Spectrum corresponding to Periodicity 2-11 for sequences MI0005962 TO MI0005979

Figure 112: Spectrum corresponding to Periodicity 2-11 for sequences MI0005980 TO MI0006011

Figure 113: Spectrum corresponding to Periodicity 2-11 for sequences MI0006017 TO MI0006128

Figure 114: Spectrum corresponding to Periodicity 2-11 for sequences MI0006173 TO MI0006287

Figure 115: Spectrum corresponding to Periodicity 2-11 for sequences MI0006306 TO MI0006444

Figure 116: Spectrum corresponding to Periodicity 2-11 for sequences MI0006486 TO MI0006515

Figure 117: Spectrum corresponding to Periodicity 2-11 for sequences MI0006519 TO MI0006578

Figure 118: Spectrum corresponding to Periodicity 2-11 for sequences MI0006579 TO MI0006870

Figure 119: Spectrum corresponding to Periodicity 2-11 for sequences MI0006871 TO MI0007034

Figure 120: Spectrum corresponding to Periodicity 2-11 for sequences MI0007045 TO MI0007595

Figure 121: Spectrum corresponding to Periodicity 2-11 for sequences MI0007599 TO MI0007631

Figure 122: Spectrum corresponding to Periodicity 2-11 for sequences MI0007632 TO MI0007730

Figure 123: Spectrum corresponding to Periodicity 2-11 for sequences MI0007735 TO MI0008198

Figure 124: Spectrum corresponding to Periodicity 2-11 for sequences MI0008199 TO MI0008266

Figure 125: Spectrum corresponding to Periodicity 2-11 for sequences MI0008282 TO MI0008356

Figure 126: Spectrum corresponding to Periodicity 2-11 for sequences MI0008357 TO MI0008900

Figure 127: Spectrum corresponding to Periodicity 2-11 for sequences MI0008901 TO MI0008913

Figure 128: Spectrum corresponding to Periodicity 2-11 for sequences MI0008915 TO MI0008933

Figure 129: Spectrum corresponding to Periodicity 2-11 for sequences MI0008934 TO MI0009922

Figure 130: Spectrum corresponding to Periodicity 2-11 for sequences MI0009923 TO MI0009977

Figure 131: Spectrum corresponding to Periodicity 2-11 for sequences MI0009979 TO MI0010038

Figure 132: Spectrum corresponding to Periodicity 2-11 for sequences MI0010148 TO MI0010531

Figure 133: Spectrum corresponding to Periodicity 2-11 for sequences MI0010532 TO MI0010625

Figure 134: Spectrum corresponding to Periodicity 2-11 for sequences MI00010635 TO MI0010898

Figure 135: Spectrum corresponding to Periodicity 2-11 for sequences MI00010956 TO MI0011585

Figure 136: Spectrum corresponding to Periodicity 2-11 for sequences MI00011587 TO MI0012303

Figure 137: Spectrum corresponding to Periodicity 2-11 for sequences MI00012309 TO MI0012499

Figure 138: Spectrum corresponding to Periodicity 2-11 for sequences MI00012532 TO MI0013242

Figure 139: Spectrum corresponding to Periodicity 2-11 for sequences MI00013244 TO MI0013585

Figure 140: Spectrum corresponding to Periodicity 2-11 for sequences MI00013592 TO MI0014026

Figure 141: Spectrum corresponding to Periodicity 2-11 for sequences MI00014028 TO MI0014050

Figure 142: Spectrum corresponding to Periodicity 2-11 for sequences MI00014051 TO MI0014064

Figure 143: Spectrum corresponding to Periodicity 2-11 for sequences MI00014065 TO MI0014079

Figure 144: Spectrum corresponding to Periodicity 2-11 for sequences MI00014084 TO MI0014147

Figure 145: Spectrum corresponding to Periodicity 2-11 for sequences MI00014148 TO MI0014505

Figure 146: Spectrum corresponding to Periodicity 2-11 for sequences MI00014512 TO MI0014561

Figure 147: Spectrum corresponding to Periodicity 2-11 for sequences MI00014562 TO MI0015995

Figure 148: Spectrum corresponding to Periodicity 2-11 for sequences MI00016003 TO MI0016116

Figure 149: Spectrum corresponding to Periodicity 2-11 for sequences MI00016122 TO MI0016267

Figure 150: Spectrum corresponding to Periodicity 2-11 for sequences MI0001668 TO MI0016291

Figure 151: Spectrum corresponding to Periodicity 2-11 for sequences MI00016300 TO MI0016321

Figure 152: Spectrum corresponding to Periodicity 2-11 for sequences MI00016326 TO MI0016356

Figure 153: Spectrum corresponding to Periodicity 2-11 for sequences MI00016357 TO MI0016718

Figure 154: Spectrum corresponding to Periodicity 2-11 for sequences MI00016721 TO MI0017299

Figure 155: Spectrum corresponding to Periodicity 2-11 for sequences MI00017340 TO MI0017619

Figure 156: Spectrum corresponding to Periodicity 2-11 for sequences MI00017735 TO MI0018037

Figure 157: Spectrum corresponding to Periodicity 2-11 for sequences MI00018044 TO MI0019031

Figure 158: Spectrum corresponding to Periodicity 2-11 for sequences MI00019032 TO MI0019158

Figure 159: Spectrum corresponding to Periodicity 2-11 for sequences MI00019182 TO MI0020235

Figure 160: Spectrum corresponding to Periodicity 2-11 for sequences MI00020241 TO MI0021137

Figure 161: Spectrum corresponding to Periodicity 2-11 for sequences MI00021321 TO MI0021342

Figure 162: Spectrum corresponding to Periodicity 2-11 for sequences MI00021345 TO MI0021383

Figure 163: Spectrum corresponding to Periodicity 2-11 for sequences MI00021392 TO MI0021429

Figure 164: Spectrum corresponding to Periodicity 2-11 for sequences MI00021430 TO MI0022001

Figure 165: Spectrum corresponding to Periodicity 2-11 for sequences MI00022020 TO MI0023566

LIST OF TABLES

ABBREVIATIONS

miRNAs: Micro RNAs

MST: Modified S-transform

STFT: Short-time Fourier transforms

Abstract

In the past, signal processing method has been applied on genomic data to reveal the periodicities that determine a wide variety of biological functions, ranging from exon detection to microsatellite identification. But, not much has been focused on RNA based signal processing. MicroRNAs (miRNA) are associated with gene regulation and gene silencing and thus have wide applications. The functionality of the miRNA is highly associated with its secondary structures (stem, bulge, and loop). Here, we show that signal processing method can be applied to miRNA sequences by implementing the Modified S-transform transform (MST) method to detect the periodicities between 2 and 11. We observed that these periodicities are highly correlated with the secondary structure of miRNA and possibly can be used as indicators of secondary and tertiary structure formation.

Chapter 1 - Introduction

Signal processing is a technology frequently used in various disciplines, where information can be processed and implemented containing signals. These signals are broadly designated in different physical or symbolic formats. Signal processing can be applied in any aarea of engineering and sciences to reveal the hidden information. Signal processing algorithms are based on the mathematical concepts [1]. Some of the areas where signal processing methods have been applied successfully like bioinformatics, genomics, proteomics, biosignal processing, and forensic sciences, etc.

Genomic signal processing is a specific application of signal processing implemented in the field of genomics. It is defined as the analysis, processing, and use of genomic signals for gaining biological knowledge and the translation of that knowledge into systems-based applications. In the case of DNA, different types of periodicities have been identified with the help of signal processing methods (Table 1). For example, periodicity 3 is related to the exonic repeats, periodicity 5-3 is related to the telomeric/subtelomeric repeats, and periodicity 10 is related to DNA bendability. The list of periodicities is tabulated in Table 1.

Table 1: List of periodicities identified in DNA.

Periodicity	Repeats
3	Exonic Regions Repeats
5-6	Telomeric/subtelomeric Repeats
10-11	DNA bendability(helical repeat structure)
48-50	Centromeric Repeats
68	Beta-satellite DNA
102	Nucleosomal Structure in Eukaryotes
105-106	Isochores(low G+C Content)
~135	Dimeric Alu repeats Structure
~165	A rich homopolymeric DNA sequence in Alu repeats
~171	Alpha Satellite DNA
~300	Alu repeats
~680	DNA Bend sites

1.1 MicroRNAs

MicroRNAs (miRNAs) are small single-strand, non-coding RNAs about 16 to 22 nucleotides in length, which play important role in gene regulation by targeting specific mRNAs for cleavage or translational repression [2]. The miRNAs are also involved in many important biological processes, such as affecting stability, translation of mRNAs and negatively regulating gene expression in post-transcriptional processes [2, 3].

Normal expression of miRNAs has been associated with various diseases including cancer, thus making them interesting therapeutic targets [3]. The composite of secondary structural elements that comprise miRNAs could help to design the small molecules that modulate their function [3].

1.2 RNA Secondary Structure:

The RNA is a ribonucleic acid that consists of the nucleotides or bases, A, C, G and U, where Uracil (U) is chemically similar to the thymine (T) in the DNA. RNA is generally a single stranded molecule [4]. These nucleotides A/U and C/G can form hydrogen bonded base pairs, which are commonly known as complementary base pairs. Sometimes, the bases G and U can also form pairs [4]. In a given RNA sequence, if a complementary segment exists, then these segments can form consecutive base pairs that help the RNA to fold onto itself. This complementary base pairing determines the three-dimensional structure of the RNA to a considerable extent, and the two-dimensional structure resulting from these base pairing is known as the RNA secondary structure [4]. The different secondary structures are:

- **Stem-loop:** RNA stem-loop is an essential part of the RNA secondary structure of RNA. It can guide RNA folding, determine interactions in a ribozyme, protect messenger RNA (mRNA) from degradation, serve as a recognition motif for RNA binding proteins or act as a substrate for enzymatic reactions [5] (Figure 1A).
- **Bulges and Internal Loops:** Bulges and internal loops form when two double-helical tracts are separated on either one (bulge) or both strands (internal loops) by one or more unpaired nucleotides. Internal loops containing equal numbers of bases on each strand are symmetric, whereas they are asymmetric when the number of bases is different. The presence of an internal loop or bulge reduces the thermodynamic stability, when compared to a perfect double helix, but unpaired nucleotides are more readily accessible to protein or nucleic acid ligands, which often recognize such sites [5] (Figure 1A).
- **Pseudoknots**: When complementary sequences consisting of a hairpin or internal loop and a single-stranded region interact with each other by Watson-Crick base pairing, a pseudoknot

is formed. There can be two alternative hairpin formations, when a pseudoknot forms between a hairpin loop and a complementary single-stranded region. The formation of a pseudoknot creates an extended helical region through helical stacking of the hairpin double-helical stem and the newly formed loop-loop interaction helix [6] (Figure 1B).

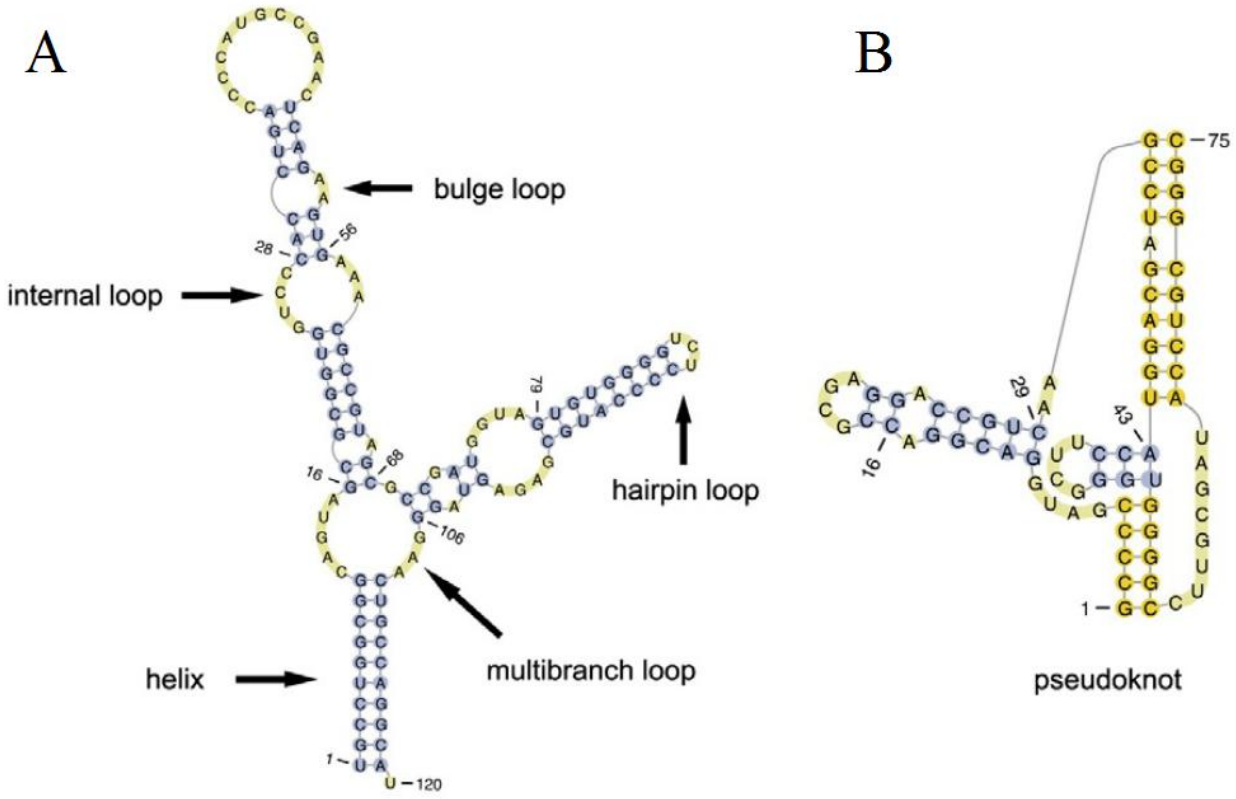


Figure 1: Secondary structure Elements of RNA. Image adapted from [7].

The RNA secondary structure is essentially governed by the base paring of nucleotides. Different computational methods have been proposed for finding the "optimal base pairing" of RNA in an efficient manner. Such algorithms are typically called RNA folding algorithm [6]. But not much has been done using signal processing methods for RNA based signal processing.

1.3 S-transform:

While, there are many RNA secondary structure algorithm available, there is no implementation of a signal processing method for detecting secondary structure. One such signal processing method that can be implemented in detection of RNA secondary structure is Modified S-Transform (MST). S-transform as a time–frequency distribution was developed in 1994 for analyzing geophysics data. In this way, the S transform is a generalization of the short-time Fourier transform (STFT), extending the continuous wavelet transform and overcoming some of its disadvantages. For one, modulation sinusoids are fixed with respect to the time axis; this localizes the scalable Gaussian window dilations and translations in S transform. Moreover, the S transform doesn't have a cross-term problem and yields a better signal clarity than Gabor transform (Gabor is the special case of short-term Fourier Transform. This is used to determine the sinusoidal frequency and the phase content of the signals as it change over time) [8].

The S-Transform of signal *x(t)* is defined as:

$$S(t,f,\sigma) = \int_{-\infty}^{\infty} x(\tau)\, G(t-\tau) e^{-2\pi f \tau} d\tau \qquad \text{(Eq. 1)}$$

Here, $G(t-\tau) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-(t-\tau)^2/2\sigma^2}$ which is Gaussian window function, σ is standard deviation, T is time, τ is central position of the window and f is frequency. The advantage of the S-transform over the other signal processing methods is that the standard deviation σ is the function of frequency, so the length of the Gaussian window varies with respect to the frequency and hence the equation above is become:

$$S(t,f) = \int_{-\infty}^{\infty} x(\tau) \left(\frac{f}{\sqrt{2\pi}}\right) e^{-(t-\tau)^2/2\sigma^2} e^{-2\pi f \tau} d\tau \qquad \text{(Eq. 2)}$$

The S-transform is also known as the short time Fourier transforms with variable window. Besides its advantages the S-transform has some faults i.e. it provides the progressive time frequency resolution but faces the problem of poor energy concentration in the time frequency plane because of the Gaussian window is inversely proportional to the frequency.

1.4 Modified S-transform:

MST is a modified version of S-transform method [8], where it has appropriate value of the parameter **γ** to control the window length for each specific frequency to improve the time and frequency resolution as well. The value of **γ** has been selected from 8.3 to 2.3 to vary the window length to efficiently capture the corresponding frequency **f** from $\frac{1}{2}$ and$\frac{1}{10}$ (period of repeated pattern 2bp to 10bp). For the higher frequencies, the window length is small and for lower frequencies, the window length is large [8]. Therefore, the MST is able to capture all the frequencies. So to get the appropriate window length to detect the corresponding period (period=1/frequency), we have to introduce the frequency-dependent control parameter γ using simulation studies, which is given by [8]:

$\gamma(f)=15f+0.8$ (Eq. 3)

Modified S-transform (MST) is defined as:

$$S(t,f) = \int_{-\infty}^{\infty} x(\tau) \left(\frac{f}{(15f+0.8)\sqrt{2\pi}}\right) e^{-(t-\tau)^2 f^2/2(15f+0.8)^2} e^{-2\pi f\tau} d\tau \quad \text{(Eq. 4)}$$

1.5 Hypothesis:

miRNAs are involved in many important biological processes, such as affecting stability, translation of mRNAs and negatively regulating gene expression in post-transcriptional processes. In this project, we have tried to apply signal processing method for the detection of periodicity present in miRNA sequences and correlated them with secondary structures of miRNA.

Chapter 2 - Material and Methods

2.1 miRNA sequences

The miRNA sequences used in the study is obtained from the miRBase v21 database [9]. These sequences are labeled as "high confidence" data. The total number of high confidence sequences has increased by 168 to 1996. That increase is partly due to our incorporation of more deep sequencing datasets.

High confidence data: The rate of deposition of novel miRNAs and the number of researchers involved in their discovery continue to increase, driven largely by small RNA deep sequencing experiments. In the face of these increases, and a range of miRNA annotation methods and criteria, maintaining the quality of the miRNA sequence data set is a significant challenge. The recent developments of the miRBase database have addressed this issue. In particular, they describe the collation and use of deep sequencing data sets to assign levels of confidence to miRBase entries. They provide the high confidence subset of miRBase entries, based on the pattern of mapped reads [9].

2.2 Mfold Prediction of RNA secondary Structures

The structures of these high confidence sequences are predicted from M-fold server [10, 11]. The 'Mfold' software for RNA folding was developed in the late 1980s. The 'M' simply refers to 'multiple'. The core algorithm predicts a minimum free energy, ΔG, as well as minimum free energies for folding that must contain any particular base pair. Any base pair, r_i–r_j, between the *i*th nucleotide and the *j*th nucleotide that is contained in a folding no more than $\delta\delta G$ from the minimum, is plotted in a triangular plot called the 'energy dot plot'. The base pair r_i–r_j is plotted in row *i* and column *j* of this matrix. The free energy increment, $\delta\delta G$, is chosen *a priori* by the

user, who selects ‘percent sub optimality’, *P*. From this, $\delta\delta G$ is computed to be $P/100\ |\Delta G|$. Base pairs within this free energy increment are chosen either automatically, or else by the user, and folding that contain the chosen base pair are computed. They have minimum free energy conditional on containing the chosen base pair [12].

2.3 Methodology

The methodology used in this study is shown in Figure 2, where the sequences obtained from mirBase are formatted and given as input to MST algorithm. The peaks obtained are analyzed and the region giving the peaks at periodicities 2 to 11 are recorded. At the same time, the same sequence is given to Mfold server and we obtain the list of predicted secondary structure elements. Further analysis was carried out by correlating the region of secondary structure to the region where peaks are observed.

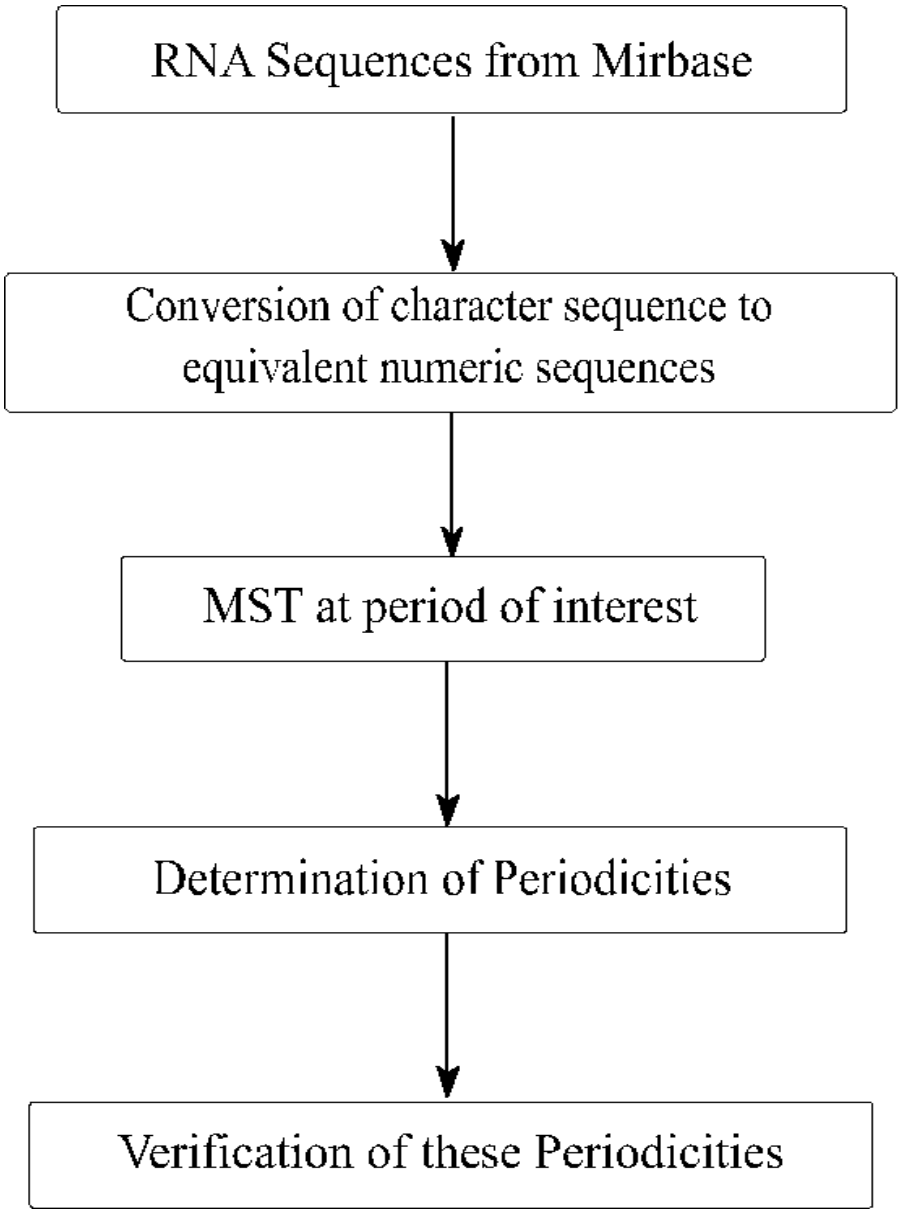


Figure 2: Outline of the method used in the Modified S-Transform.

Chapter 3 - Results

3.1 RNA Input Sequences

There are 1996 sequences of high confidence data which we obtain form mirBase site [9]. These are in one file i.e. “high_conf.dat”. This file contain all the information about these sequences which include there accession number, sequence, Pubmed id, paper name, author name etc. This file then parsed through the Perl program which give the output file that contain only the accession number and the sequences of these high confidence data. Then these sequences converted in to dat files for MST analysis. So these .dat files are used in the MST algorithm to obtain the spectrums.

3.2 MST results

The MST algorithm was run on the high confidence miRNA sequences spectrums. The spectrum obtained corresponding to the periodicities from 2 to11. The x-axis of the spectrum is the length of the sequence and the y-axis shows the periodicities form 2-11. The highlighted regions are shown by red color and they are considered as significant peaks. The spectrum for periodicity 2 to 11 are shown in Figure 3 to 169, for all the 1996 high-confidence miRNA sequences.

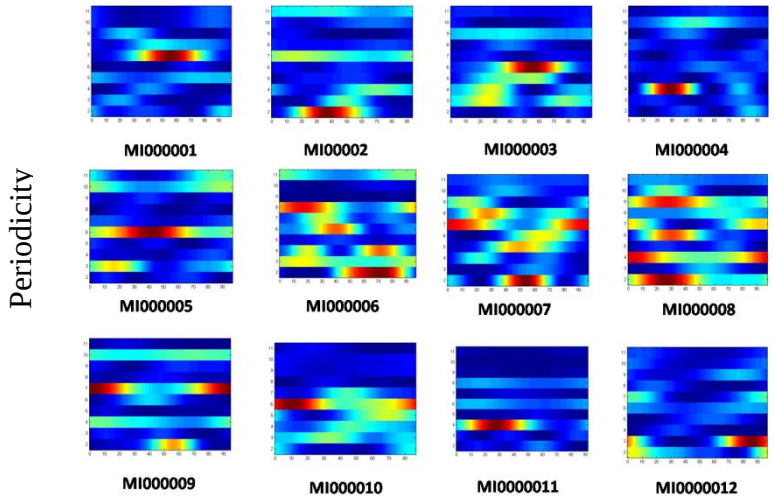


Figure 3: Spectrum corresponding to Periodicity 2-11 for sequences MI0000001 TO MI0000012

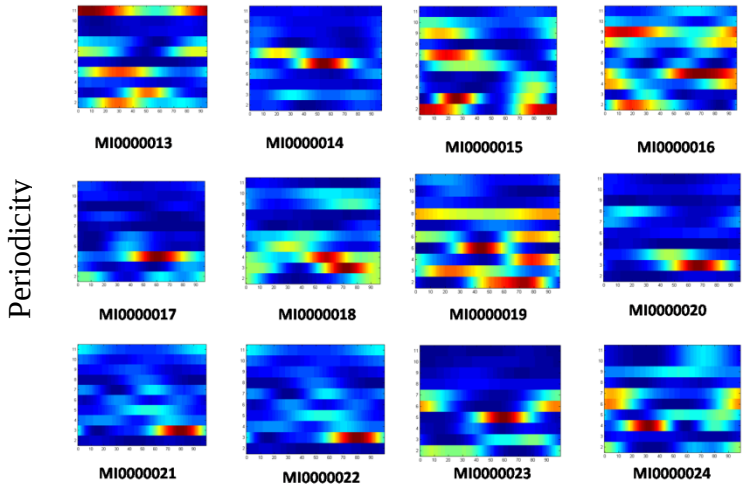


Figure 4: Spectrum corresponding to Periodicity 2-11 for sequences MI0000013 TO MI00000024

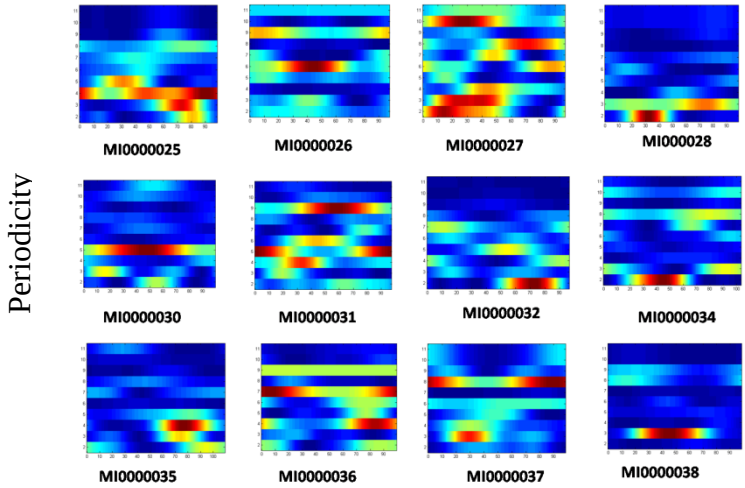


Figure 5: Spectrum corresponding to Periodicity 2-11 for sequences MI0000025 TO MI00000038

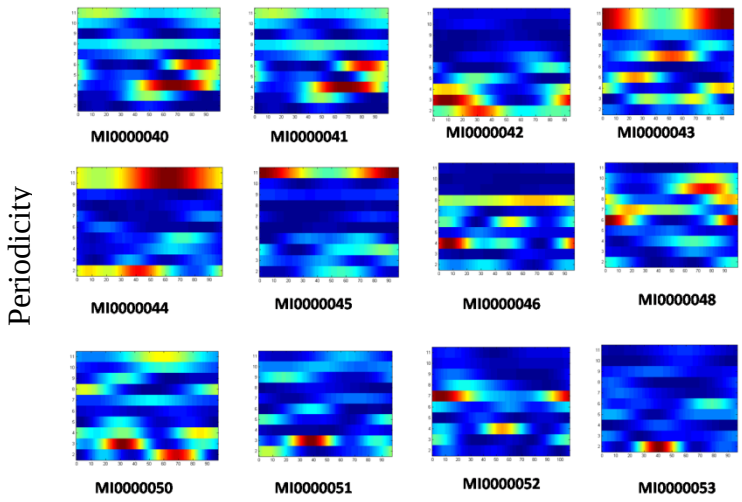


Figure 6: Spectrum corresponding to Periodicity 2-11 for sequences MI00000040 TO MI00000053

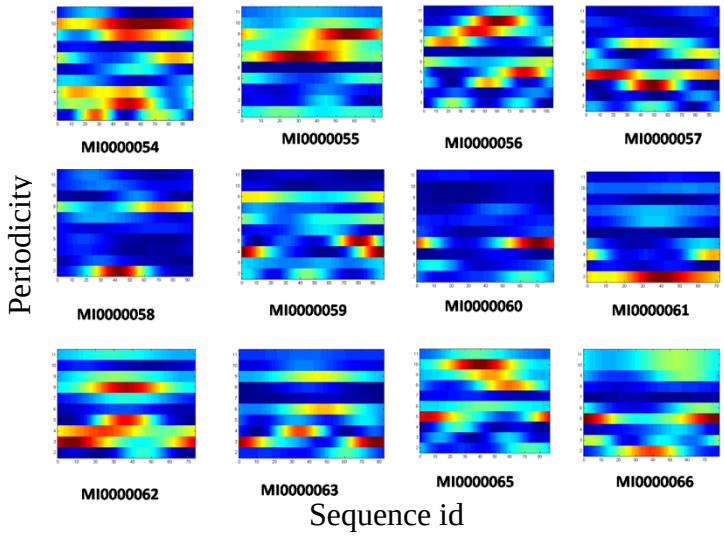


Figure 7: Spectrum corresponding to Periodicity 2-11 for sequences MI0000054 TO MI00000066

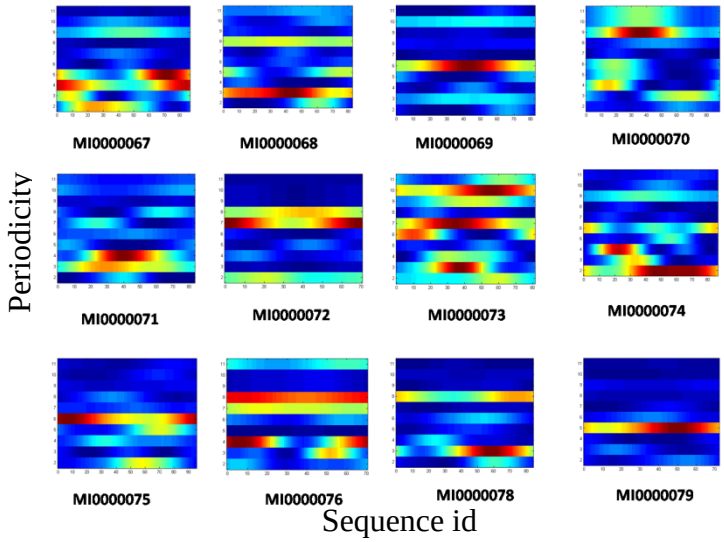


Figure 8: Spectrum corresponding to Periodicity 2-11 for sequences MI0000067 TO MI00000079

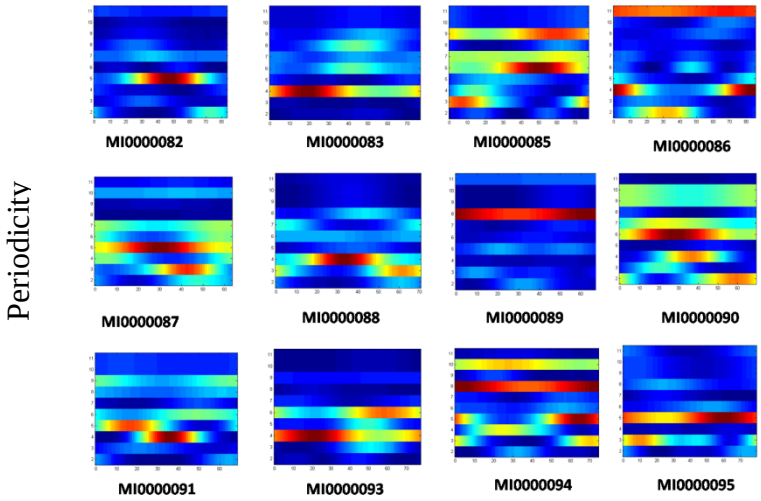


Figure 9: Spectrum corresponding to Periodicity 2-11 for sequences MI0000082 TO MI00000095

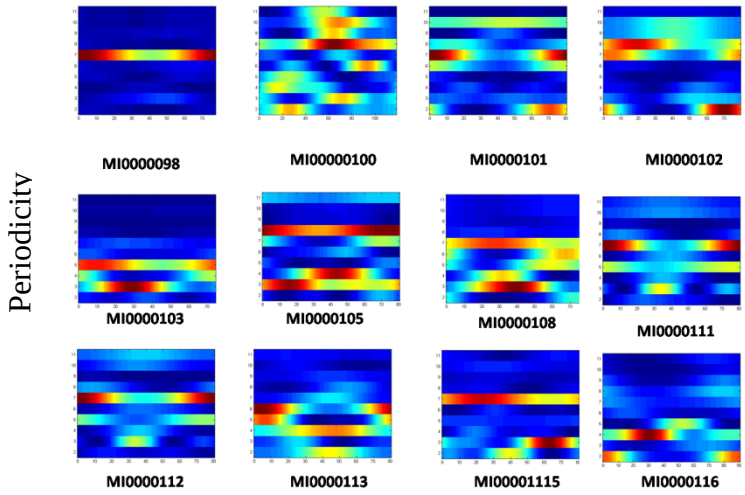


Figure 10: Spectrum corresponding to Periodicity 2-11 for sequences MI0000098 TO MI00000116

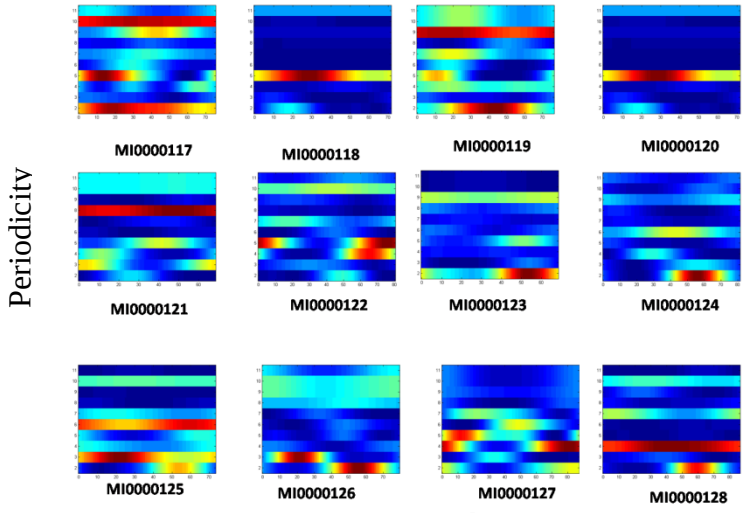


Figure 11: Spectrum corresponding to Periodicity 2-11 for sequences MI0000117 TO MI00000128

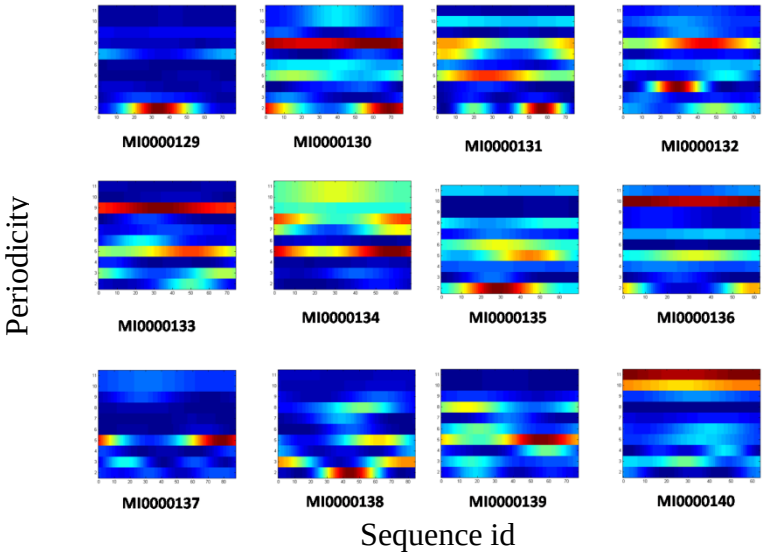


Figure 12: Spectrum corresponding to Periodicity 2-11 for sequences MI0000140 TO MI0000129

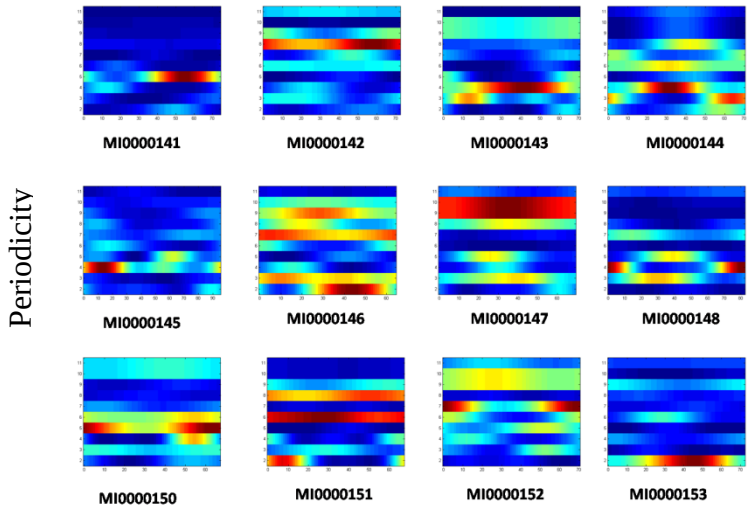


Figure 13: Spectrum corresponding to Periodicity 2-11 for sequences MI0000141 TO MI0000153

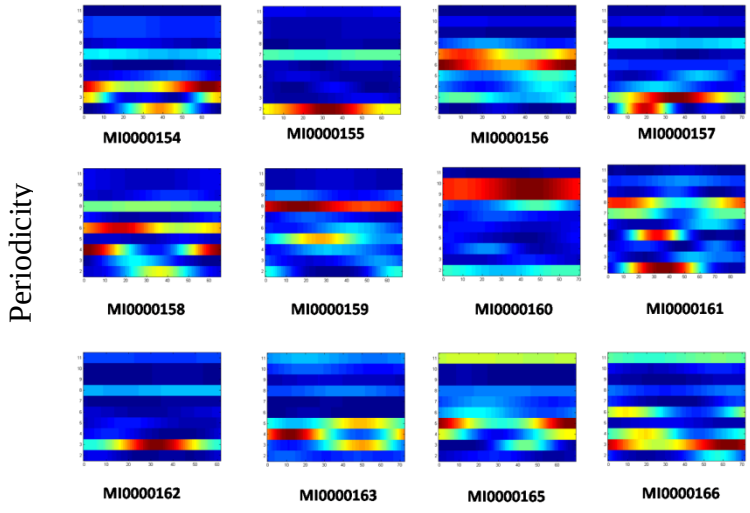


Figure 14: Spectrum corresponding to Periodicity 2-11 for sequences MI0000154 TO MI0000166

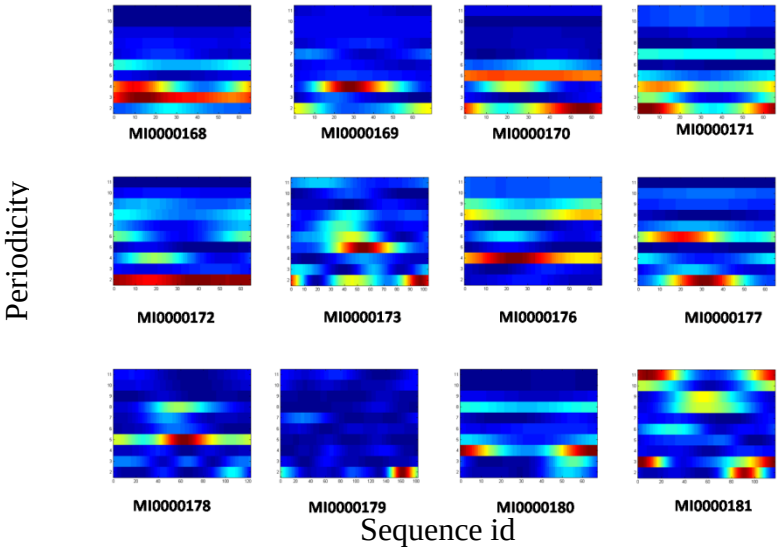


Figure 15: Spectrum corresponding to Periodicity 2-11 for sequences MI0000168 TO MI0000181

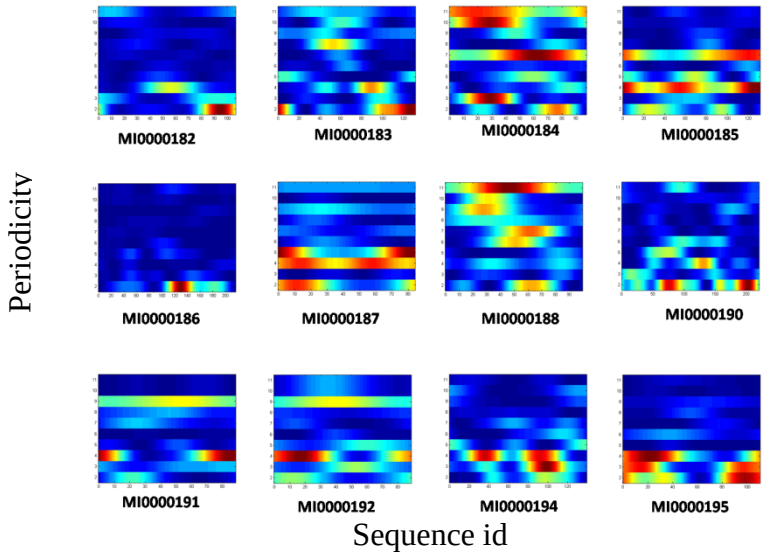


Figure 16: Spectrum corresponding to Periodicity 2-11 for sequences MI0000182 TO MI0000195

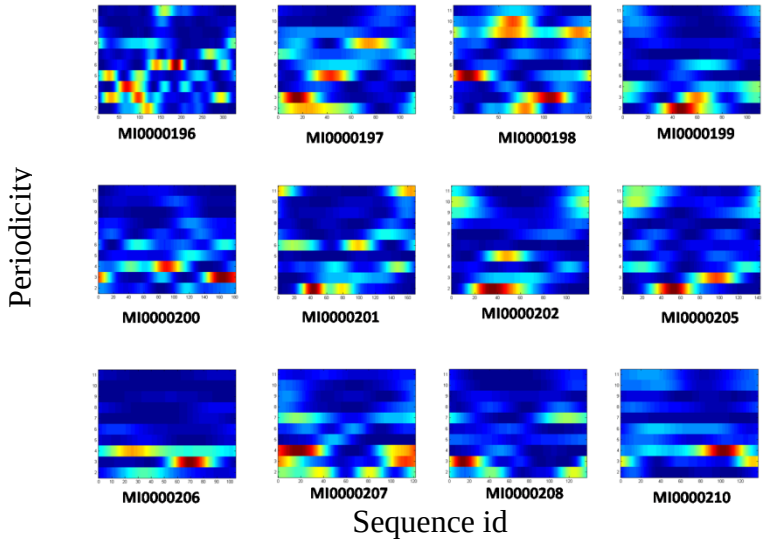


Figure 17: Spectrum corresponding to Periodicity 2-11 for sequences MI0000196 TO MI0000210

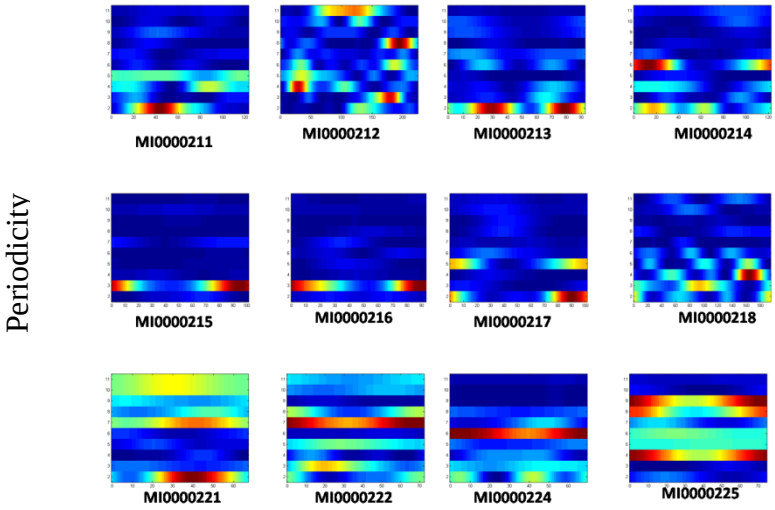


Figure 18: Spectrum corresponding to Periodicity 2-11 for sequences MI0000211 TO MI0000225

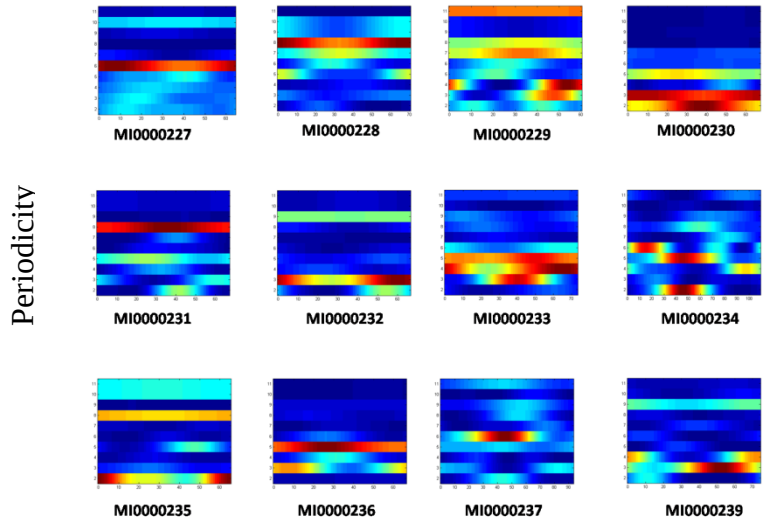


Figure 19: Spectrum corresponding to Periodicity 2-11 for sequences MI0000227 TO MI0000239

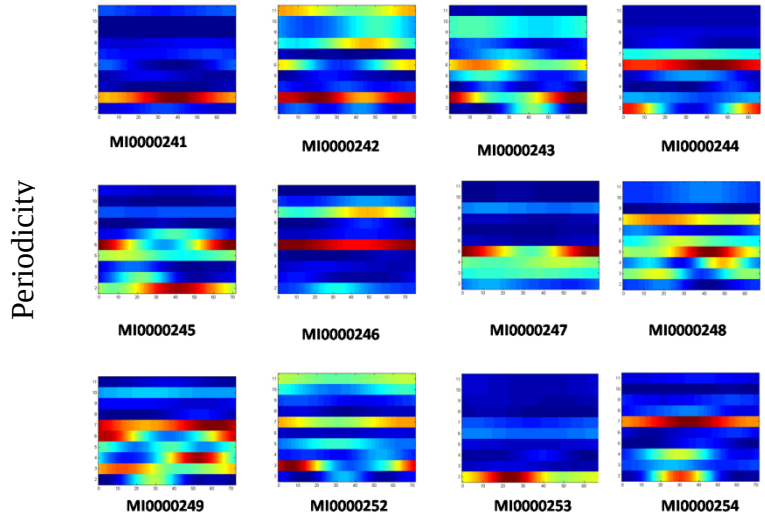


Figure 20: Spectrum corresponding to Periodicity 2-11 for sequences MI0000241 TO MI0000254

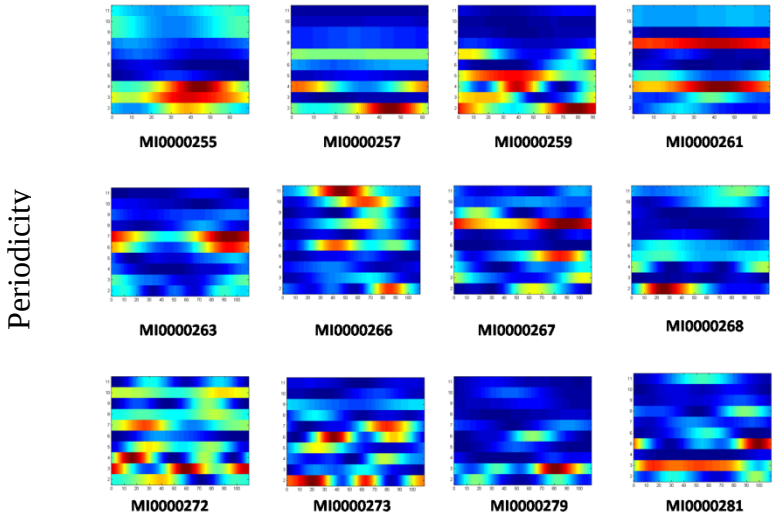


Figure 21: Spectrum corresponding to Periodicity 2-11 for sequences MI0000255 TO MI0000281

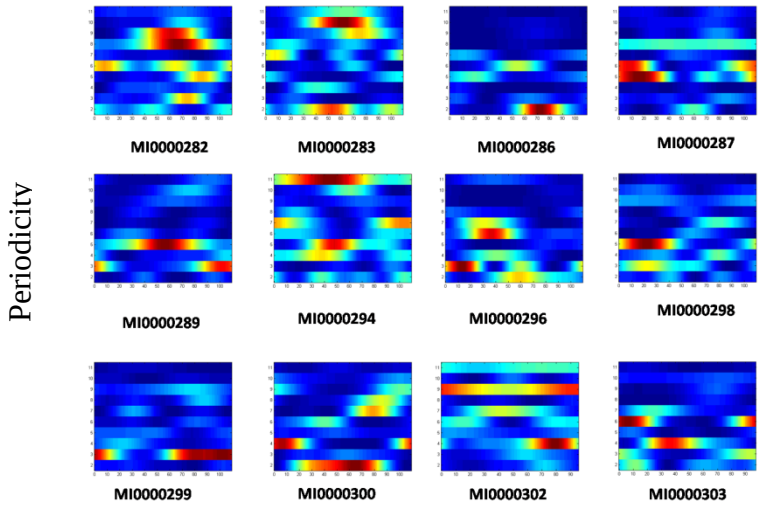


Figure 22: Spectrum corresponding to Periodicity 2-11 for sequences MI0000282 TO MI0000303

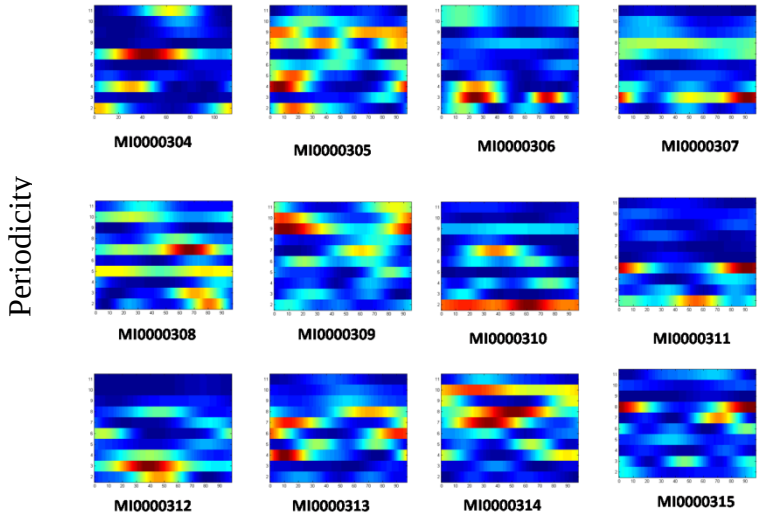


Figure23: Spectrum corresponding to Periodicity 2-11 for sequences MI0000304 TO MI0000315

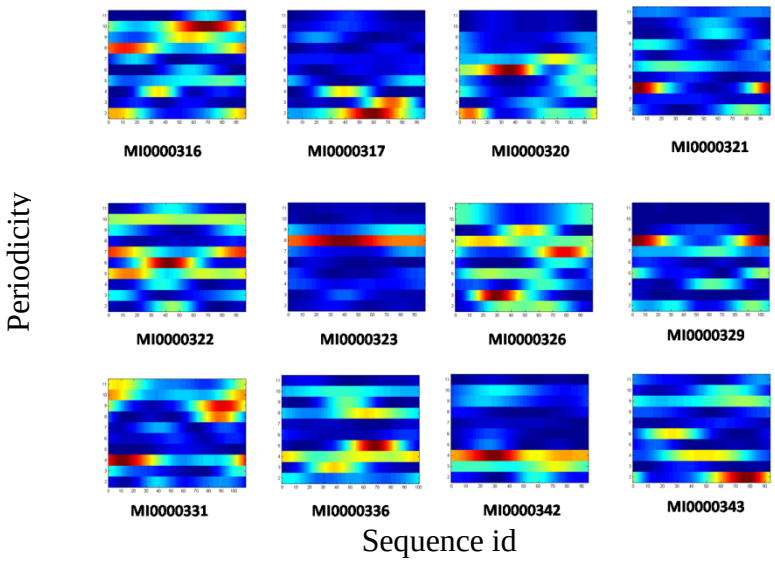


Figure24: Spectrum corresponding to Periodicity 2-11 for sequences MI0000316 TO MI0000343

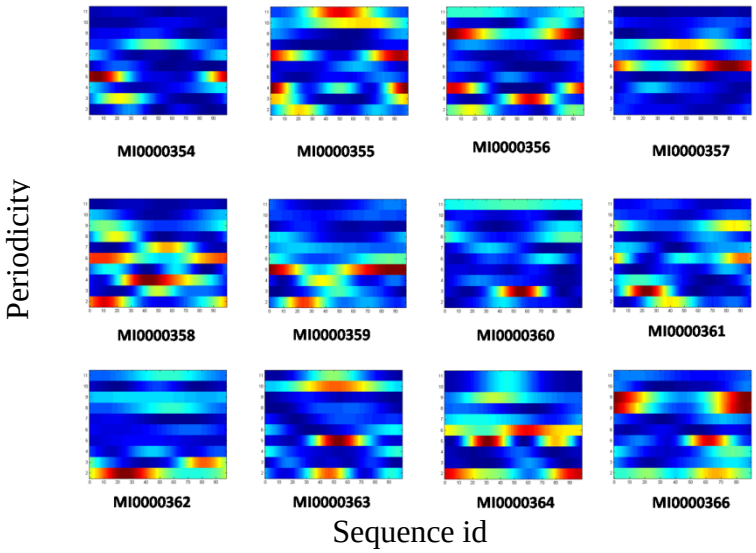


Figure25: Spectrum corresponding to Periodicity 2-11 for sequences MI0000354 TO MI0000366

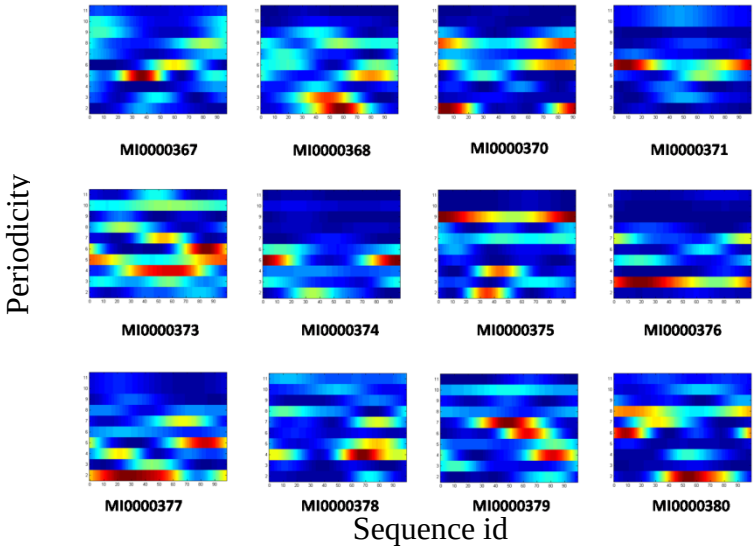


Figure26: Spectrum corresponding to Periodicity 2-11 for sequences MI0000367 TO MI0000380

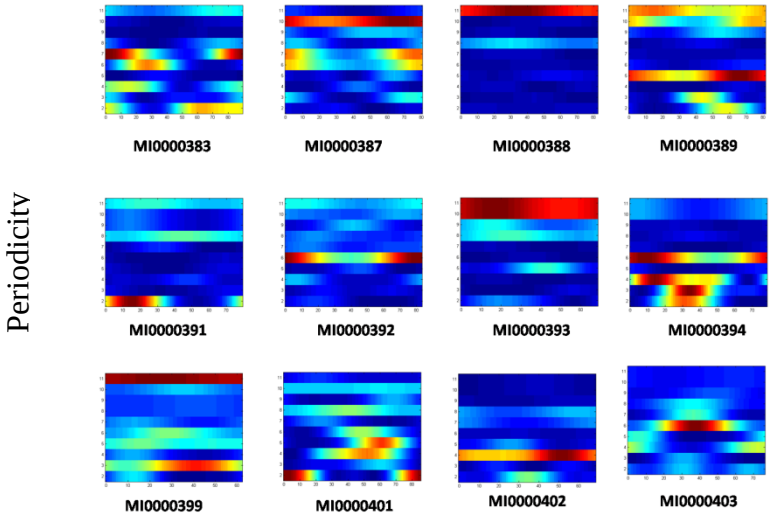


Figure 27: Spectrum corresponding to Periodicity 2-11 for sequences MI0000383 TO MI0000403

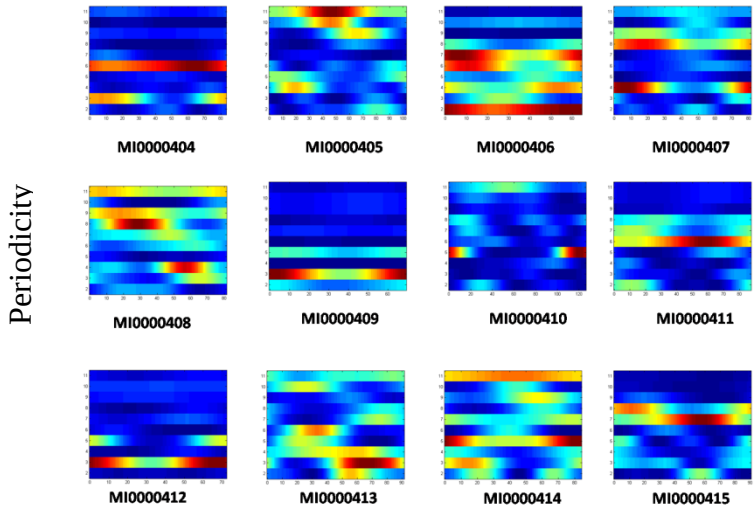


Figure 28: Spectrum corresponding to Periodicity 2-11 for sequences MI0000404 TO MI0000415

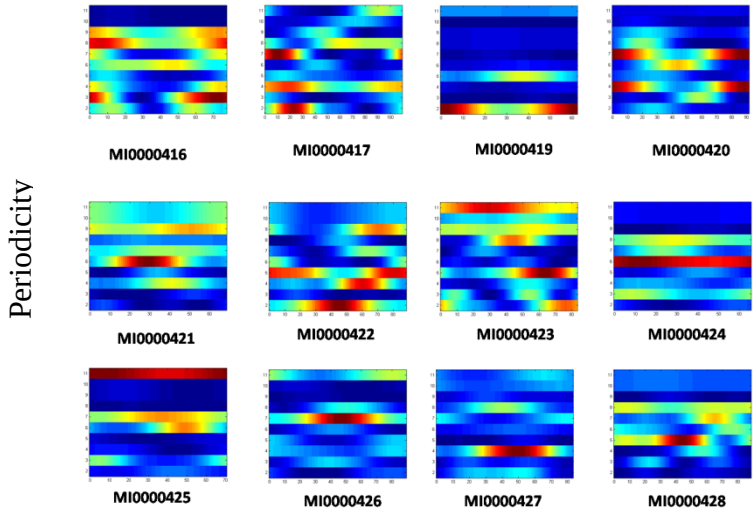


Figure 29: Spectrum corresponding to Periodicity 2-11 for sequences MI0000416 TO MI0000428

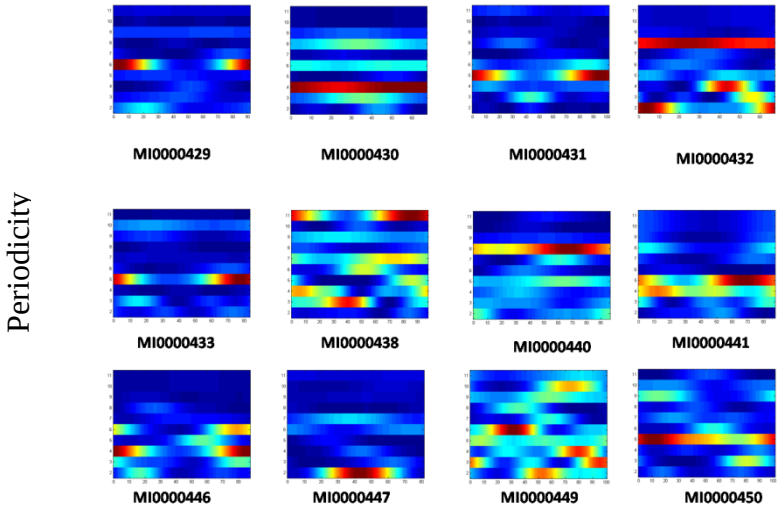


Figure 30: Spectrum corresponding to Periodicity 2-11 for sequences MI00004429 TO MI0000450

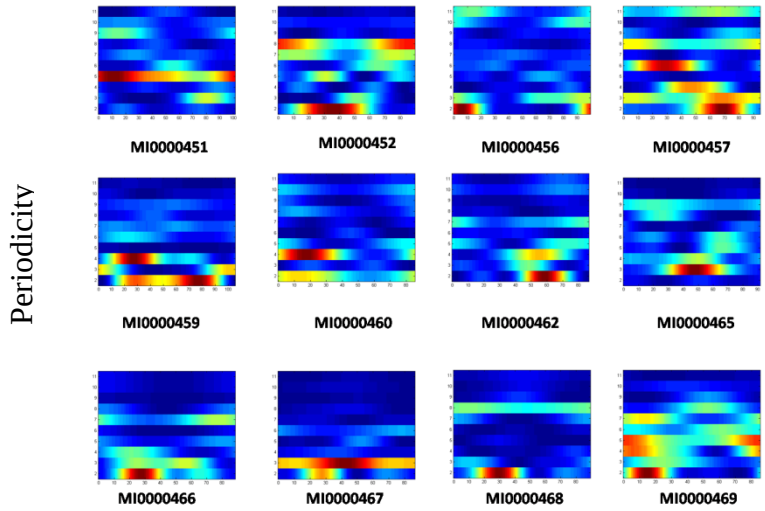


Figure 31: Spectrum corresponding to Periodicity 2-11 for sequences MI0000451 TO MI0000469

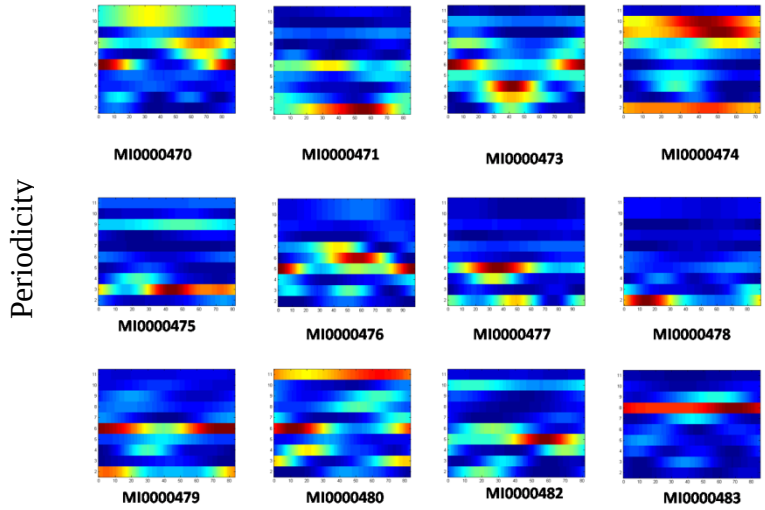


Figure 32: Spectrum corresponding to Periodicity 2-11 for sequences MI0000470 TO MI0000483

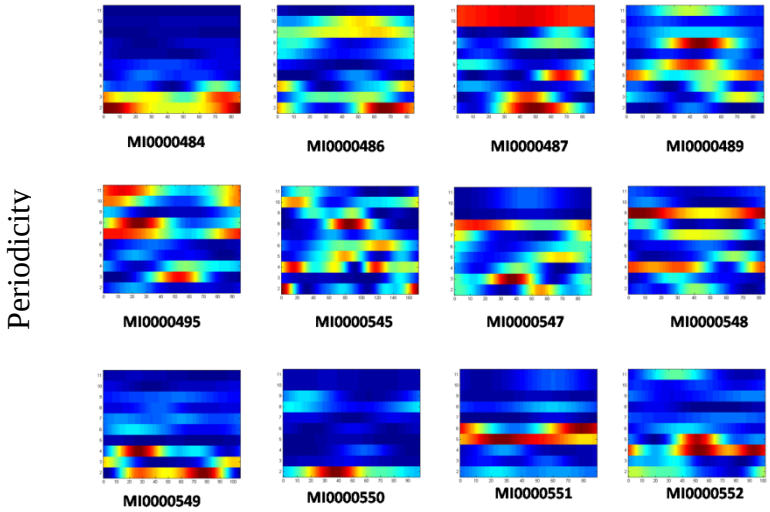


Figure 33: Spectrum corresponding to Periodicity 2-11 for sequences MI0000484 TO MI0000552

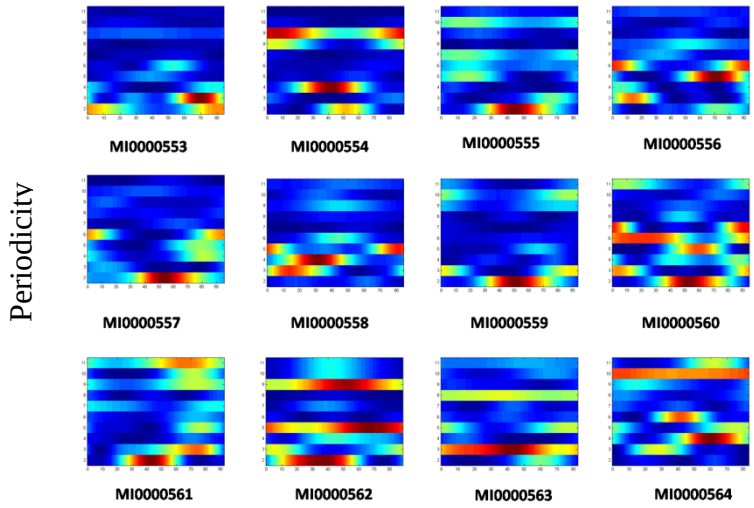


Figure 34: Spectrum corresponding to Periodicity 2-11 for sequences MI0000553 TO MI0000564

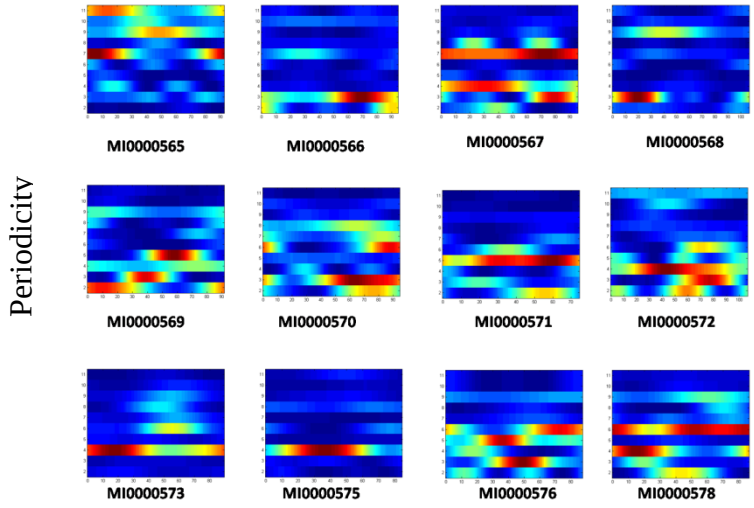


Figure 35: Spectrum corresponding to Periodicity 2-11 for sequences MI0000565 TO MI0000578

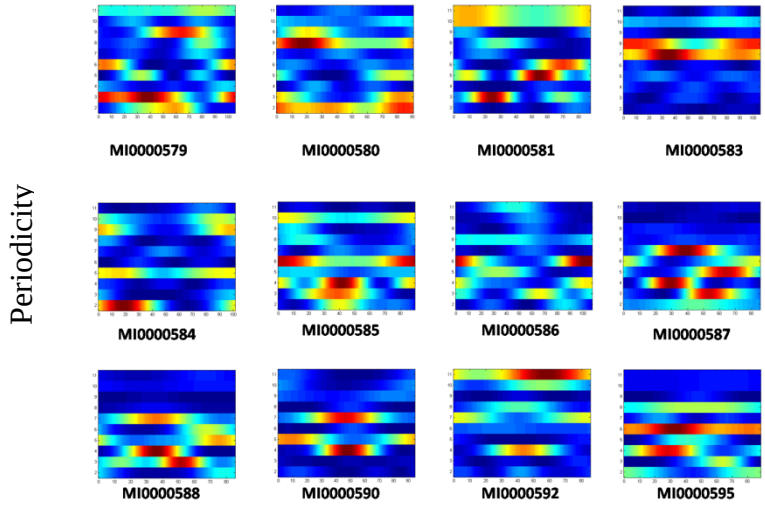


Figure 36: Spectrum corresponding to Periodicity 2-11 for sequences MI0000579 TO MI0000595

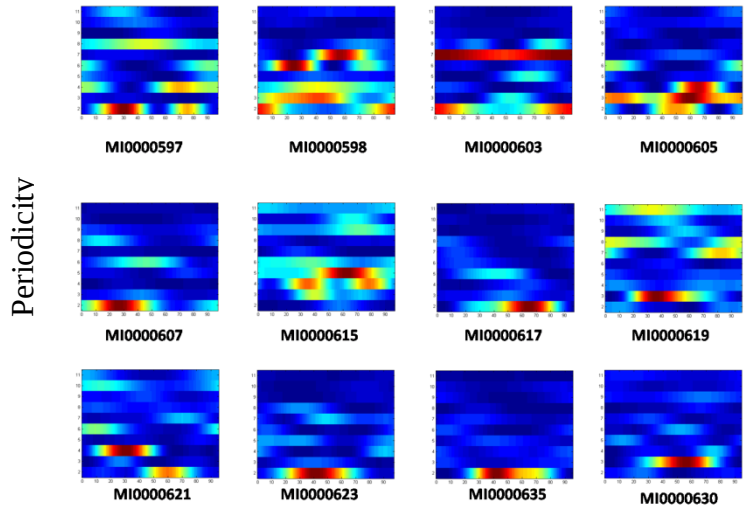


Figure 37: Spectrum corresponding to Periodicity 2-11 for sequences MI0000597 TO MI0000630

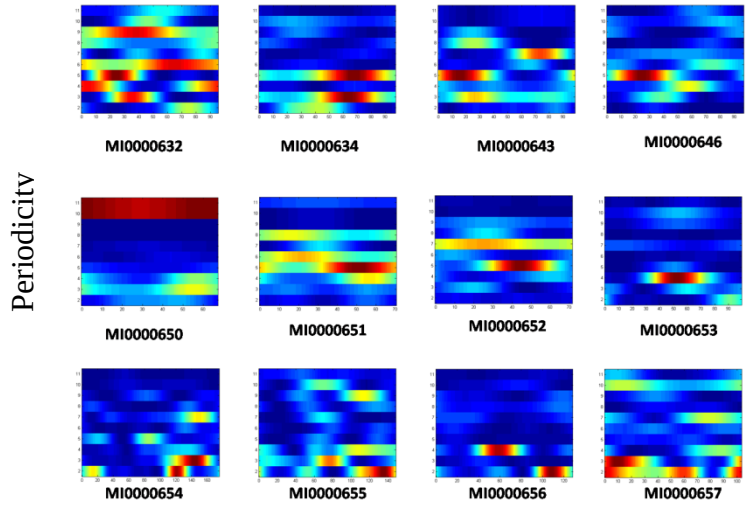


Figure 38: Spectrum corresponding to Periodicity 2-11 for sequences MI0000632 TO MI0000657

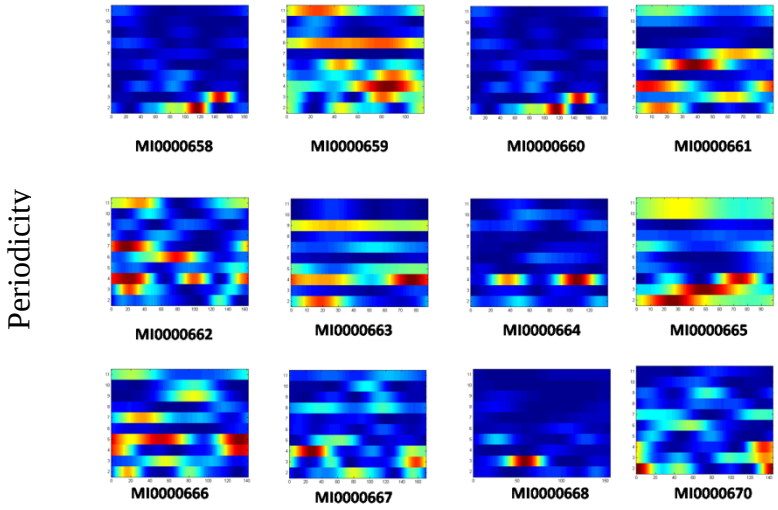


Figure 39: Spectrum corresponding to Periodicity 2-11 for sequences MI0000658 TO MI0000670

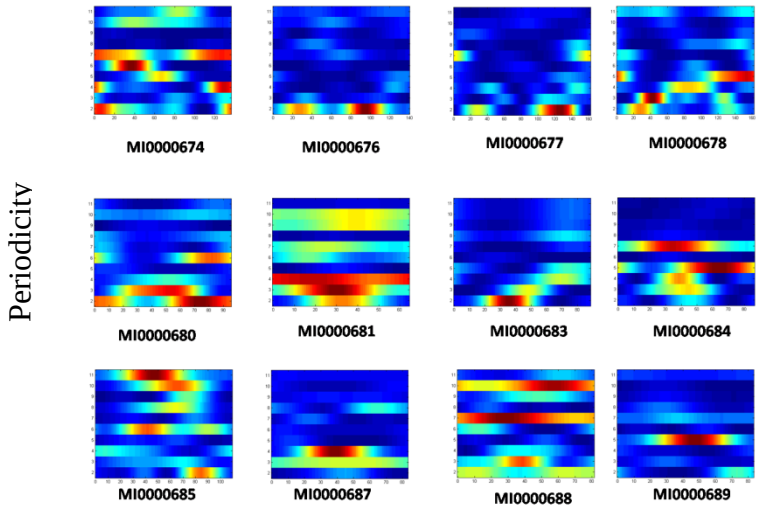


Figure 40: Spectrum corresponding to Periodicity 2-11 for sequences MI0000674 TO MI0000689

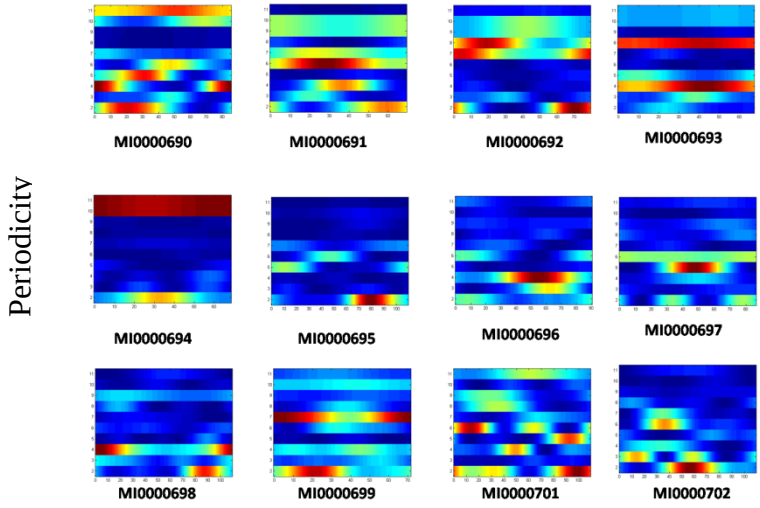


Figure 41: Spectrum corresponding to Periodicity 2-11 for sequences MI0000690 TO MI0000702

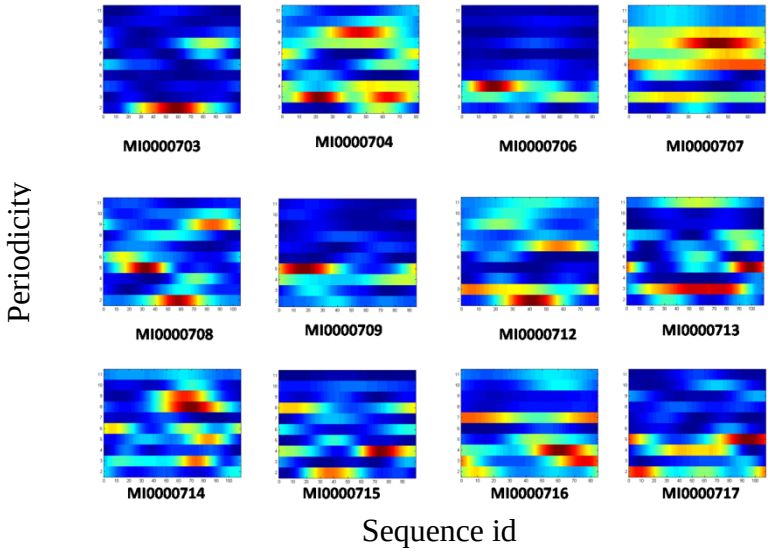


Figure 42: Spectrum corresponding to Periodicity 2-11 for sequences MI0000703 TO MI0000717

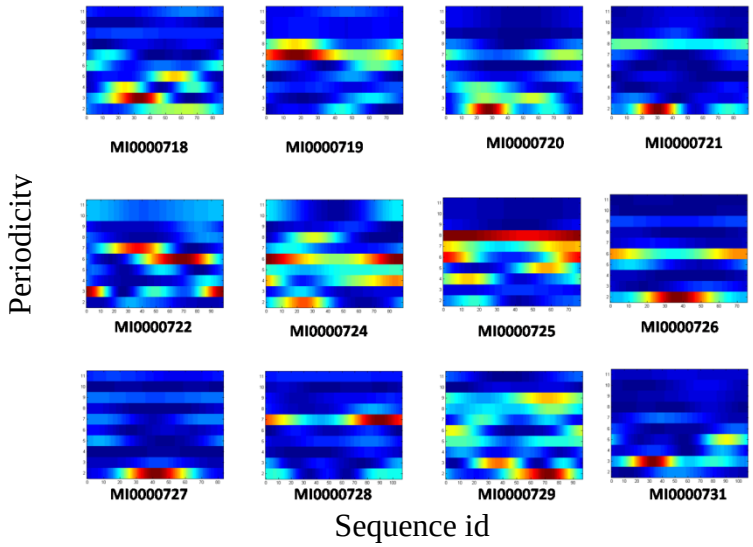


Figure 43: Spectrum corresponding to Periodicity 2-11 for sequences MI0000718 TO MI0000731

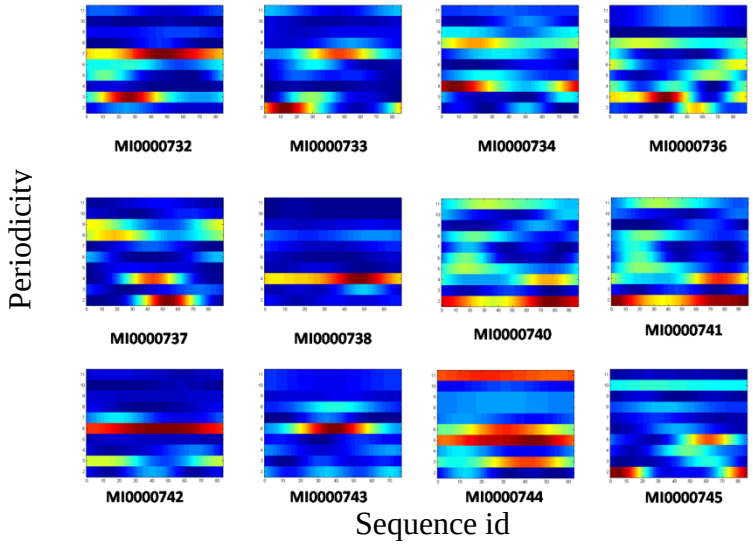


Figure 44: Spectrum corresponding to Periodicity 2-11 for sequences MI0000732 TO MI0000745

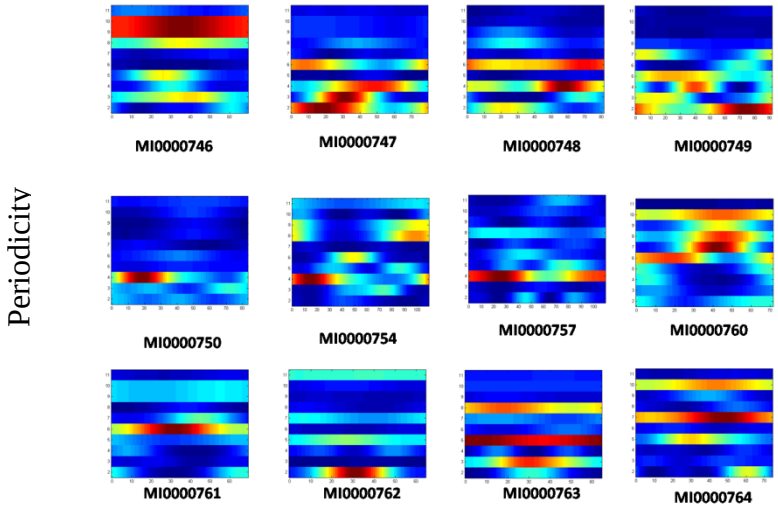


Figure 45: Spectrum corresponding to Periodicity 2-11 for sequences MI0000746 TO MI0000764

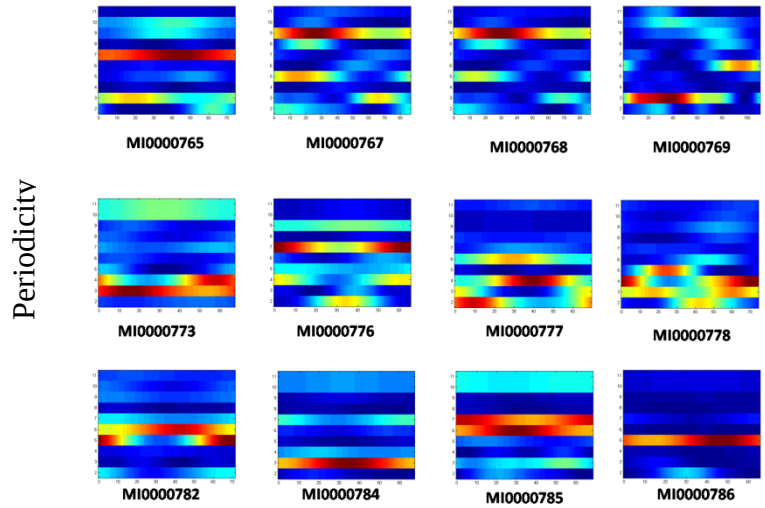


Figure 46: Spectrum corresponding to Periodicity 2-11 for sequences MI0000765 TO MI0000786

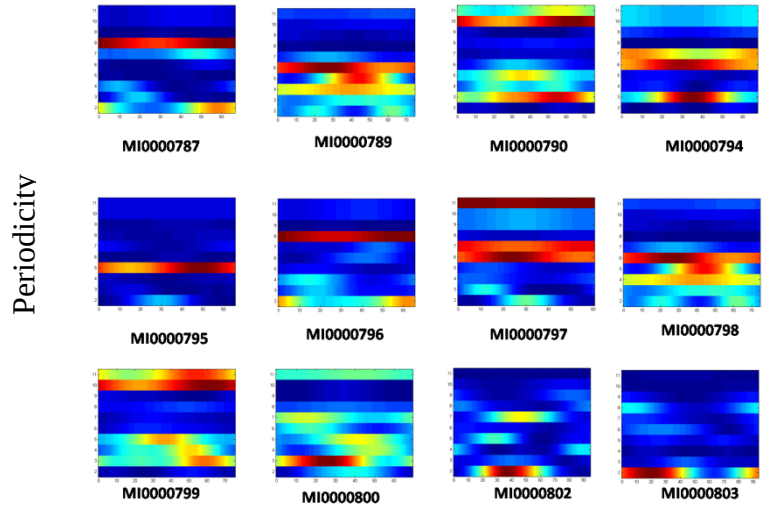


Figure 47: Spectrum corresponding to Periodicity 2-11 for sequences MI0000787 TO MI0000803

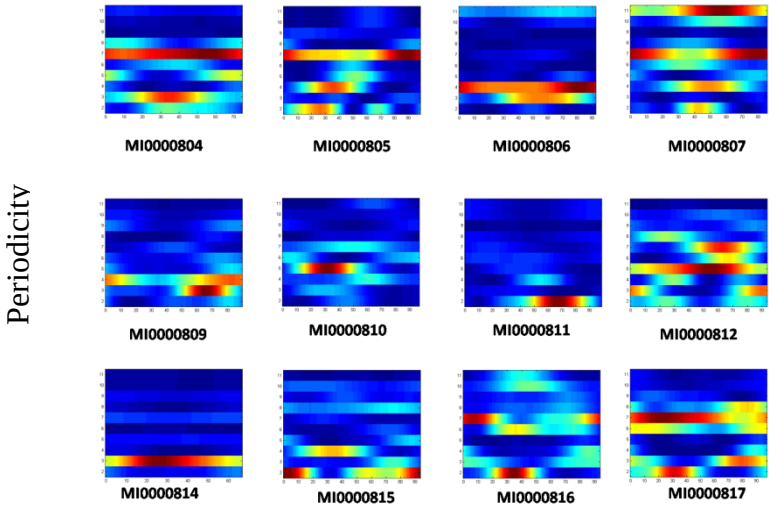


Figure 48: Spectrum corresponding to Periodicity 2-11 for sequences MI0000804 TO MI0000817

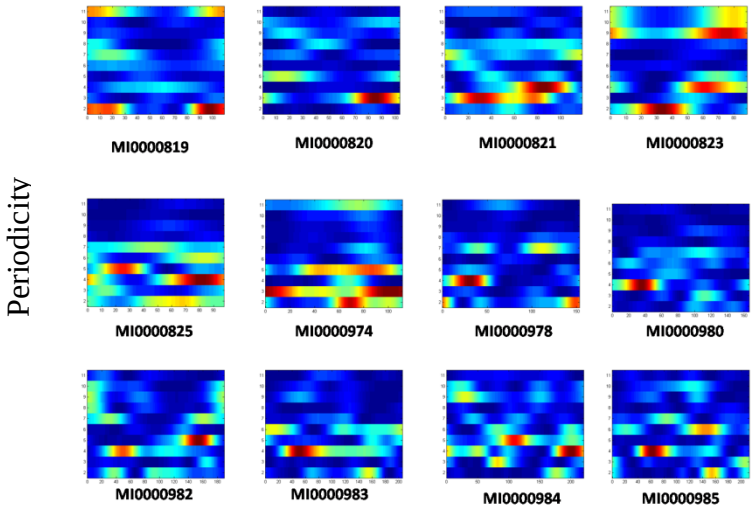


Figure 49: Spectrum corresponding to Periodicity 2-11 for sequences MI0000819 TO MI0000985

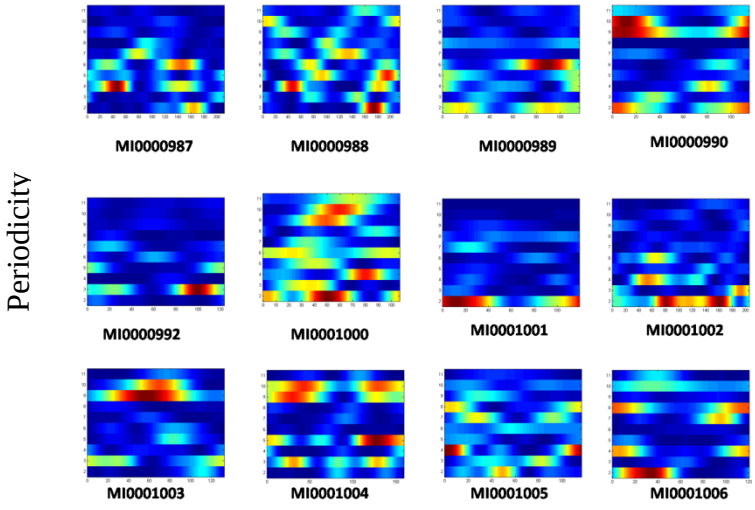


Figure 50: Spectrum corresponding to Periodicity 2-11 for sequences MI0000987 TO MI0001006

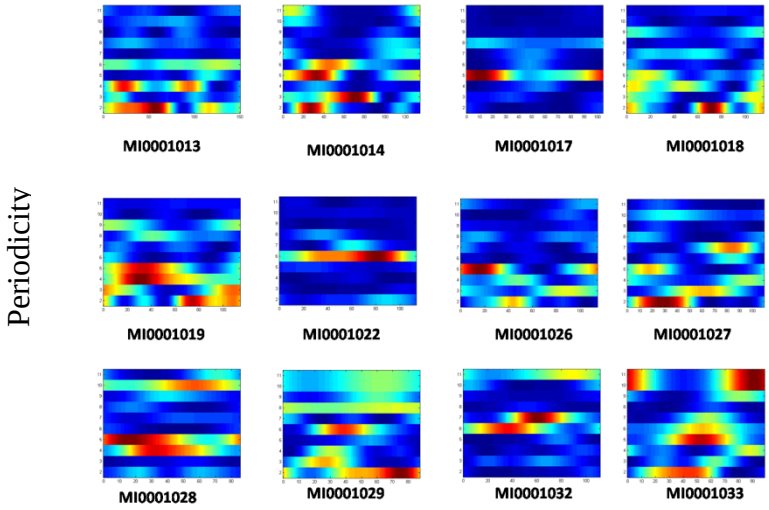


Figure 51: Spectrum corresponding to Periodicity 2-11 for sequences MI0001013 TO MI0001033

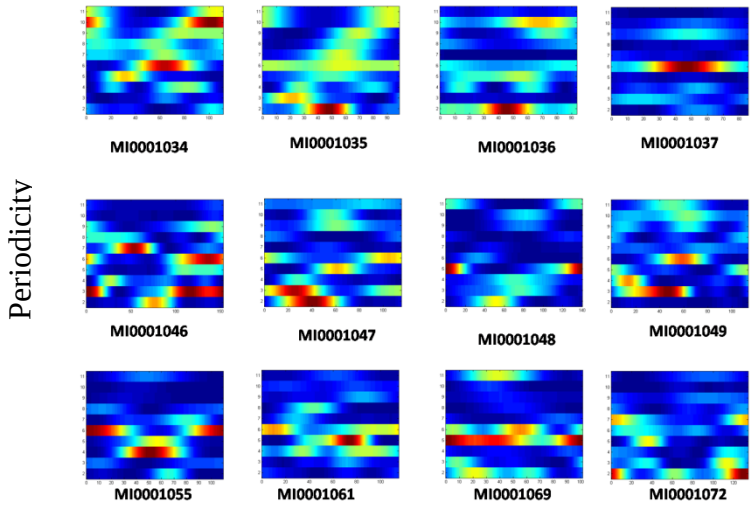


Figure 52: Spectrum corresponding to Periodicity 2-11 for sequences MI0001034 TO MI0001072

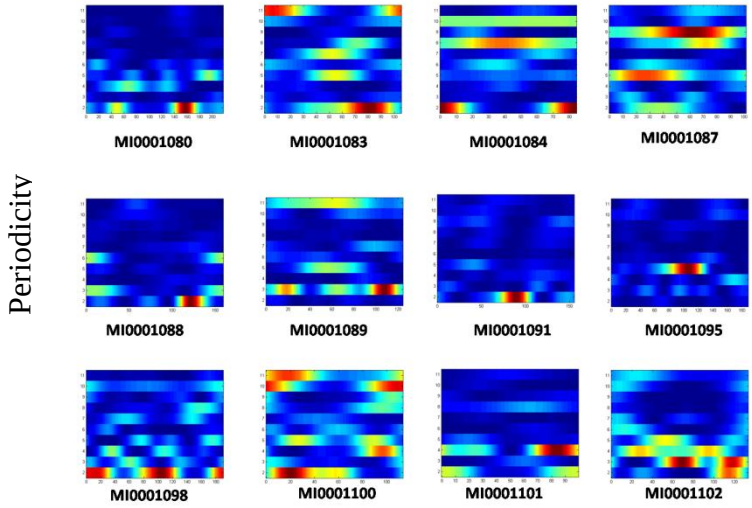


Figure 53: Spectrum corresponding to Periodicity 2-11 for sequences MI0001080 TO MI0001102

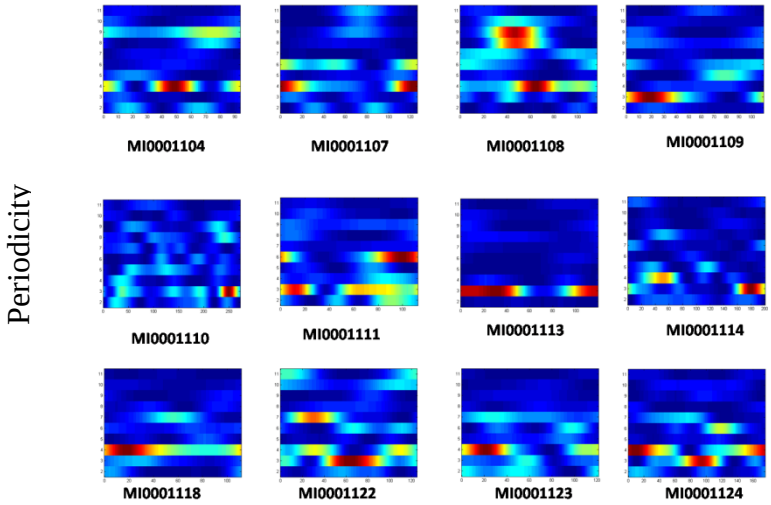


Figure 54: Spectrum corresponding to Periodicity 2-11 for sequences MI0001104 TO MI0001124

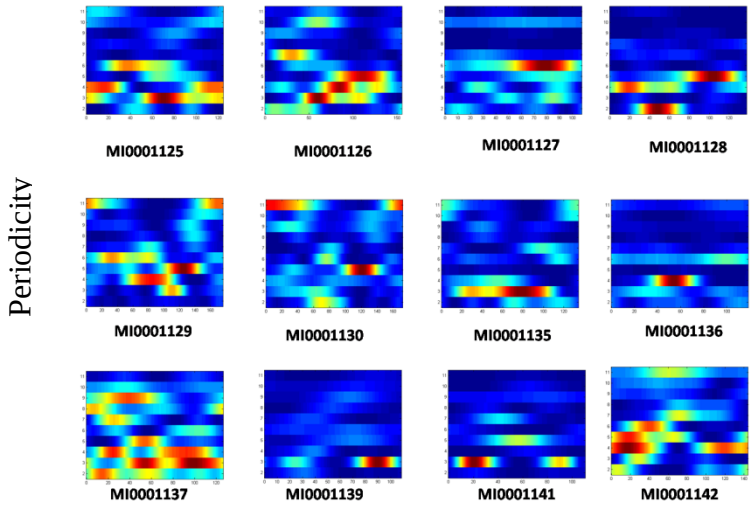


Figure 55: Spectrum corresponding to Periodicity 2-11 for sequences MI0001125TO MI0001142

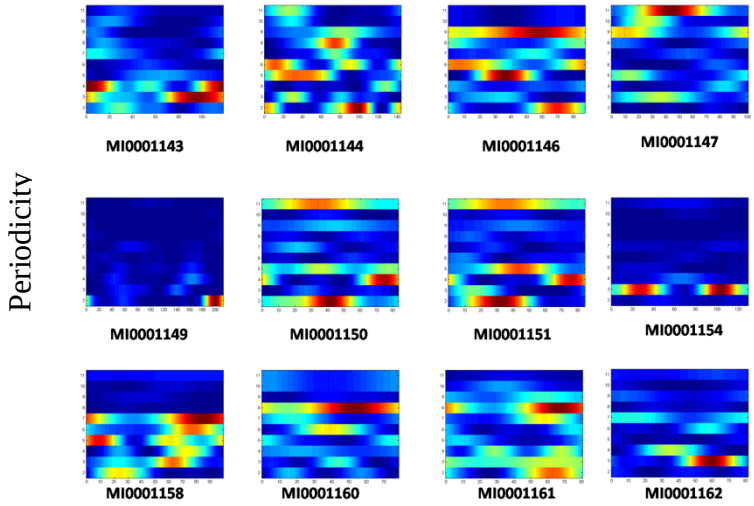


Figure 56: Spectrum corresponding to Periodicity 2-11 for sequences MI0001143 TO MI0001162

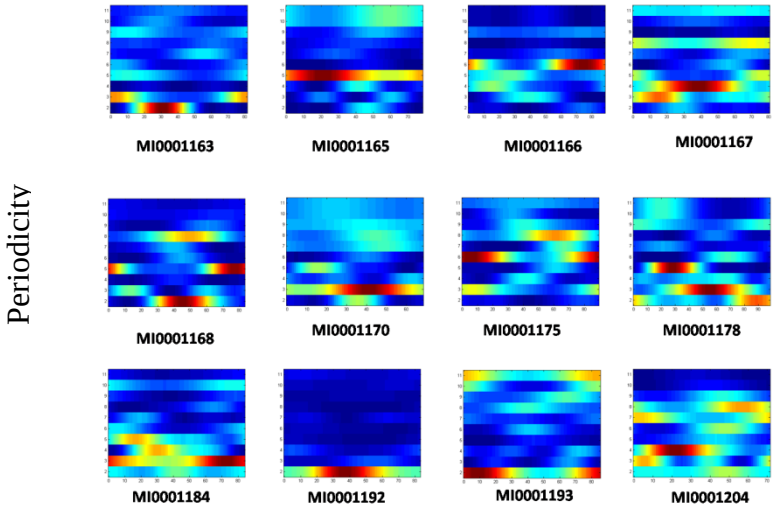


Figure 57: Spectrum corresponding to Periodicity 2-11 for sequences MI0001163 TO MI0001204

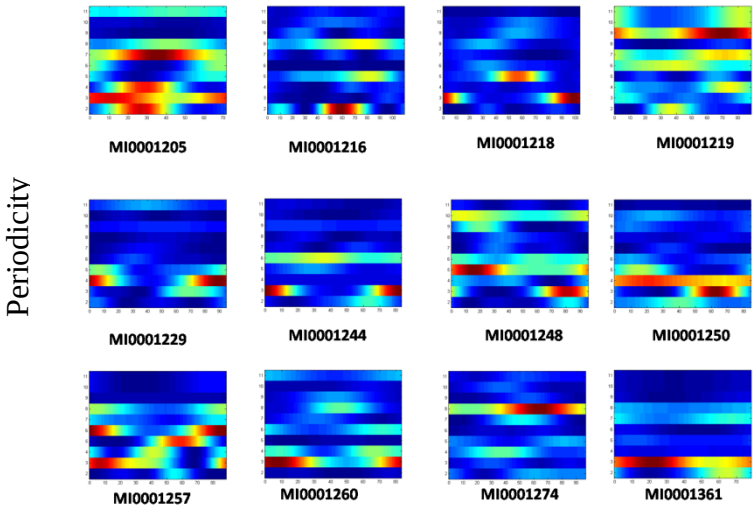


Figure 58: Spectrum corresponding to Periodicity 2-11 for sequences MI0001205 TO MI0001361

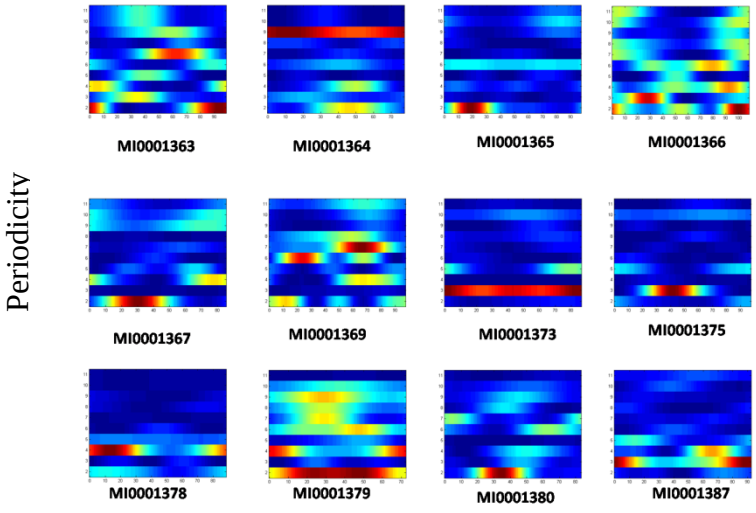


Figure 59: Spectrum corresponding to Periodicity 2-11 for sequences MI0001363 TO MI0001387

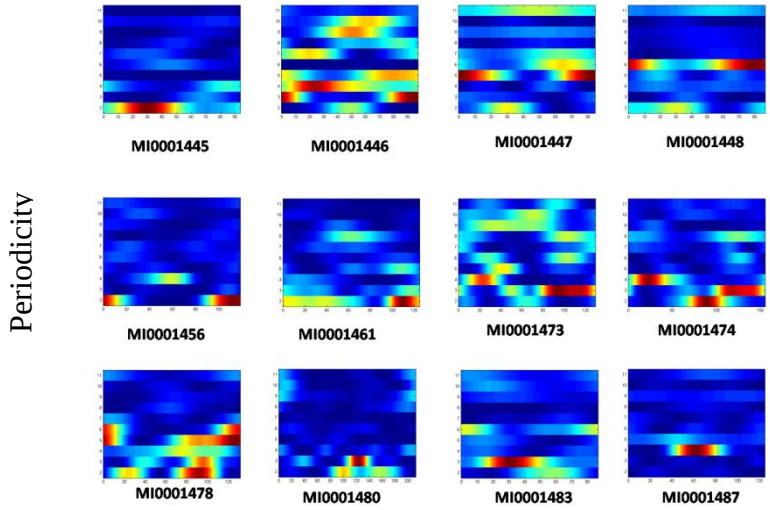


Figure 60: Spectrum corresponding to Periodicity 2-11 for sequences MI0001445 TO MI0001487

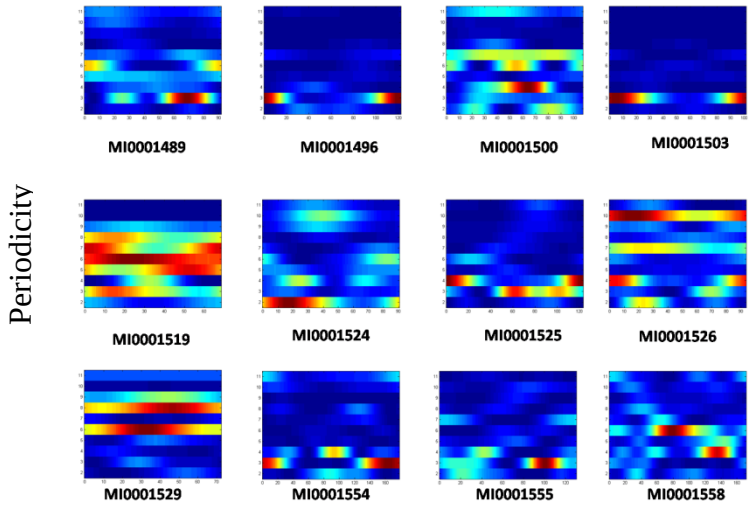


Figure 61: Spectrum corresponding to Periodicity 2-11 for sequences MI0001489 TO MI0001558

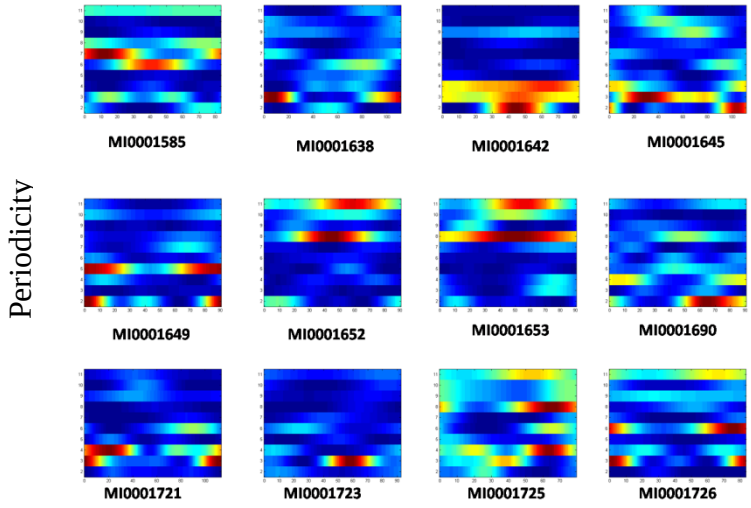


Figure 62: Spectrum corresponding to Periodicity 2-11 for sequences MI0001585 TO MI0001726

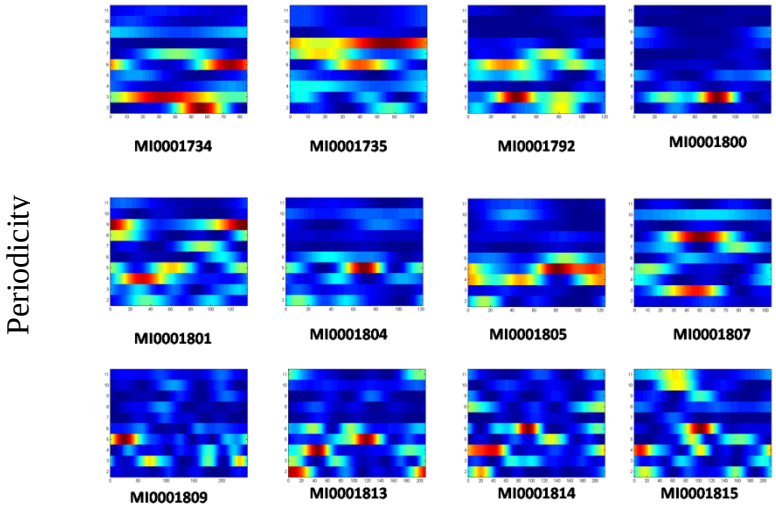


Figure 63: Spectrum corresponding to Periodicity 2-11 for sequences MI0001734 TO MI0001815

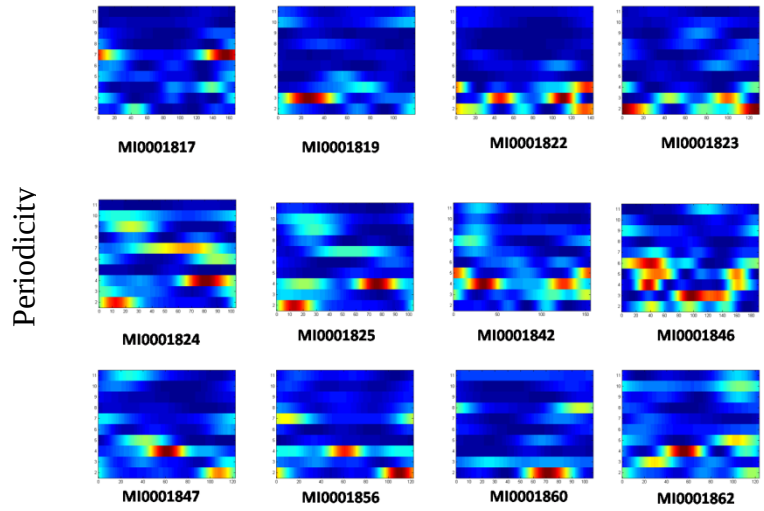


Figure 64: Spectrum corresponding to Periodicity 2-11 for sequences MI0001817 TO MI0001862

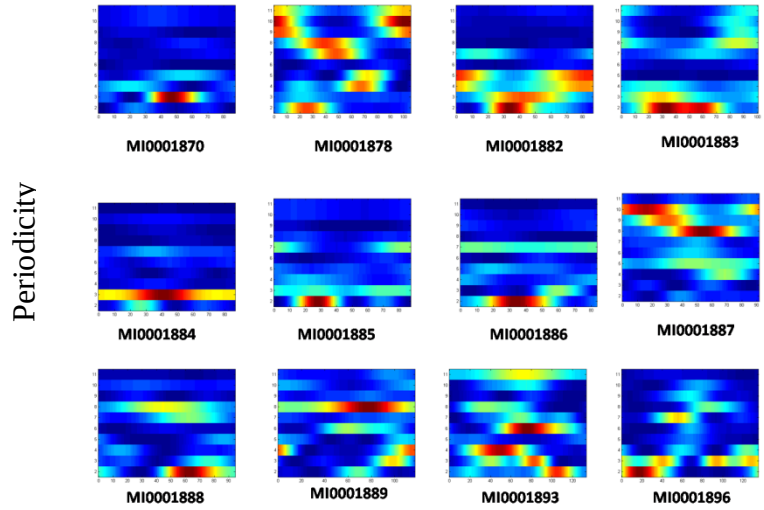


Figure 65: Spectrum corresponding to Periodicity 2-11 for sequences MI0001870 TO MI0001896

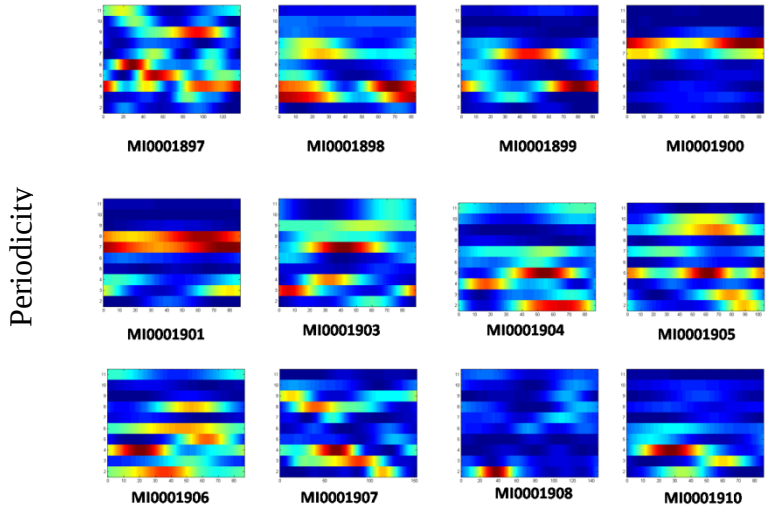


Figure 66: Spectrum corresponding to Periodicity 2-11 for sequences MI0001897 TO MI0001910

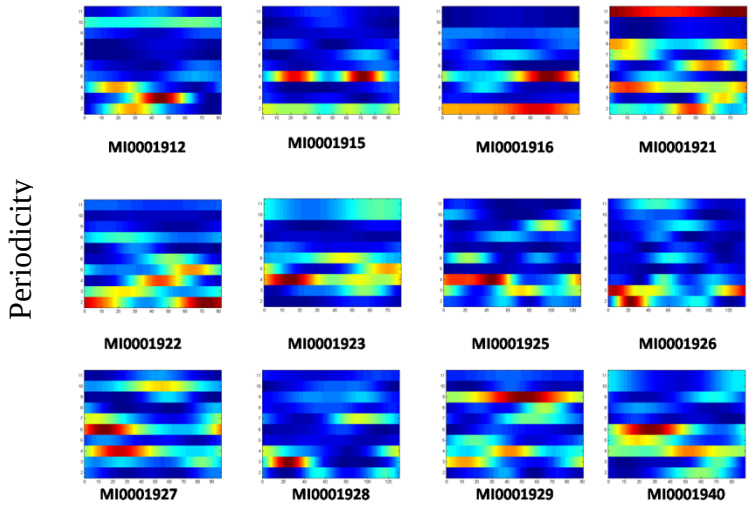


Figure 67: Spectrum corresponding to Periodicity 2-11 for sequences MI0001912 TO MI0001940

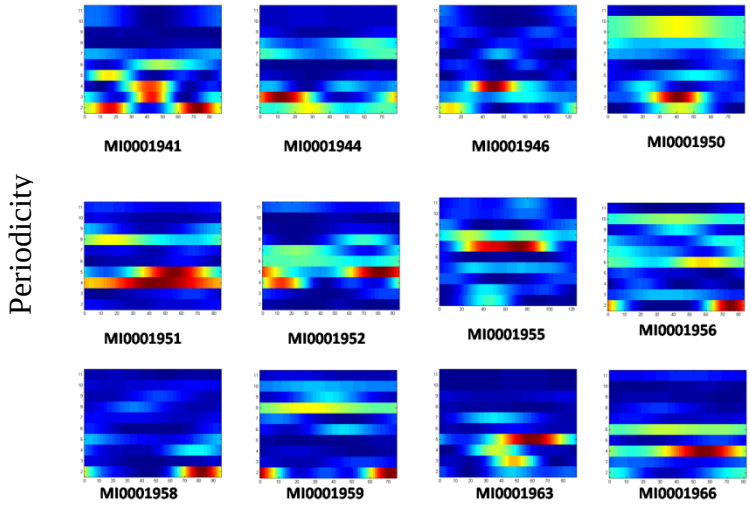


Figure 68: Spectrum corresponding to Periodicity 2-11 for sequences MI0001941 TO MI0001966

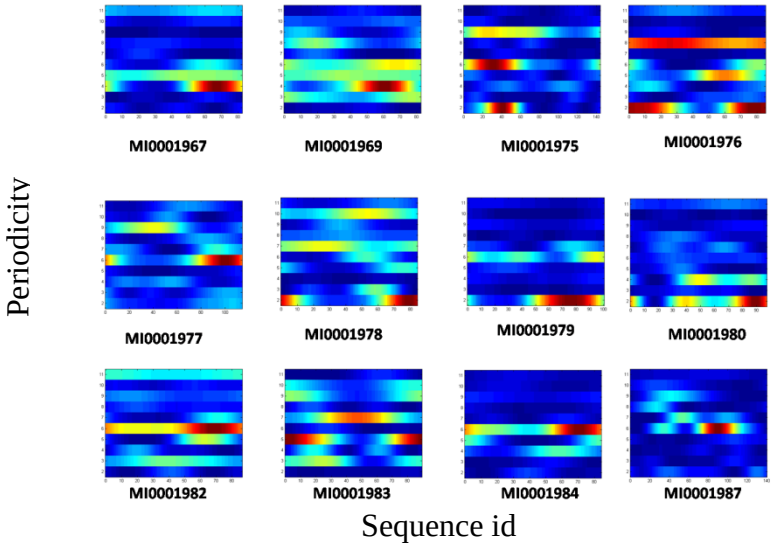


Figure 69: Spectrum corresponding to Periodicity 2-11 for sequences MI0001967 TO MI0001987

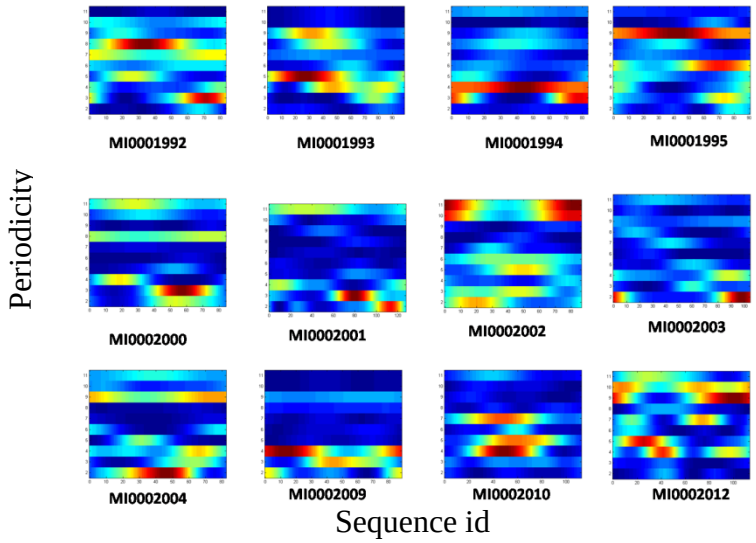


Figure 70: Spectrum corresponding to Periodicity 2-11 for sequences MI0001992 TO MI0002012

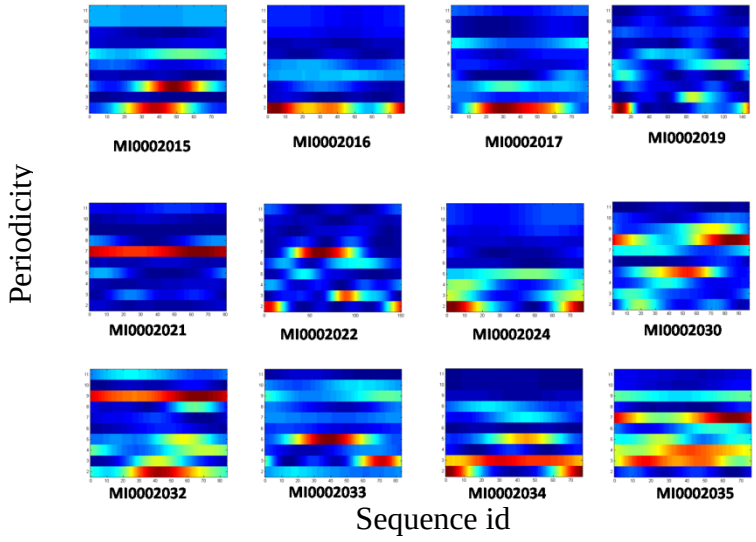


Figure 71: Spectrum corresponding to Periodicity 2-11 for sequences MI0002015 TO MI0002035

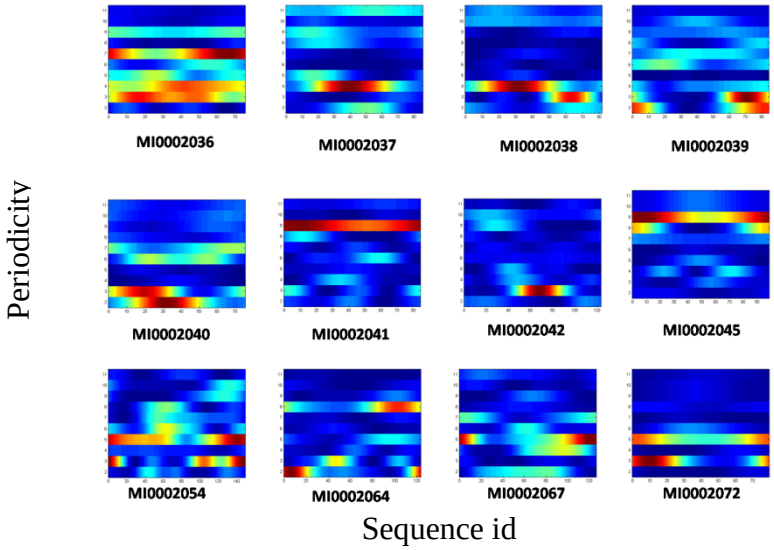


Figure 72: Spectrum corresponding to Periodicity 2-11 for sequences MI0002036 TO MI0002072

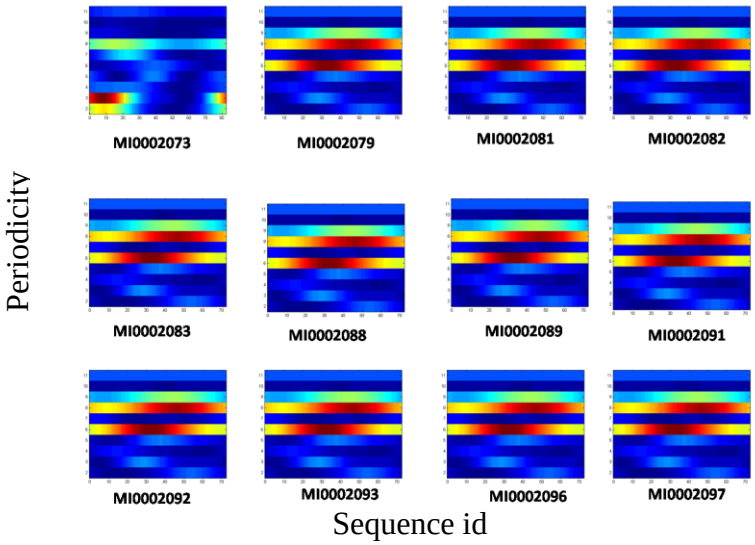


Figure 73: Spectrum corresponding to Periodicity 2-11 for sequences MI0002073 TO MI0002097

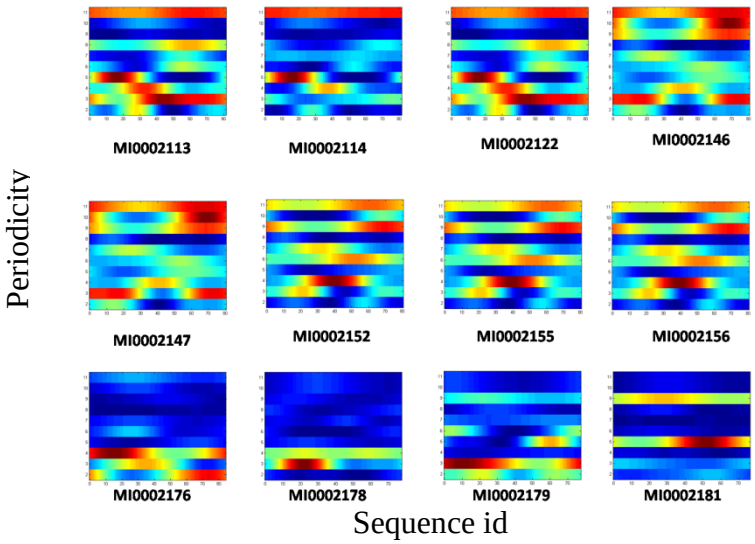


Figure 74: Spectrum corresponding to Periodicity 2-11 for sequences MI0002113 TO MI0002181

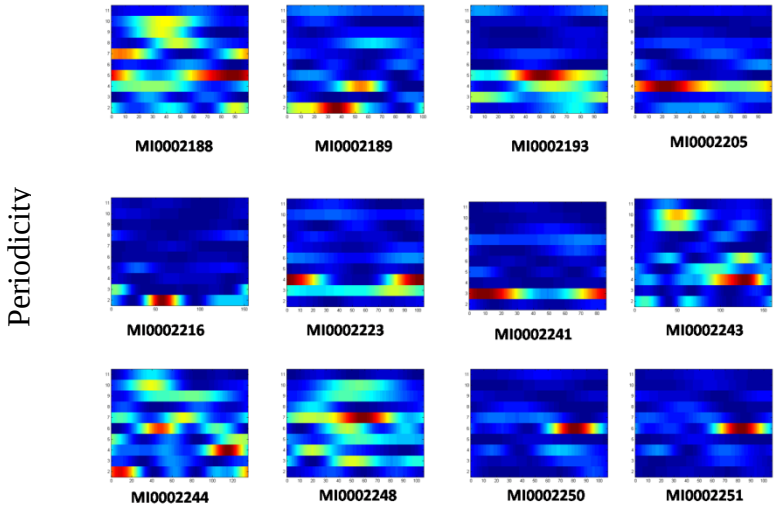


Figure 75: Spectrum corresponding to Periodicity 2-11 for sequences MI0002188 TO MI0002251

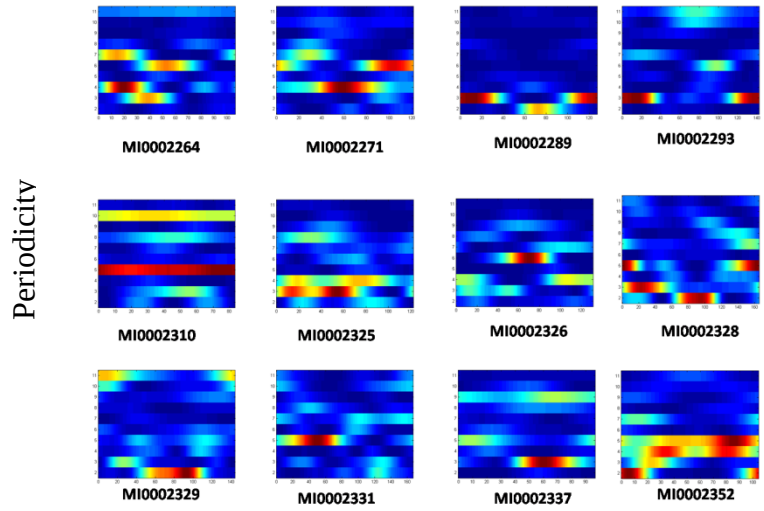


Figure 76: Spectrum corresponding to Periodicity 2-11 for sequences MI0002264 TO MI0002352

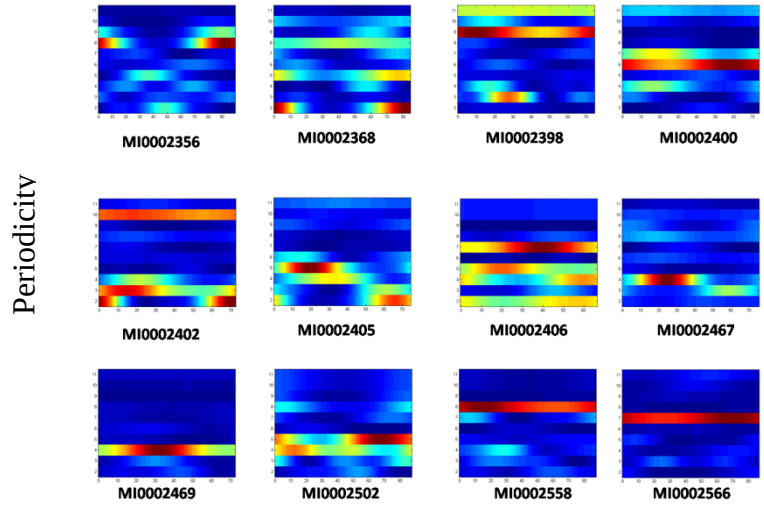


Figure 77: Spectrum corresponding to Periodicity 2-11 for sequences MI0002356 TO MI0002566

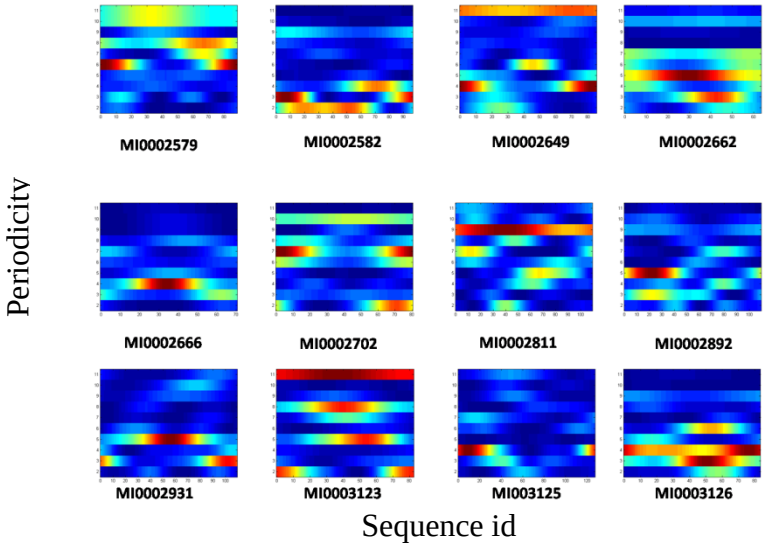


Figure 78: Spectrum corresponding to Periodicity 2-11 for sequences MI0002579 TO MI0003126

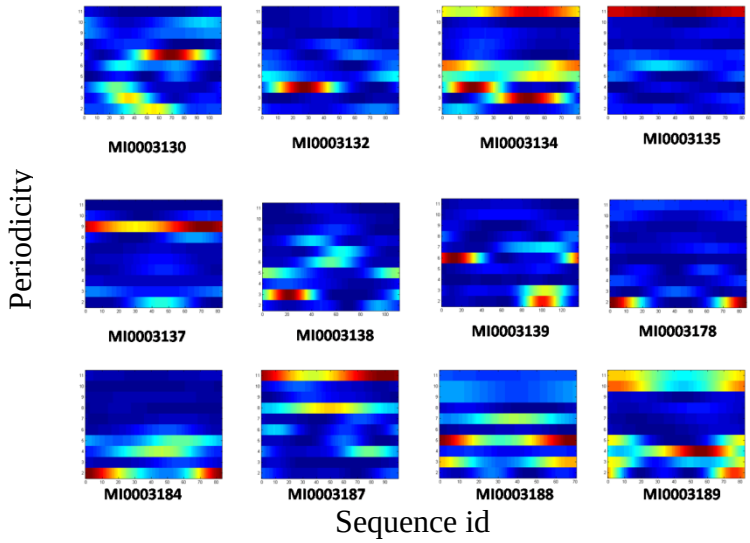


Figure 79: Spectrum corresponding to Periodicity 2-11 for sequences MI0003130 TO MI0003189

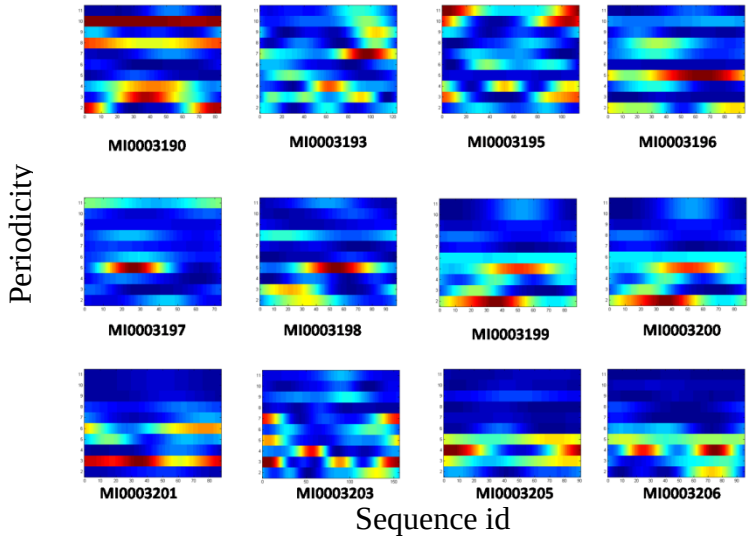


Figure 80: Spectrum corresponding to Periodicity 2-11 for sequences MI0003190 TO MI0003206

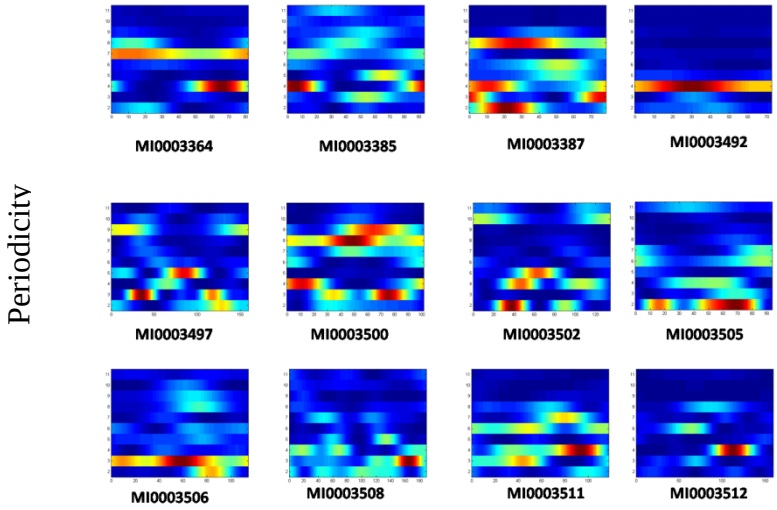


Figure 81: Spectrum corresponding to Periodicity 2-11 for sequences MI0003364 TO MI0003512

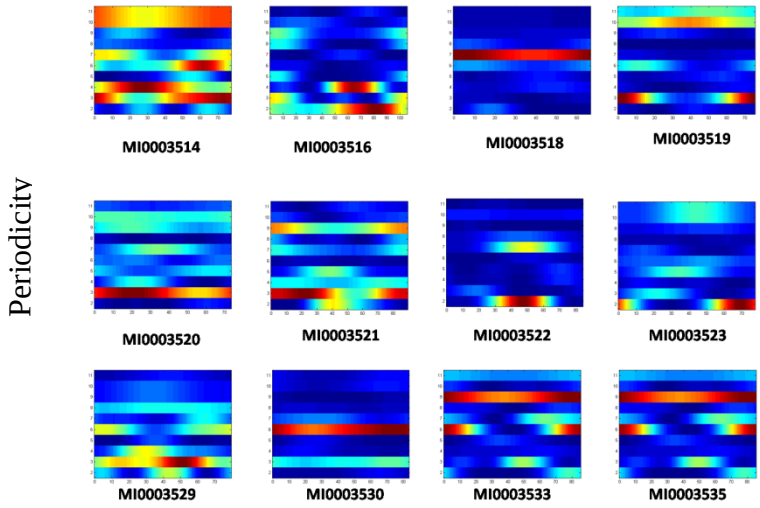


Figure 82: Spectrum corresponding to Periodicity 2-11 for sequences MI0003514 TO MI0003535

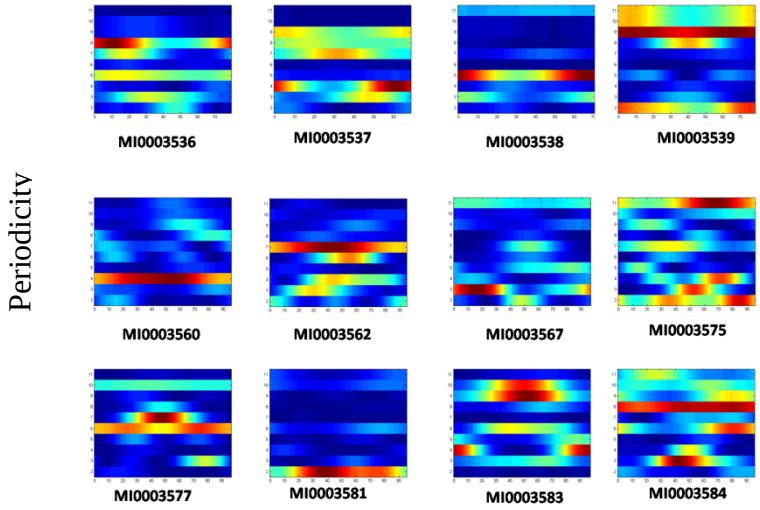


Figure 83: Spectrum corresponding to Periodicity 2-11 for sequences MI0003536 TO MI0003584

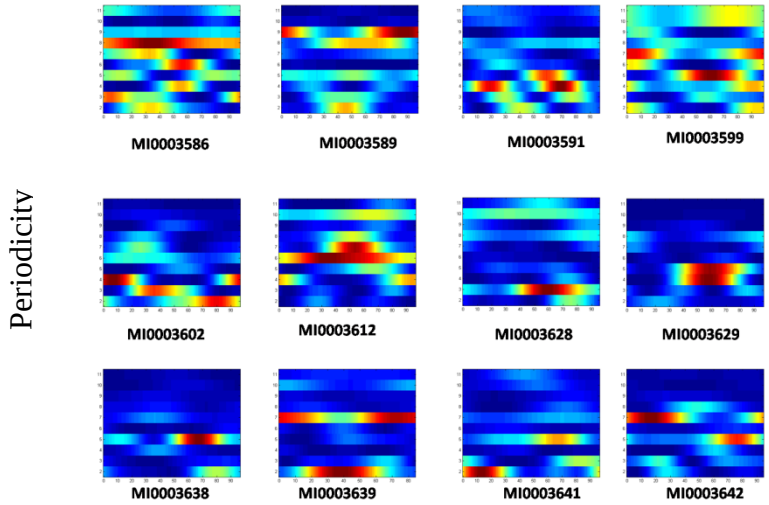


Figure 84: Spectrum corresponding to Periodicity 2-11 for sequences MI0003586 TO MI0003642

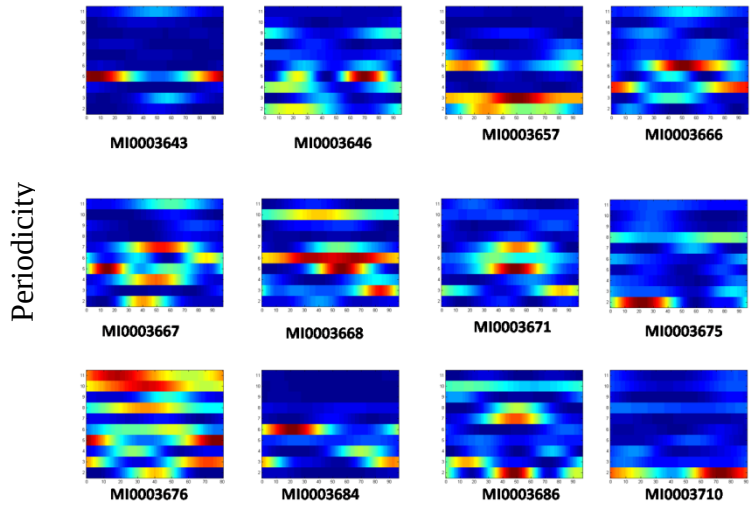


Figure 85: Spectrum corresponding to Periodicity 2-11 for sequences MI0003643 TO MI0003710

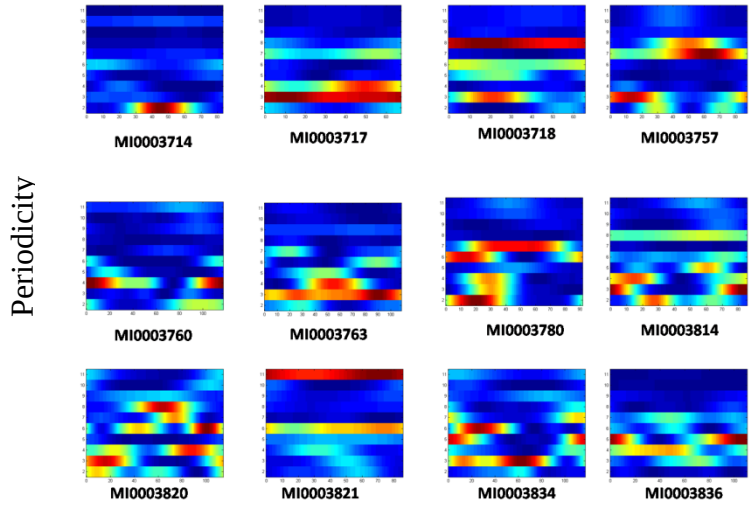


Figure 86: Spectrum corresponding to Periodicity 2-11 for sequences MI0003714 TO MI0003836

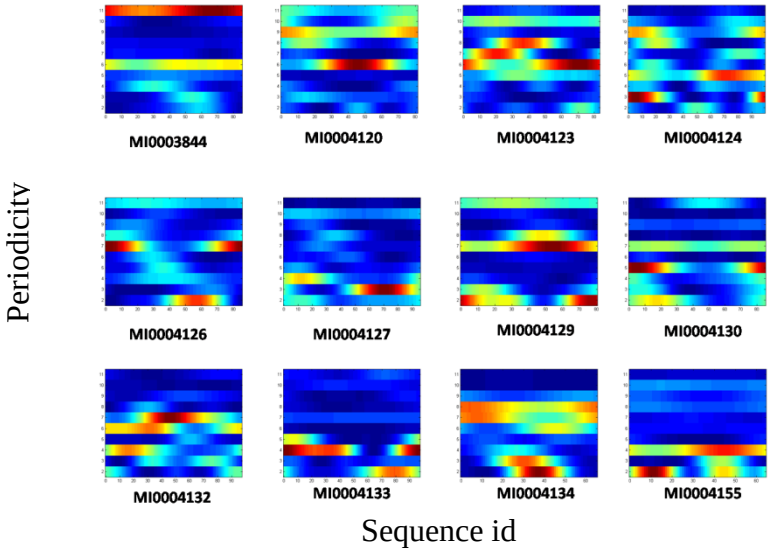


Figure 87: Spectrum corresponding to Periodicity 2-11 for sequences MI0003844 TO MI0004155

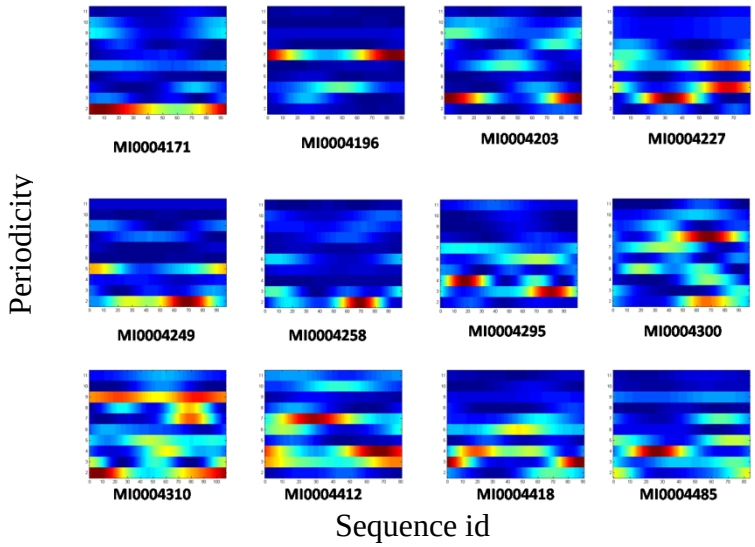


Figure 88: Spectrum corresponding to Periodicity 2-11 for sequences MI0004171 TO MI0004485

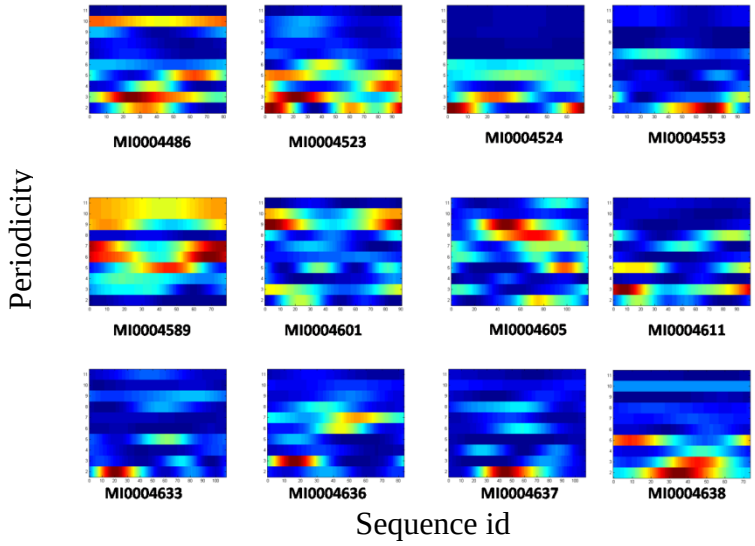


Figure 89: Spectrum corresponding to Periodicity 2-11 for sequences MI0004486 TO MI0004638

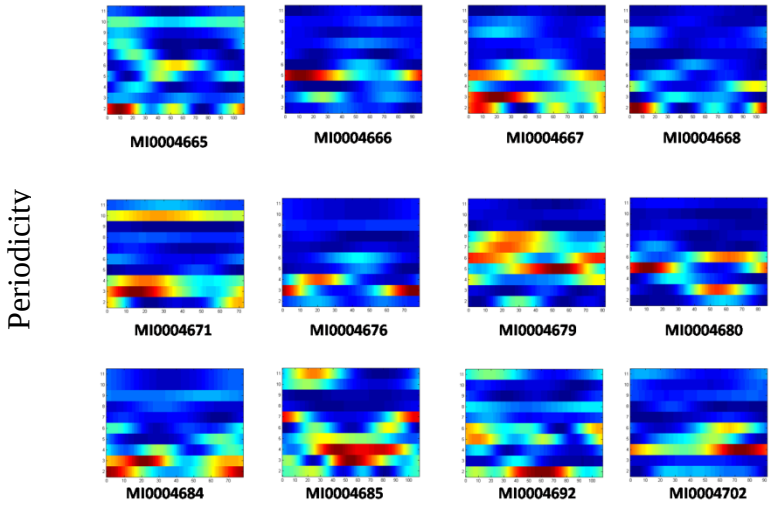


Figure 90: Spectrum corresponding to Periodicity 2-11 for sequences MI0004665 TO MI0004702

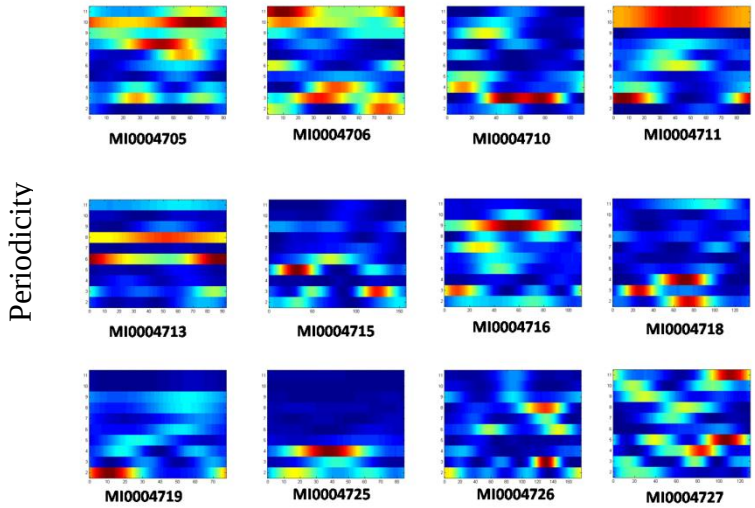


Figure 91: Spectrum corresponding to Periodicity 2-11 for sequences MI0004705 TO MI0004727

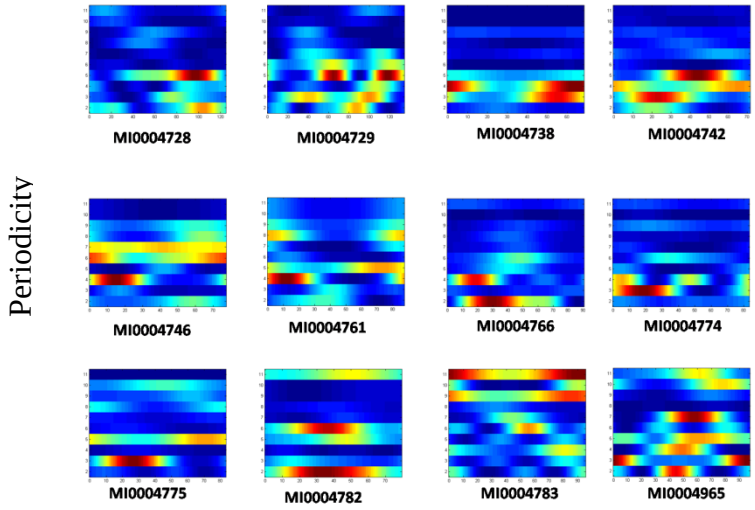


Figure 92: Spectrum corresponding to Periodicity 2-11 for sequences MI0004728 TO MI0004965

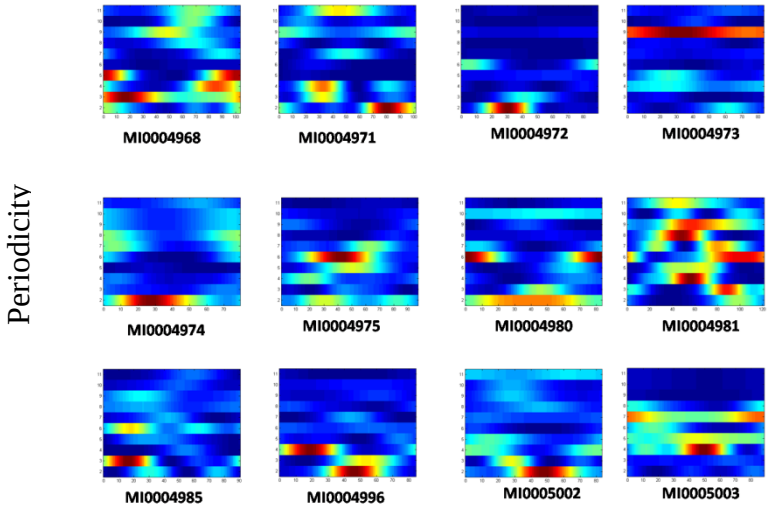


Figure 93: Spectrum corresponding to Periodicity 2-11 for sequences MI0004968 TO MI0005003

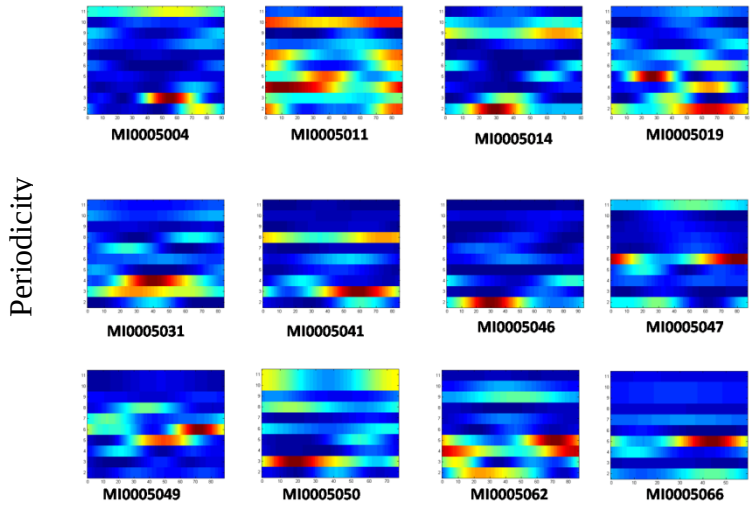


Figure 94: Spectrum corresponding to Periodicity 2-11 for sequences MI0005004 TO MI0005066

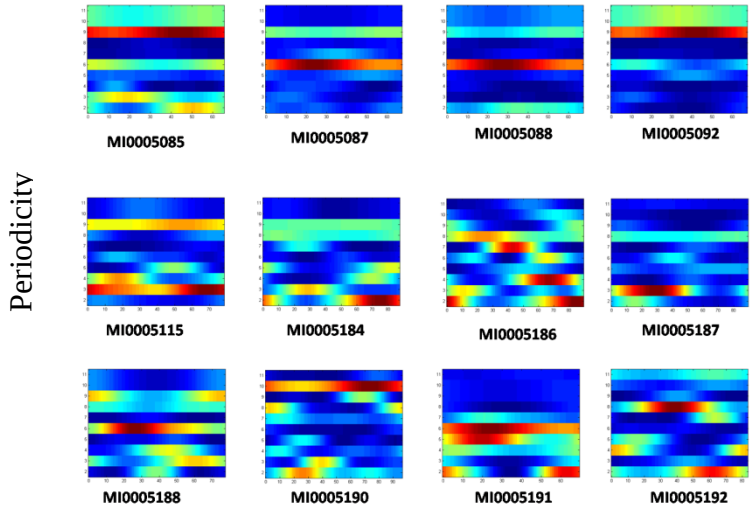


Figure 95: Spectrum corresponding to Periodicity 2-11 for sequences MI0005085 TO MI0005192

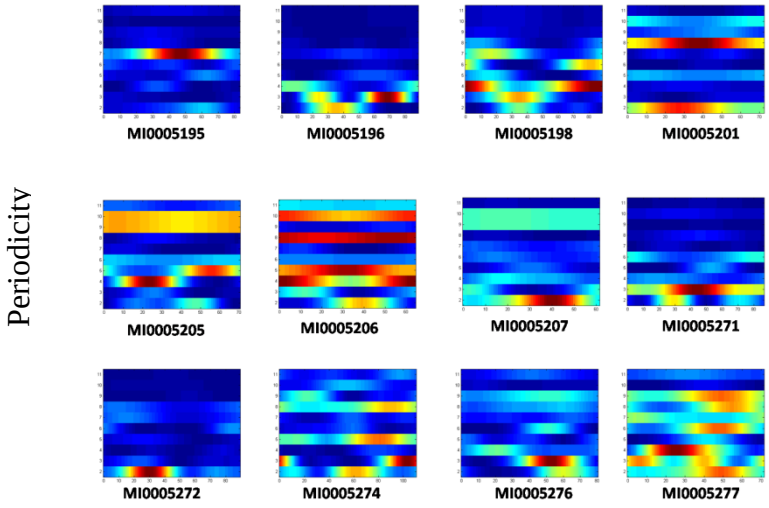


Figure 96: Spectrum corresponding to Periodicity 2-11 for sequences MI0005195 TO MI0005277

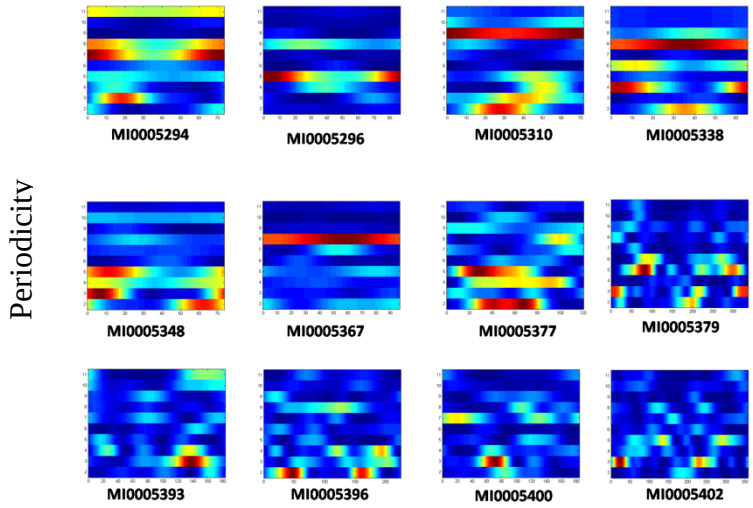


Figure 97: Spectrum corresponding to Periodicity 2-11 for sequences MI0005294 TO MI0005402

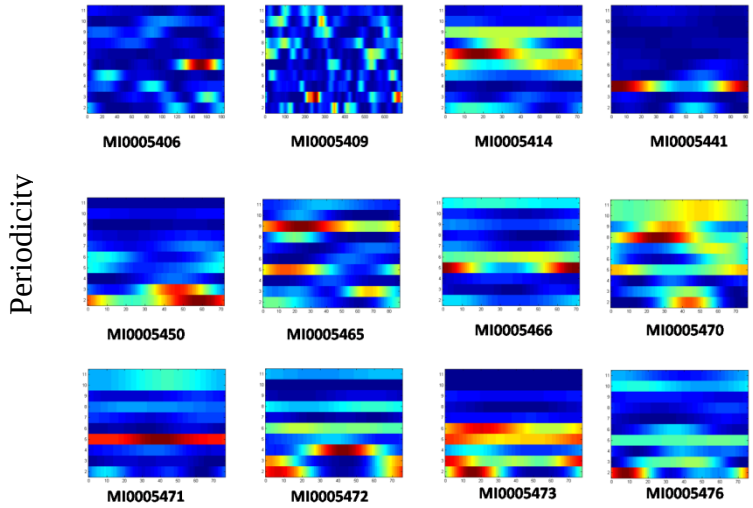


Figure 98: Spectrum corresponding to Periodicity 2-11 for sequences MI0005406 TO MI0005476

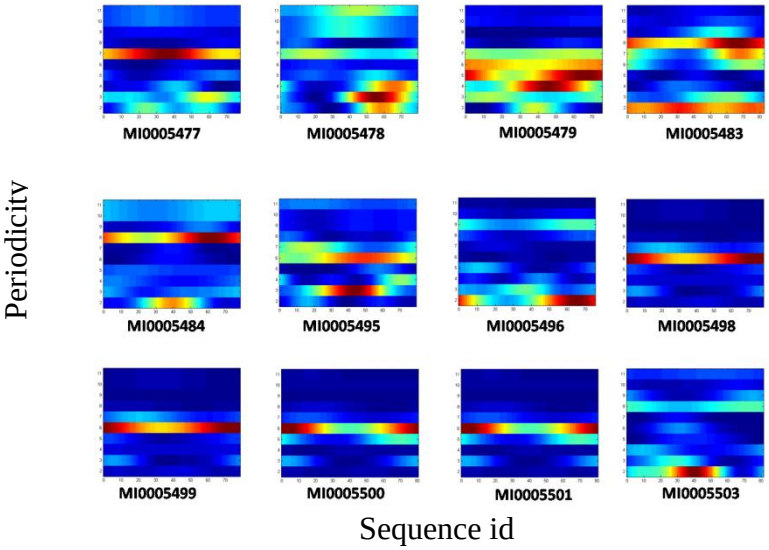


Figure 99: Spectrum corresponding to Periodicity 2-11 for sequences MI0005477 TO MI0005503

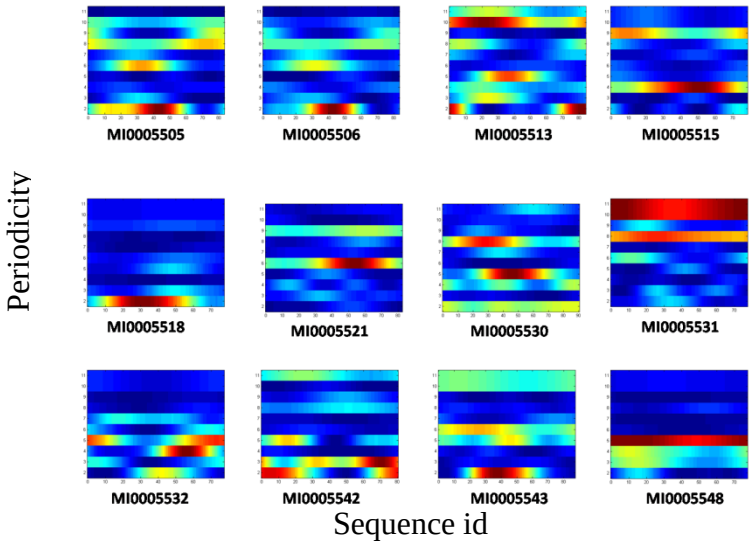


Figure 100: Spectrum corresponding to Periodicity 2-11 for sequences MI0005505 TO MI0005548

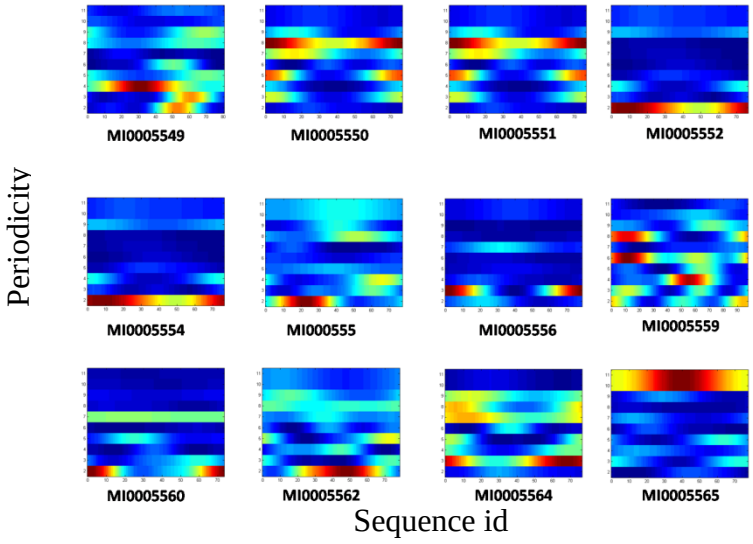


Figure 101: Spectrum corresponding to Periodicity 2-11 for sequences MI0005549 TO MI0005565

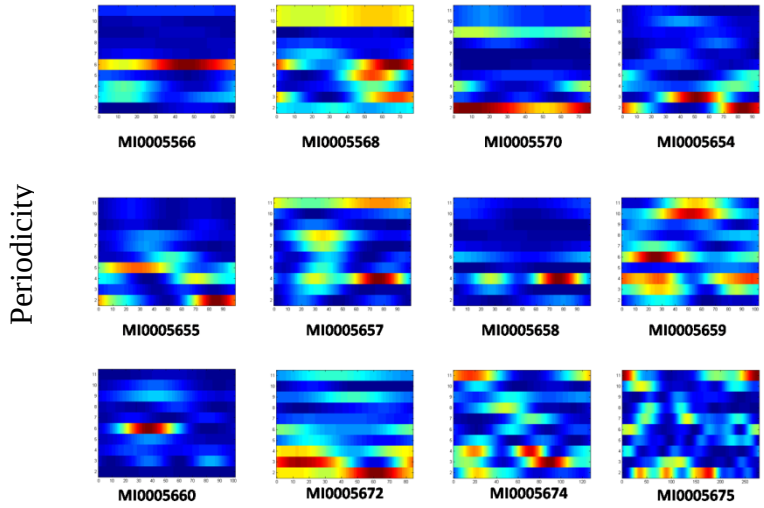


Figure 102: Spectrum corresponding to Periodicity 2-11 for sequences MI0005566 TO MI0005675

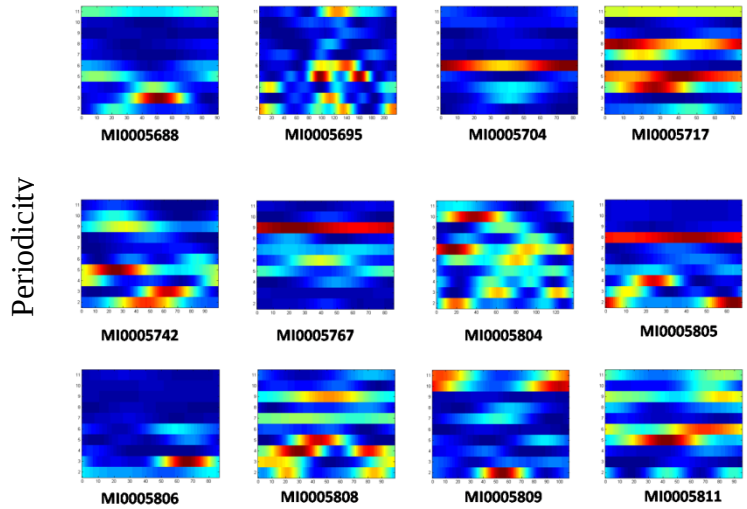


Figure 103: Spectrum corresponding to Periodicity 2-11 for sequences MI0005688 TO MI0005811

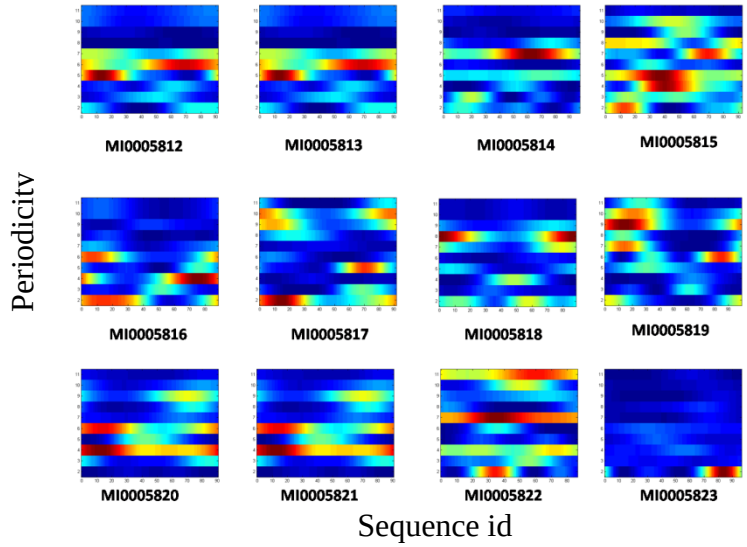


Figure 104: Spectrum corresponding to Periodicity 2-11 for sequences MI0005812 TO MI0005823

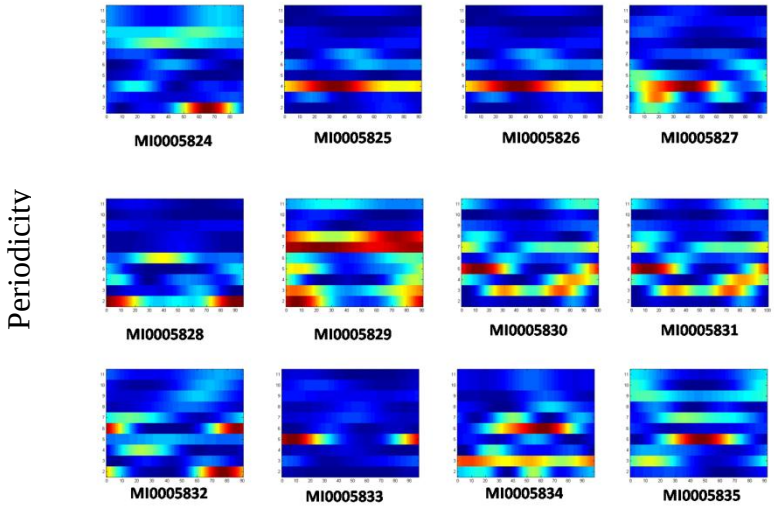


Figure 105: Spectrum corresponding to Periodicity 2-11 for sequences MI0005824 TO MI0005835

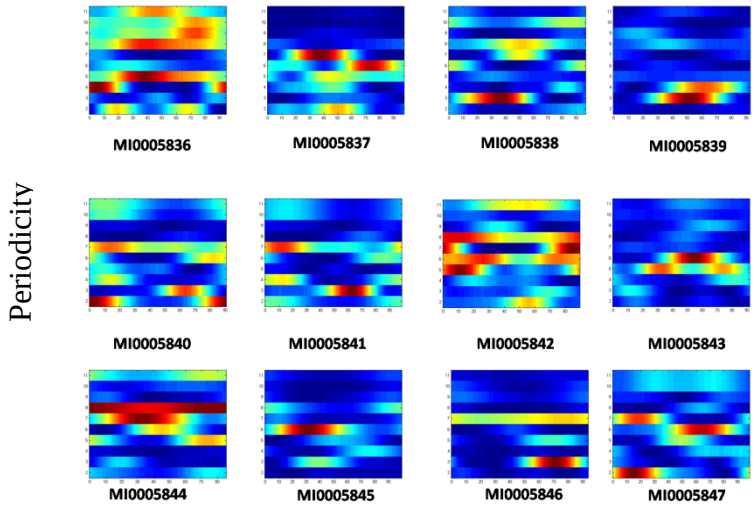


Figure 106: Spectrum corresponding to Periodicity 2-11 for sequences MI0005836 TO MI0005847

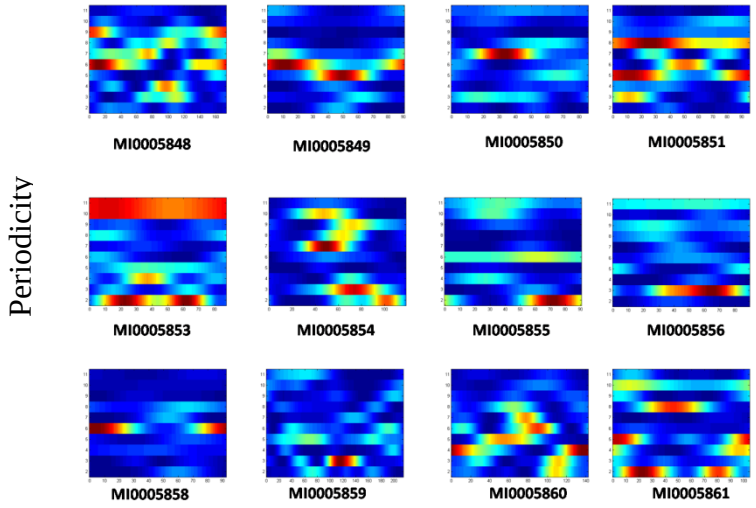


Figure 107: Spectrum corresponding to Periodicity 2-11 for sequences MI0005848 TO MI0005861

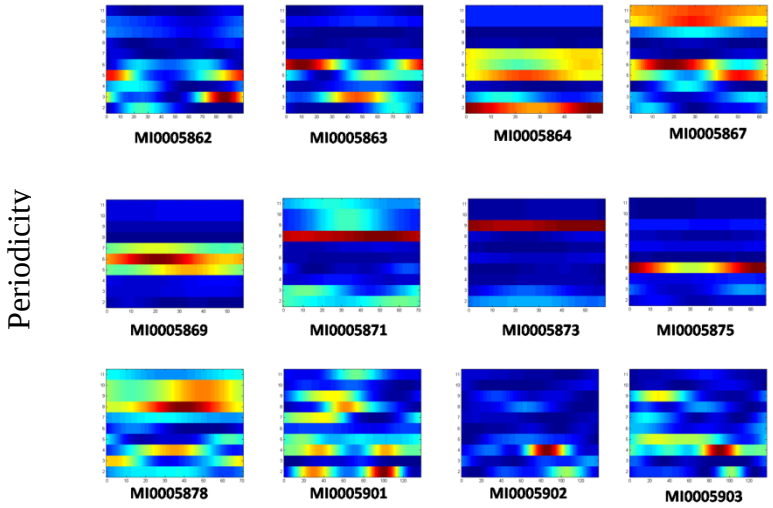


Figure 108: Spectrum corresponding to Periodicity 2-11 for sequences MI0005862 TO MI0005903

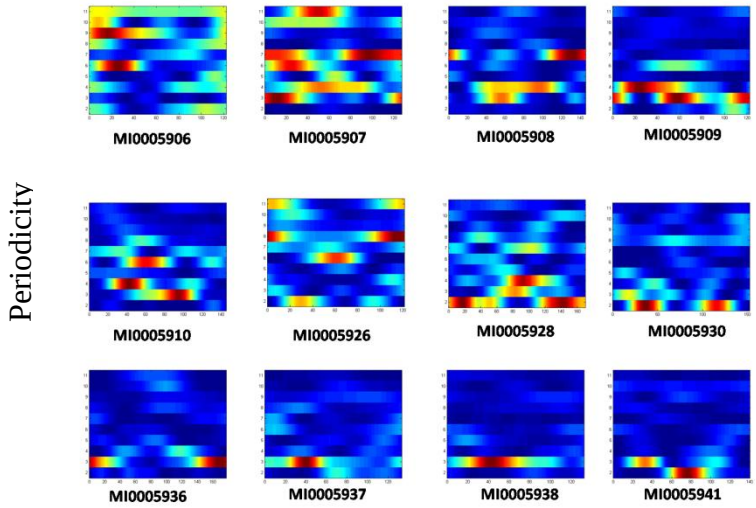


Figure 109: Spectrum corresponding to Periodicity 2-11 for sequences MI0005906 TO MI0005941

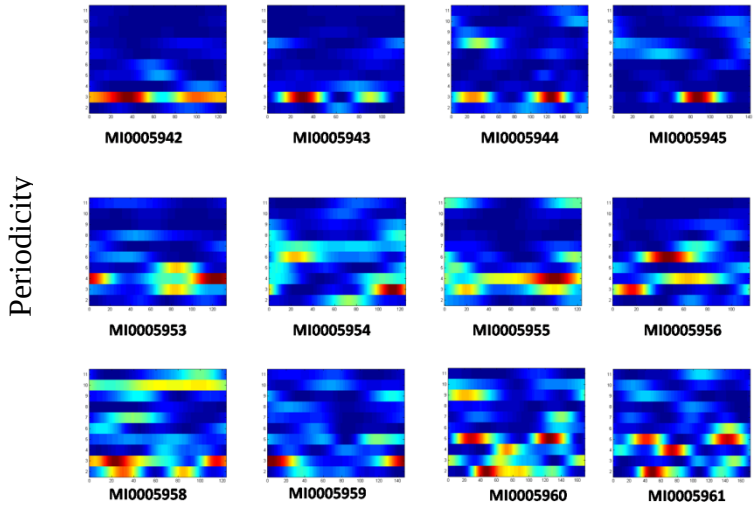


Figure 110: Spectrum corresponding to Periodicity 2-11 for sequences MI0005942 TO MI0005961

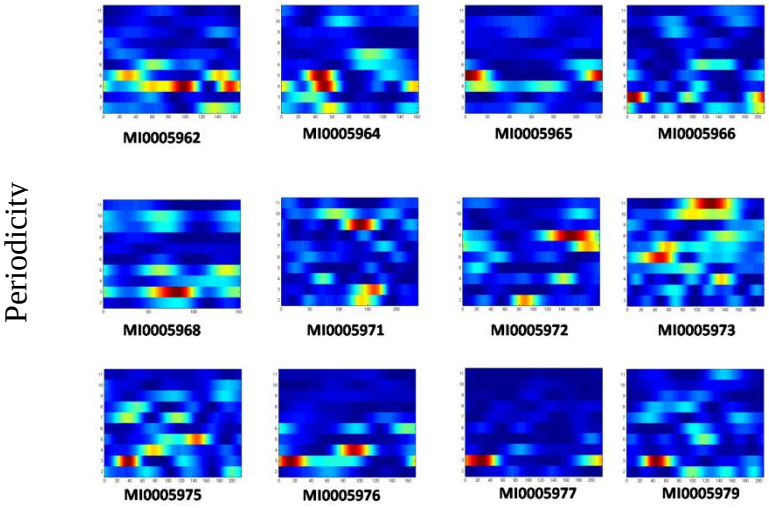


Figure 111: Spectrum corresponding to Periodicity 2-11 for sequences MI0005962 TO MI0005979

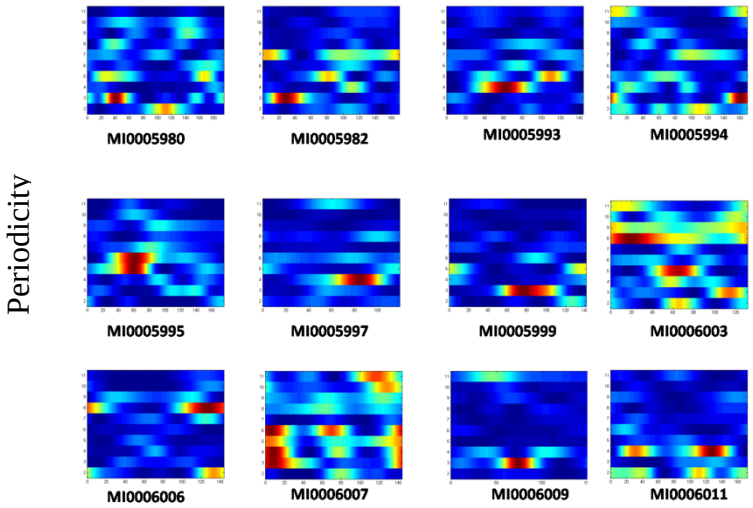


Figure 112: Spectrum corresponding to Periodicity 2-11 for sequences MI0005980 TO MI0006011

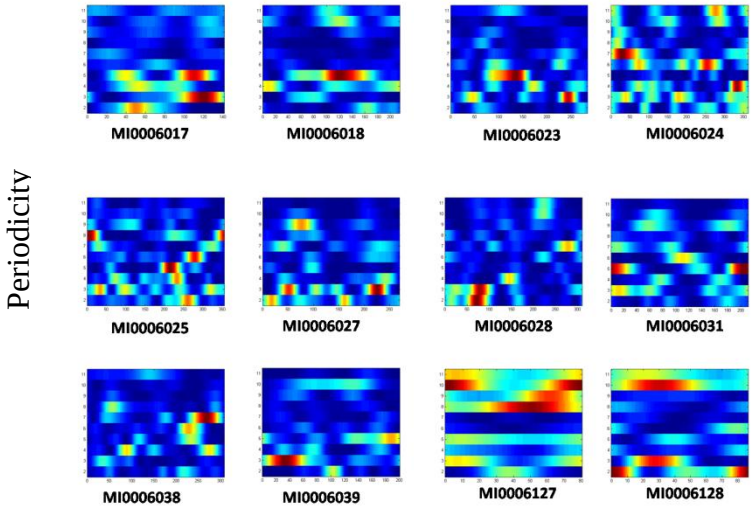


Figure 113: Spectrum corresponding to Periodicity 2-11 for sequences MI0006017 TO MI0006128

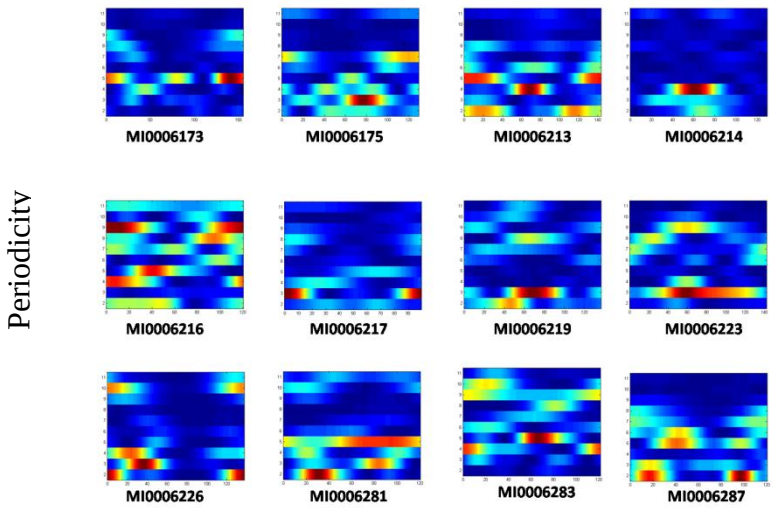


Figure 114: Spectrum corresponding to Periodicity 2-11 for sequences MI0006173 TO MI0006287

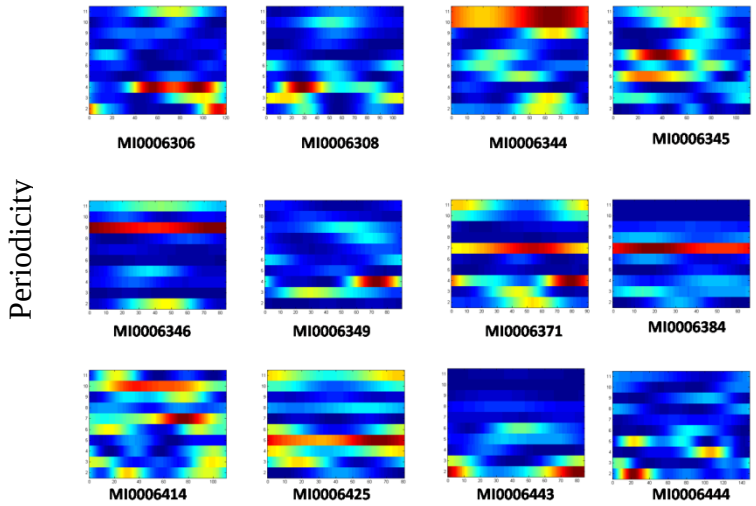


Figure 115: Spectrum corresponding to Periodicity 2-11 for sequences MI0006306 TO MI0006444

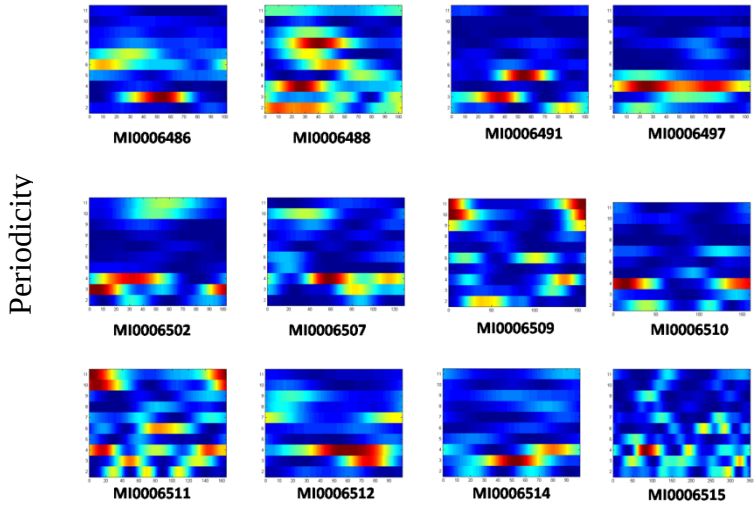


Figure 116: Spectrum corresponding to Periodicity 2-11 for sequences MI0006486 TO MI0006515

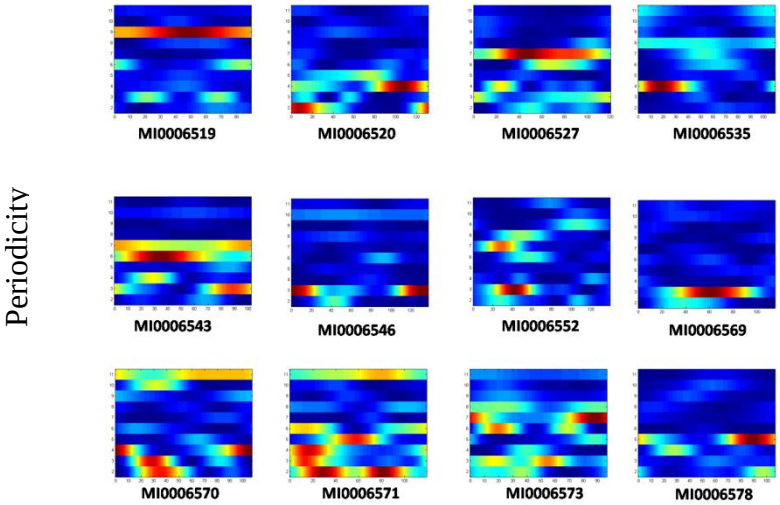


Figure 117: Spectrum corresponding to Periodicity 2-11 for sequences MI0006519 TO MI0006578

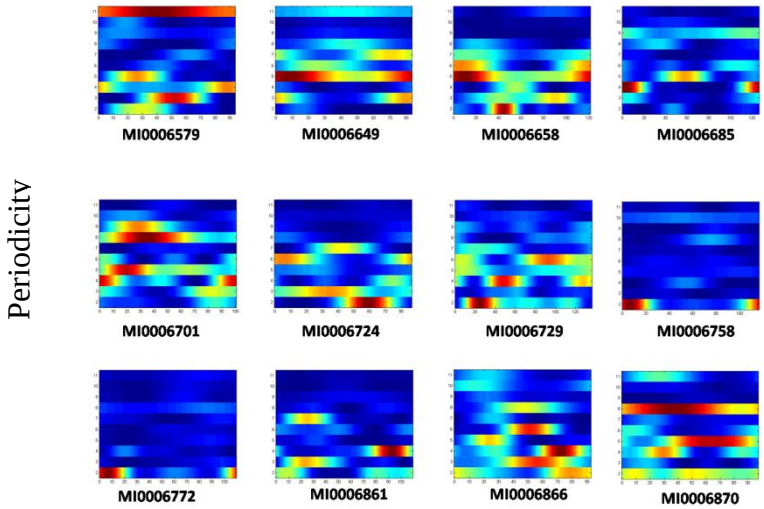


Figure 118: Spectrum corresponding to Periodicity 2-11 for sequences MI0006579 TO MI0006870

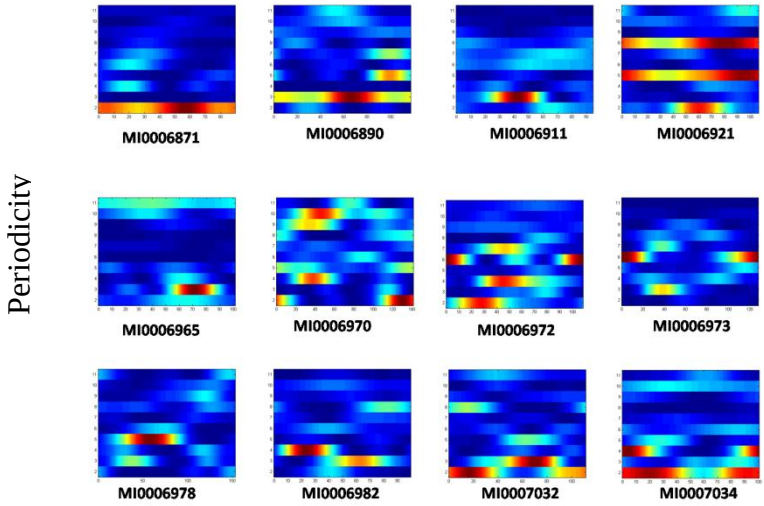


Figure 119: Spectrum corresponding to Periodicity 2-11 for sequences MI0006871 TO MI0007034

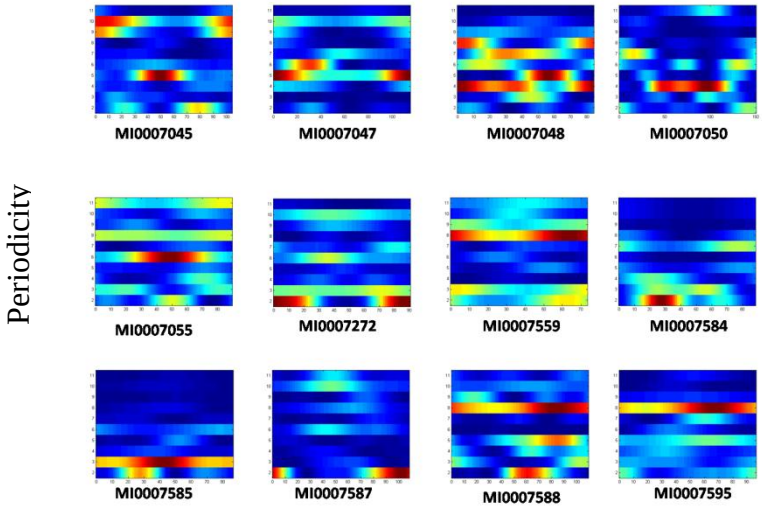


Figure 120: Spectrum corresponding to Periodicity 2-11 for sequences MI0007045 TO MI0007595

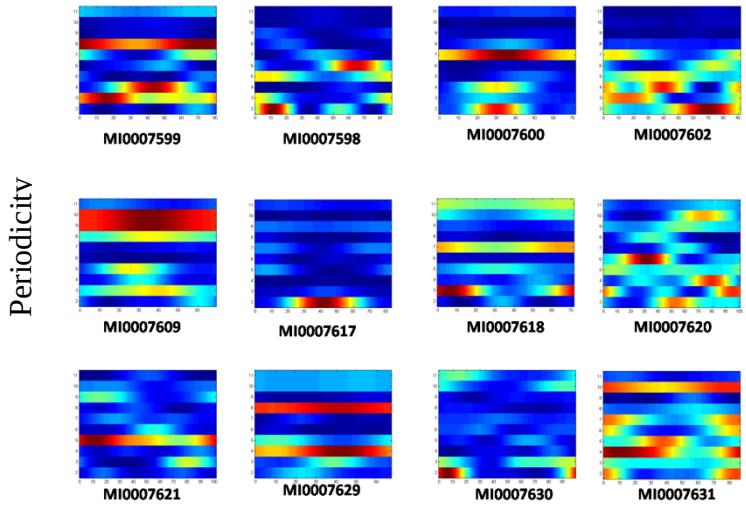


Figure 121: Spectrum corresponding to Periodicity 2-11 for sequences MI0007599 TO MI0007631

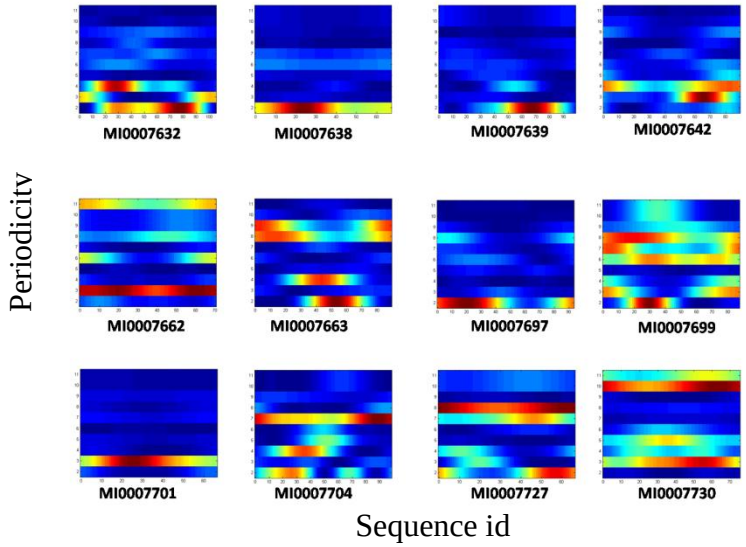


Figure 122: Spectrum corresponding to Periodicity 2-11 for sequences MI0007632 TO MI0007730

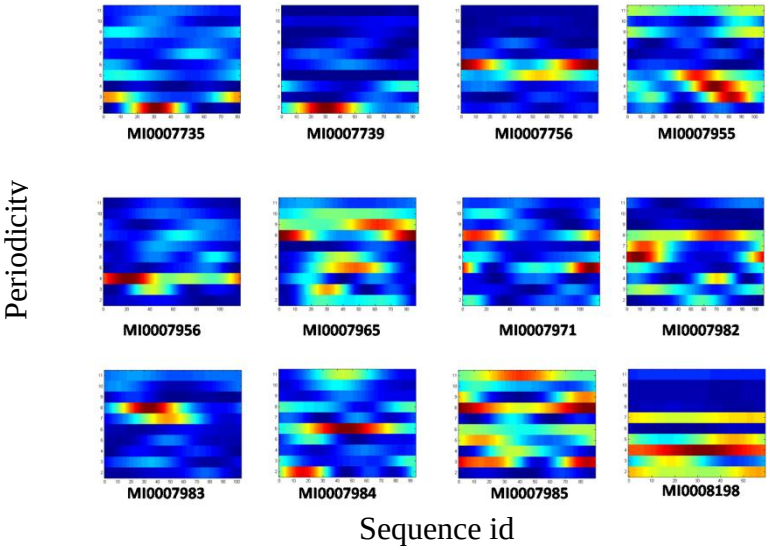


Figure 123: Spectrum corresponding to Periodicity 2-11 for sequences MI0007735 TO MI0008198

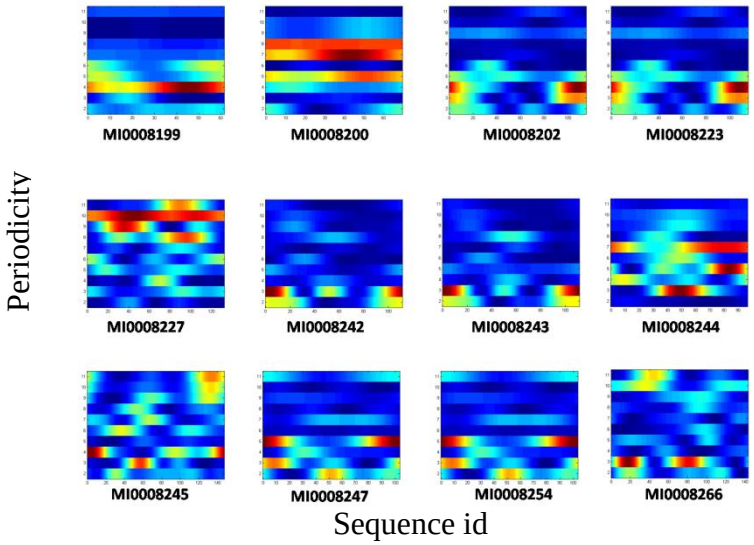


Figure 124: Spectrum corresponding to Periodicity 2-11 for sequences MI0008199 TO MI0008266

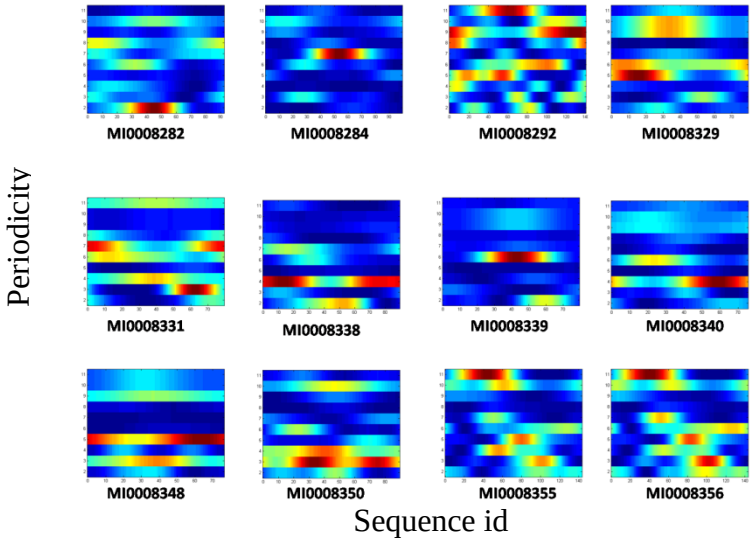


Figure 125: Spectrum corresponding to Periodicity 2-11 for sequences MI0008282 TO MI0008356

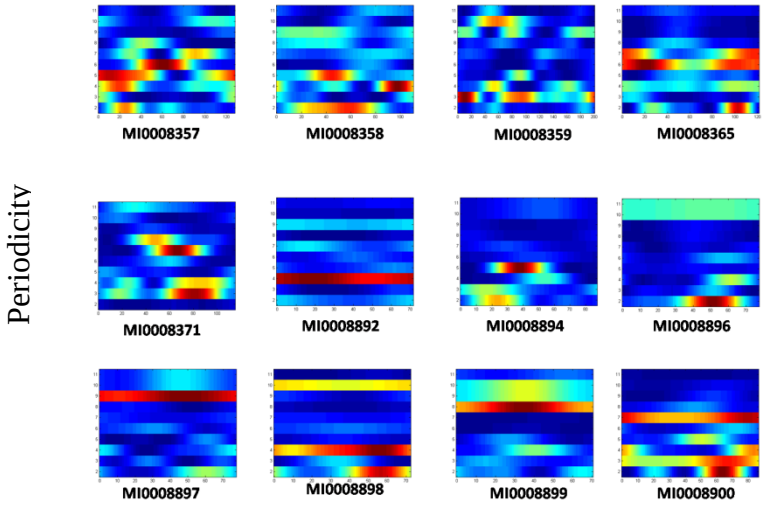


Figure 126: Spectrum corresponding to Periodicity 2-11 for sequences MI0008357 TO MI0008900

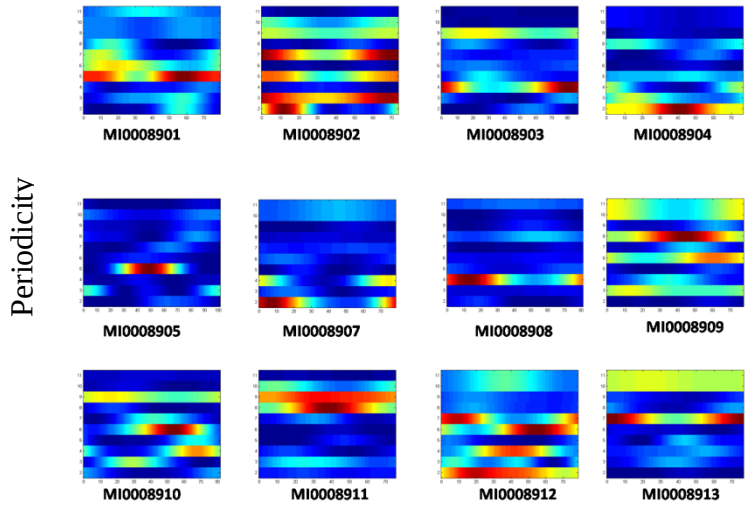


Figure 127: Spectrum corresponding to Periodicity 2-11 for sequences MI0008901 TO MI0008913

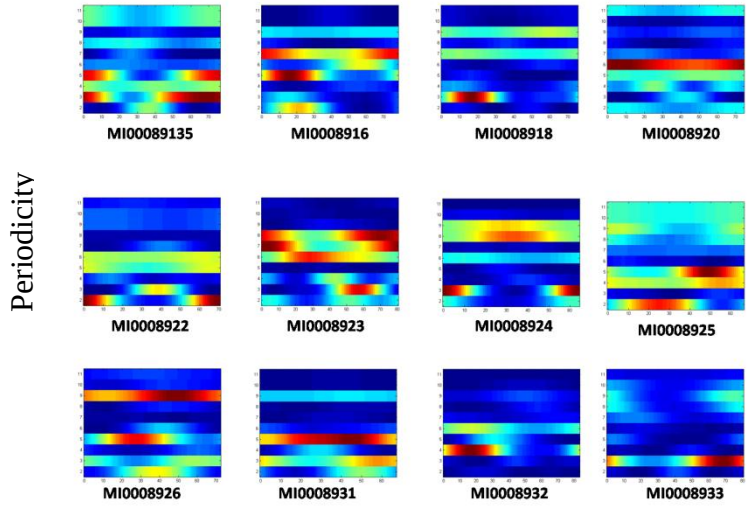


Figure 128: Spectrum corresponding to Periodicity 2-11 for sequences MI0008915 TO MI0008933

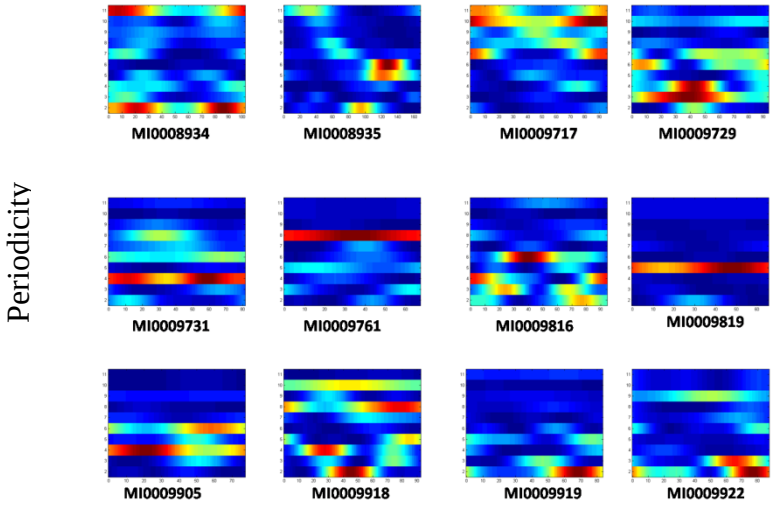


Figure 129: Spectrum corresponding to Periodicity 2-11 for sequences MI0008934 TO MI0009922

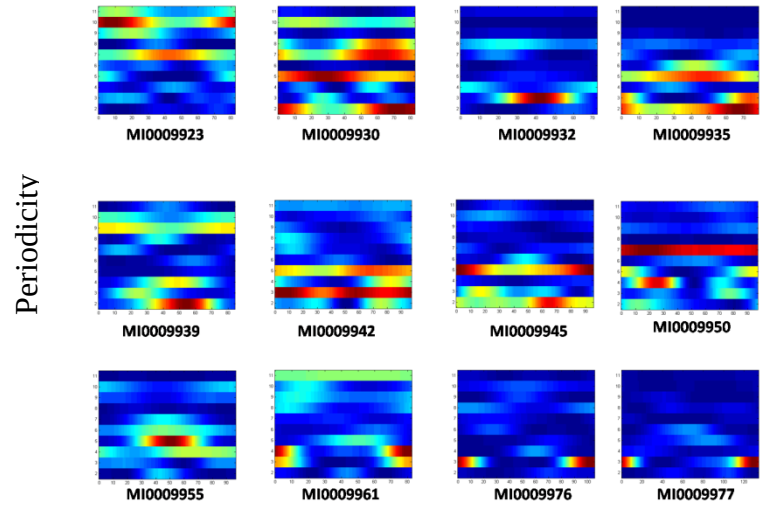


Figure 130: Spectrum corresponding to Periodicity 2-11 for sequences MI0009923 TO MI0009977

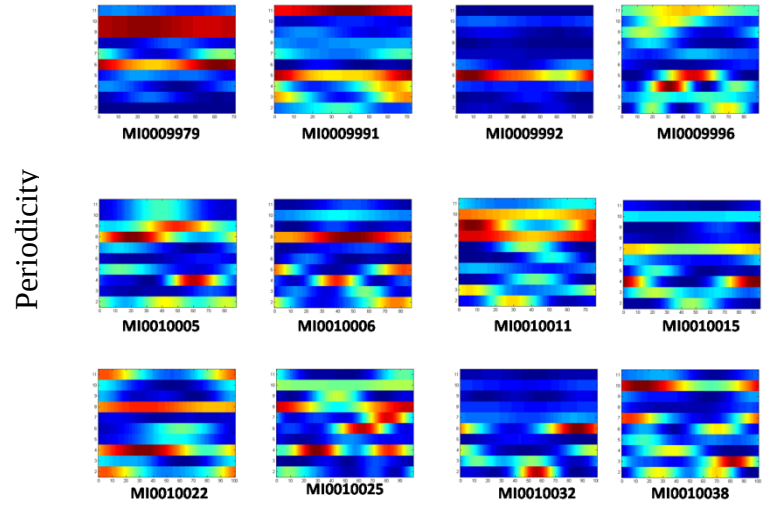


Figure 131: Spectrum corresponding to Periodicity 2-11 for sequences MI0009979 TO MI0010038

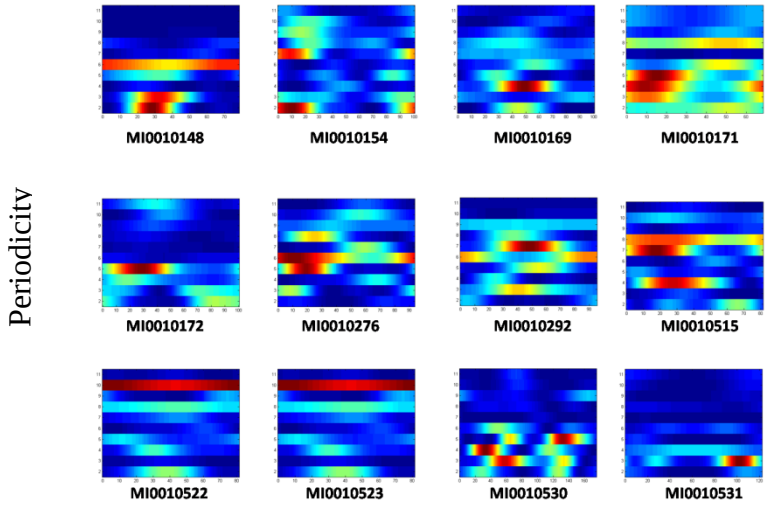


Figure 132: Spectrum corresponding to Periodicity 2-11 for sequences MI0010148 TO MI0010531

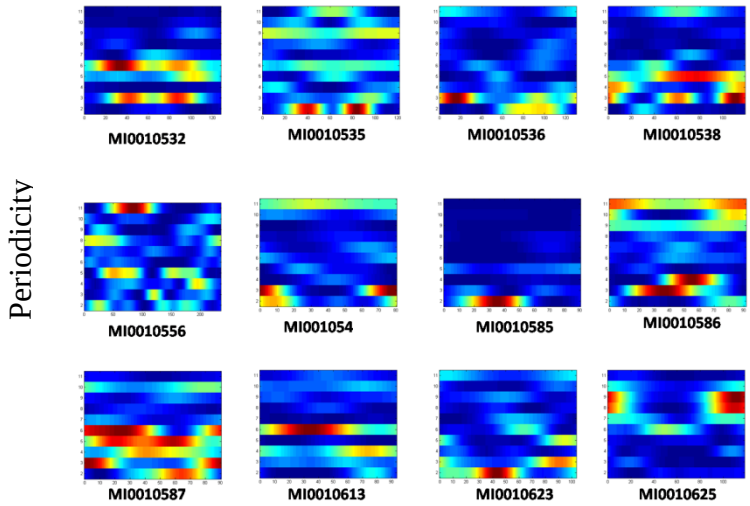


Figure 133: Spectrum corresponding to Periodicity 2-11 for sequences MI0010532 TO MI0010625

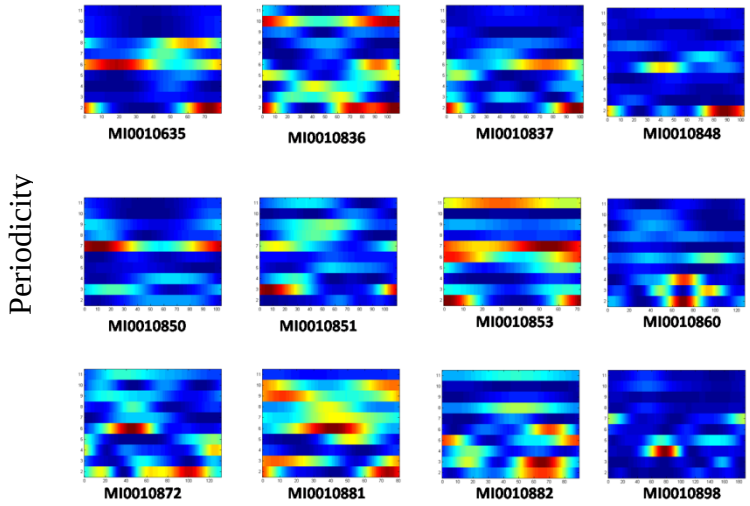


Figure 134: Spectrum corresponding to Periodicity 2-11 for sequences MI00010635 TO MI0010898

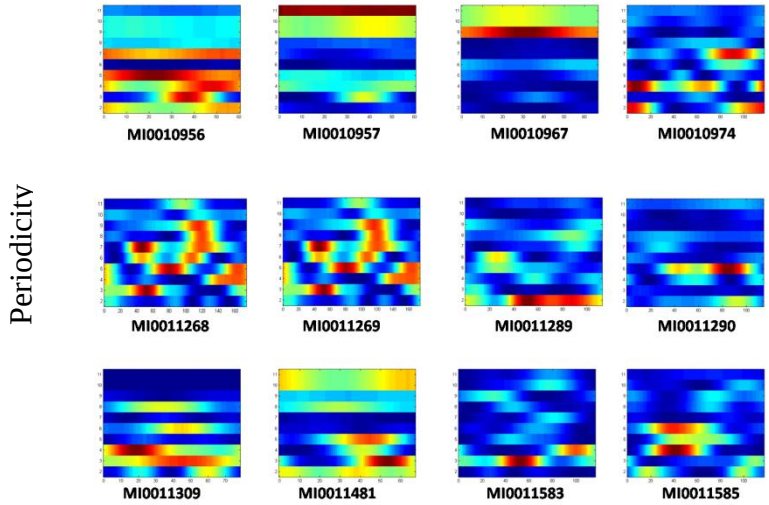


Figure 135: Spectrum corresponding to Periodicity 2-11 for sequences MI00010956 TO MI0011585

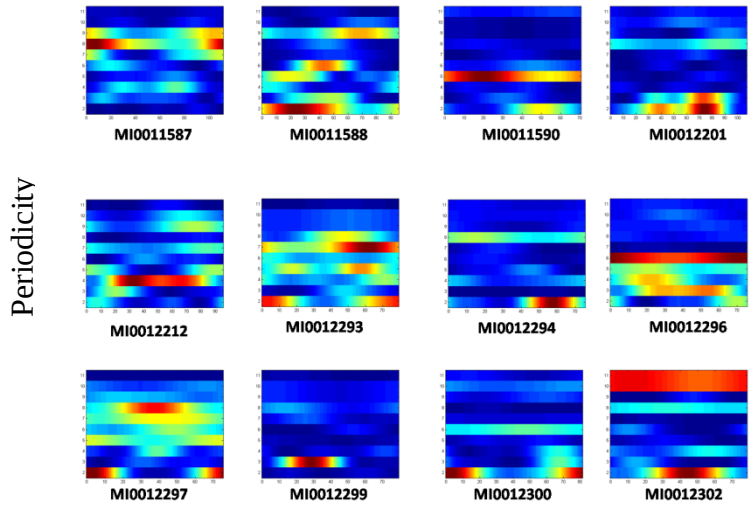


Figure 136: Spectrum corresponding to Periodicity 2-11 for sequences MI00011587 TO MI0012303

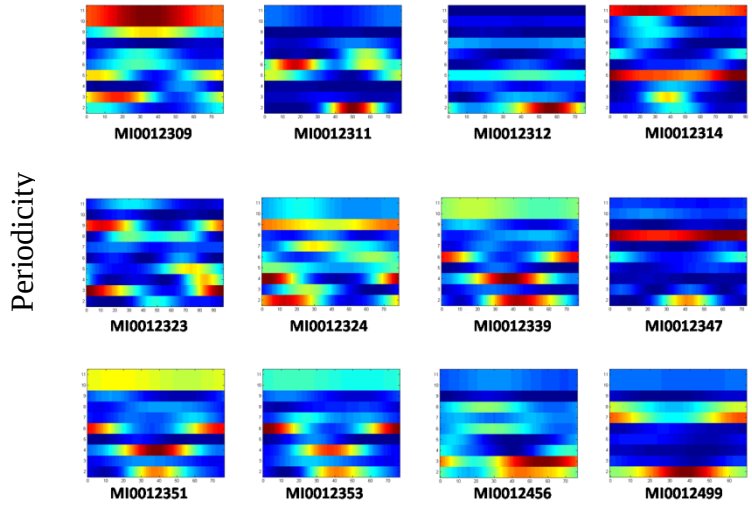


Figure 137: Spectrum corresponding to Periodicity 2-11 for sequences MI00012309 TO MI0012499

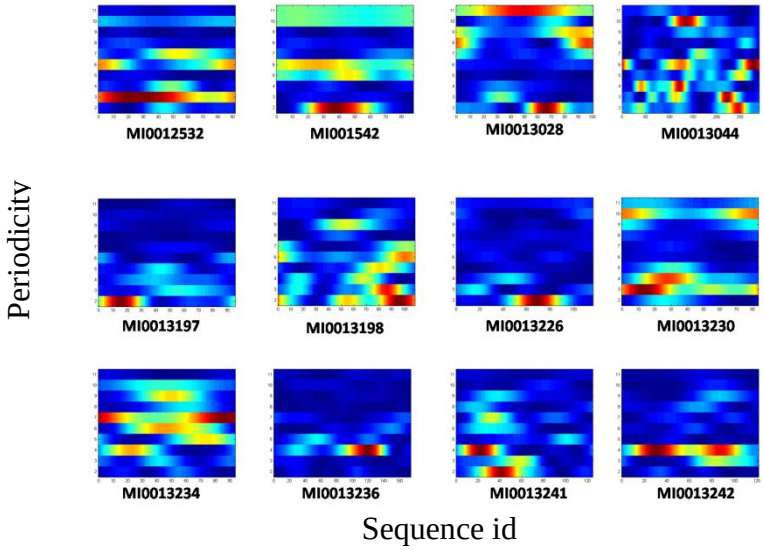


Figure 138: Spectrum corresponding to Periodicity 2-11 for sequences MI00012532 TO MI0013242

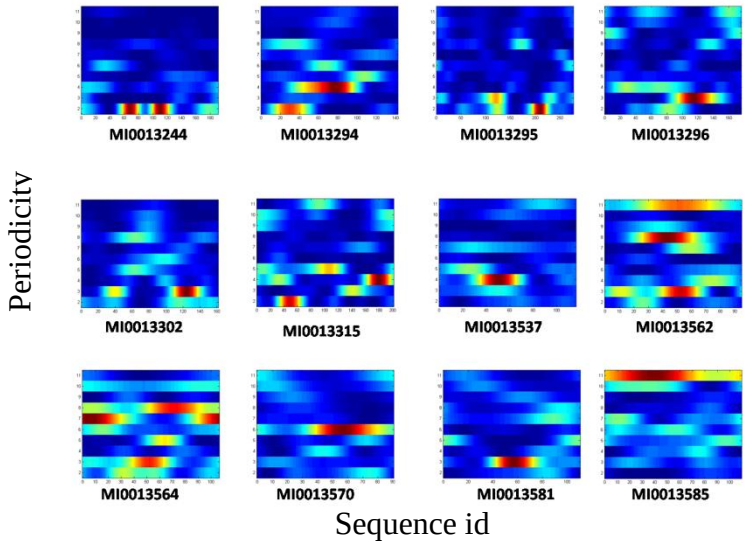


Figure 139: Spectrum corresponding to Periodicity 2-11 for sequences MI00013244 TO MI0013585

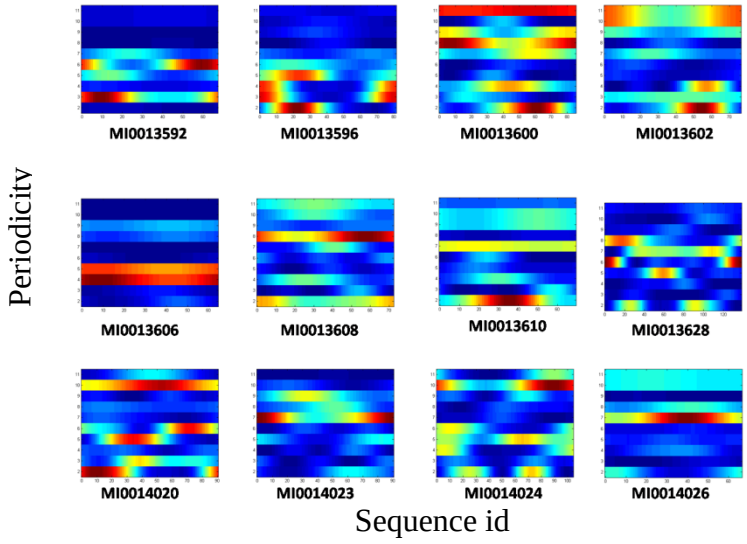


Figure 140: Spectrum corresponding to Periodicity 2-11 for sequences MI00013592 TO MI0014026

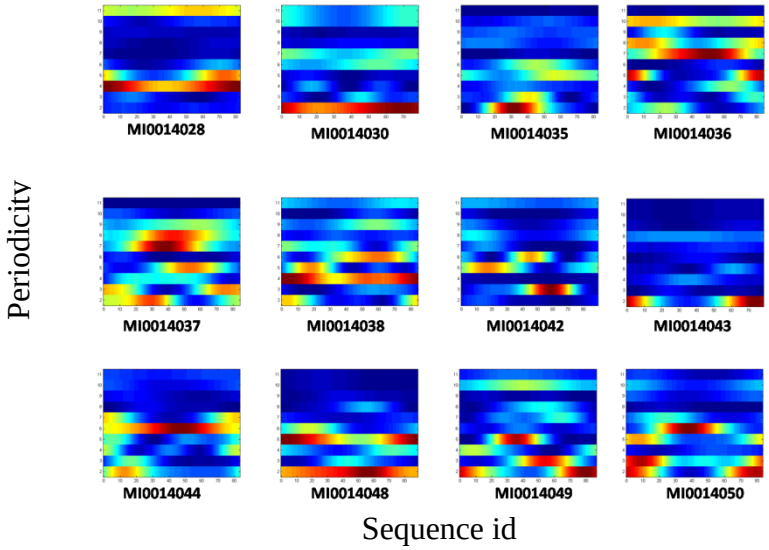


Figure 141: Spectrum corresponding to Periodicity 2-11 for sequences MI00014028 TO MI0014050

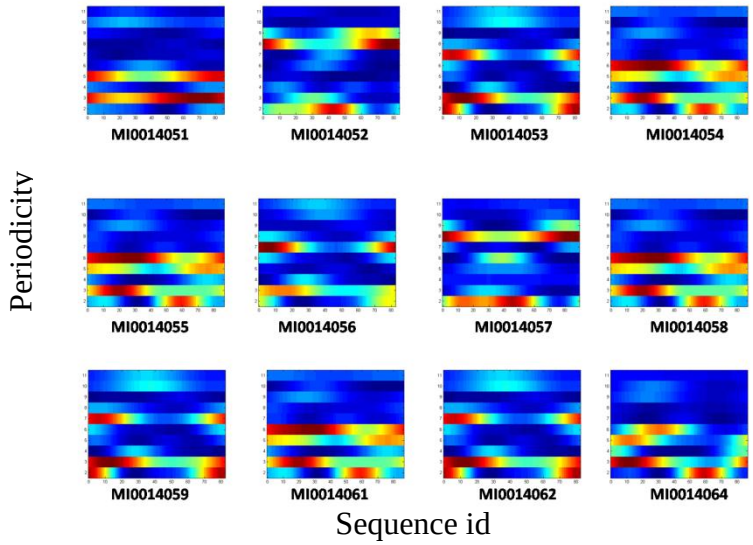


Figure 142: Spectrum corresponding to Periodicity 2-11 for sequences MI00014051 TO MI0014064

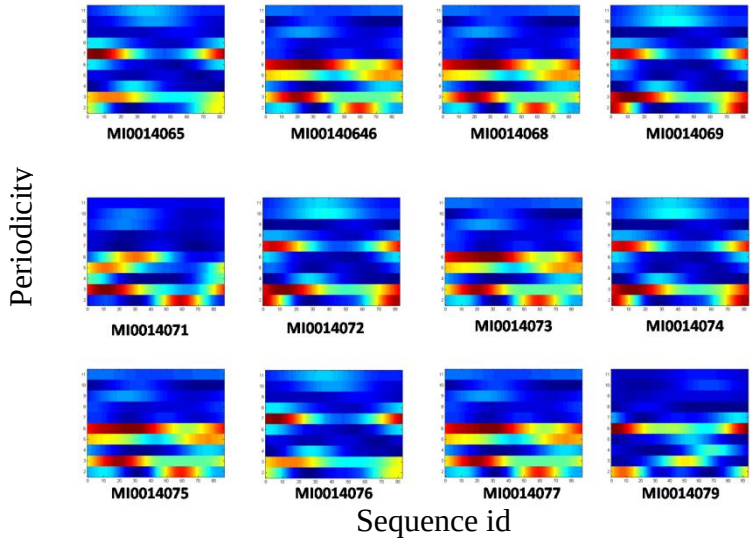


Figure 143: Spectrum corresponding to Periodicity 2-11 for sequences MI00014065 TO MI0014079

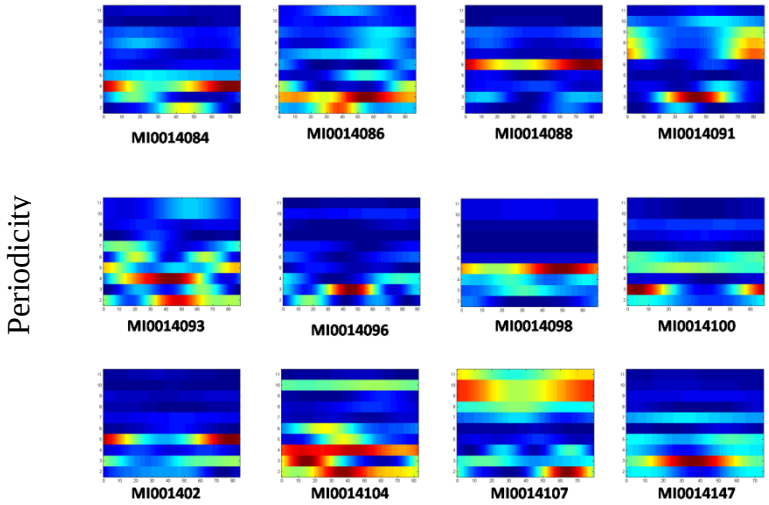


Figure 144: Spectrum corresponding to Periodicity 2-11 for sequences MI00014084 TO MI0014147

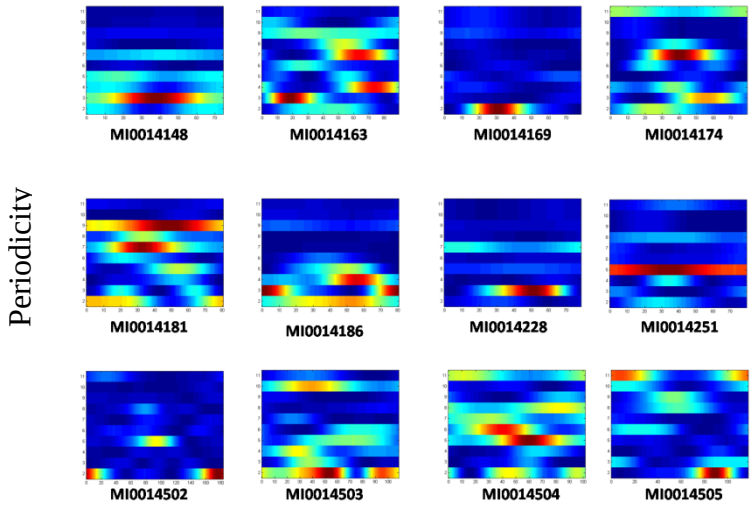


Figure 145: Spectrum corresponding to Periodicity 2-11 for sequences MI00014148 TO MI0014505

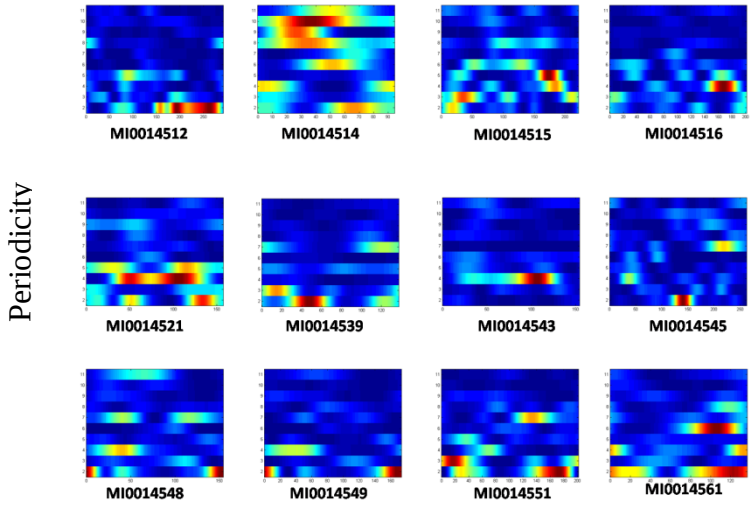


Figure 146: Spectrum corresponding to Periodicity 2-11 for sequences MI00014512 TO MI0014561

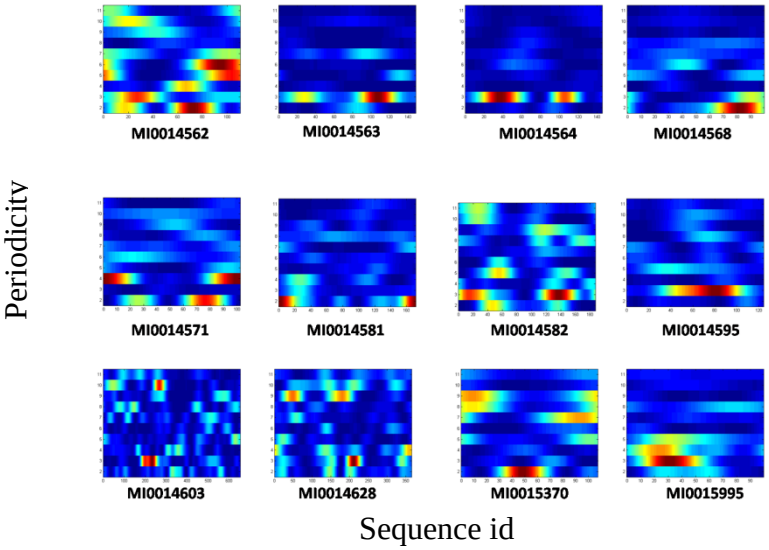


Figure 147: Spectrum corresponding to Periodicity 2-11 for sequences MI00014562 TO MI0015995

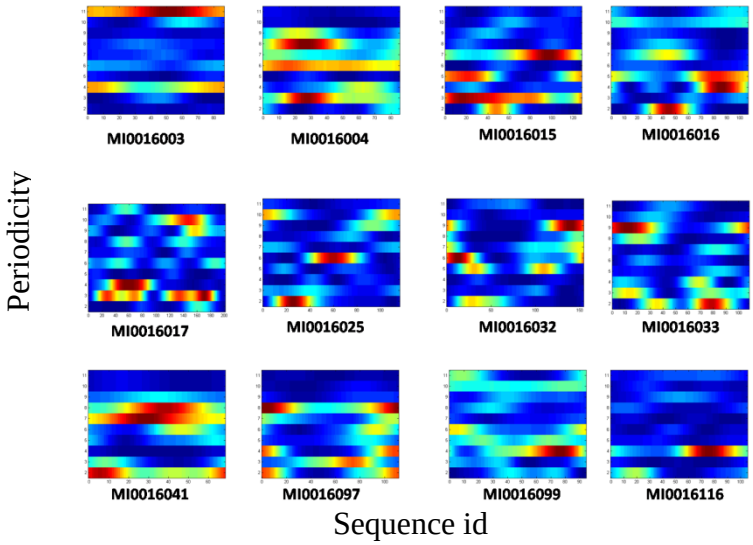


Figure 148: Spectrum corresponding to Periodicity 2-11 for sequences MI00016003 TO MI0016116

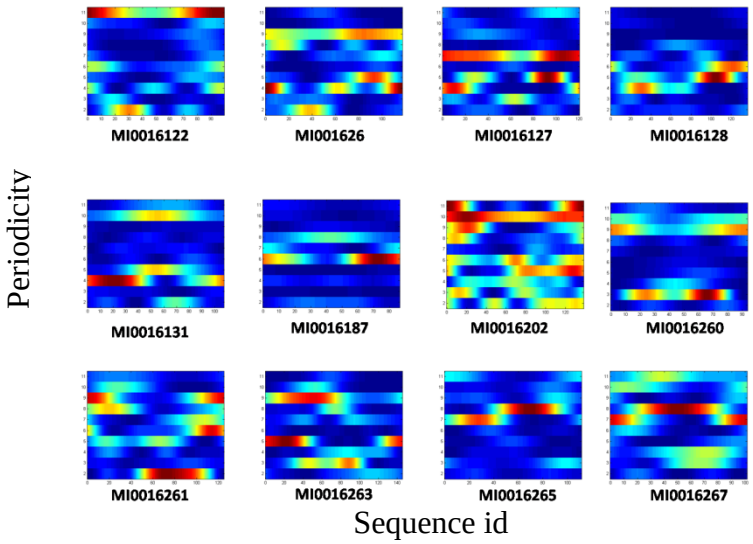


Figure 149: Spectrum corresponding to Periodicity 2-11 for sequences MI00016122 TO MI0016267

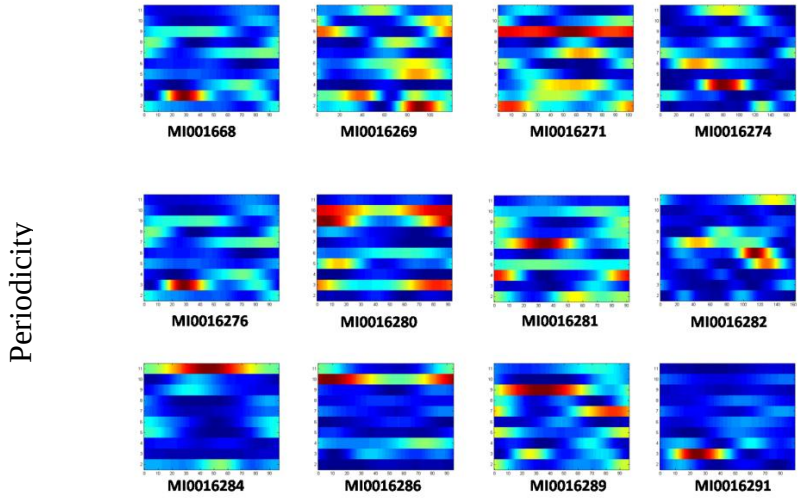


Figure 150: Spectrum corresponding to Periodicity 2-11 for sequences MI0001668 TO MI0016291

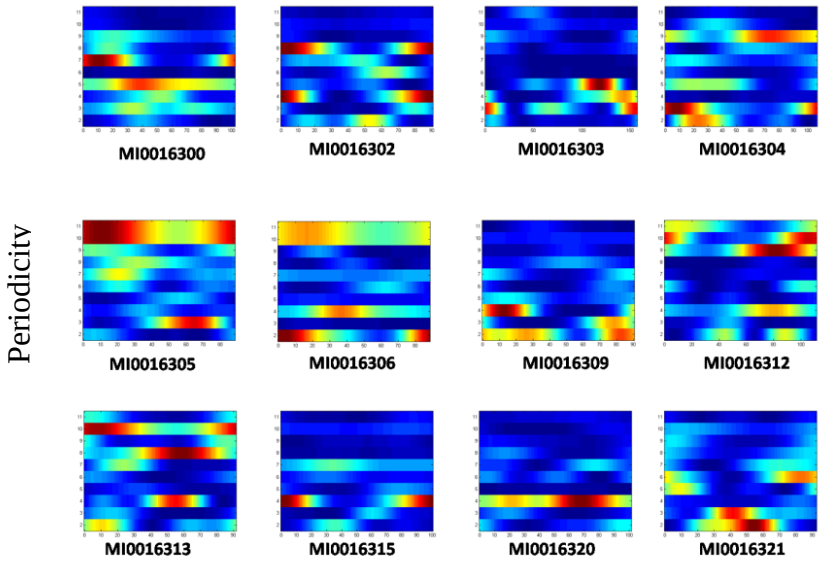


Figure 151: Spectrum corresponding to Periodicity 2-11 for sequences MI00016300 TO MI0016321

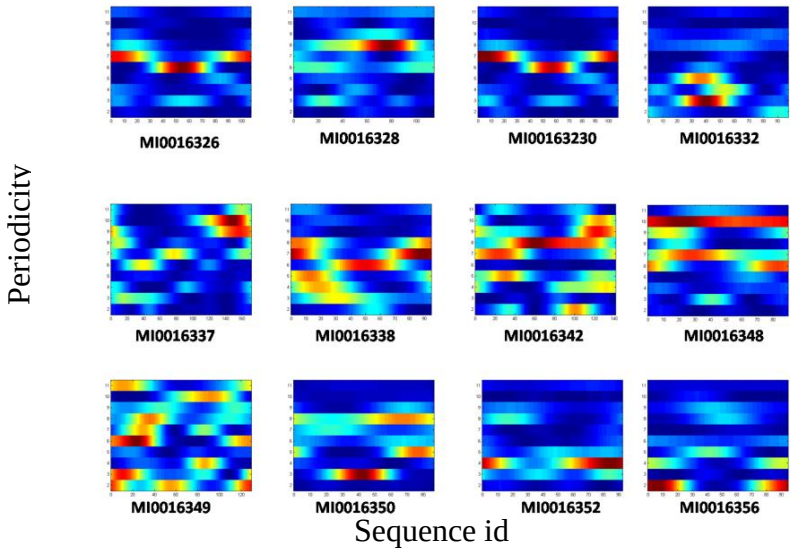


Figure 152: Spectrum corresponding to Periodicity 2-11 for sequences MI00016326 TO MI0016356

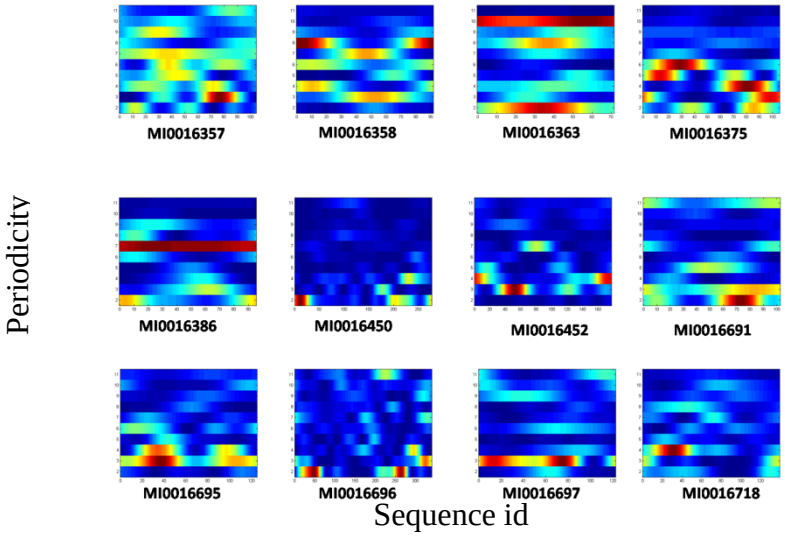


Figure 153: Spectrum corresponding to Periodicity 2-11 for sequences MI00016357 TO MI0016718

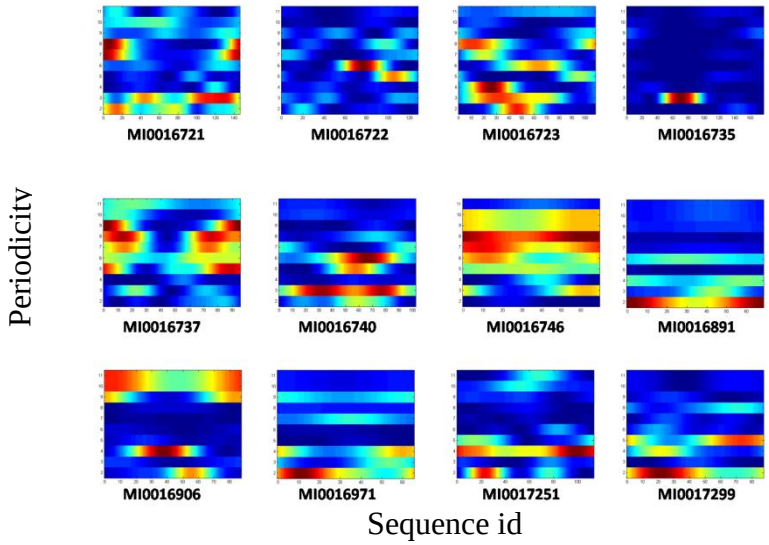


Figure 154: Spectrum corresponding to Periodicity 2-11 for sequences MI00016721 TO MI0017299

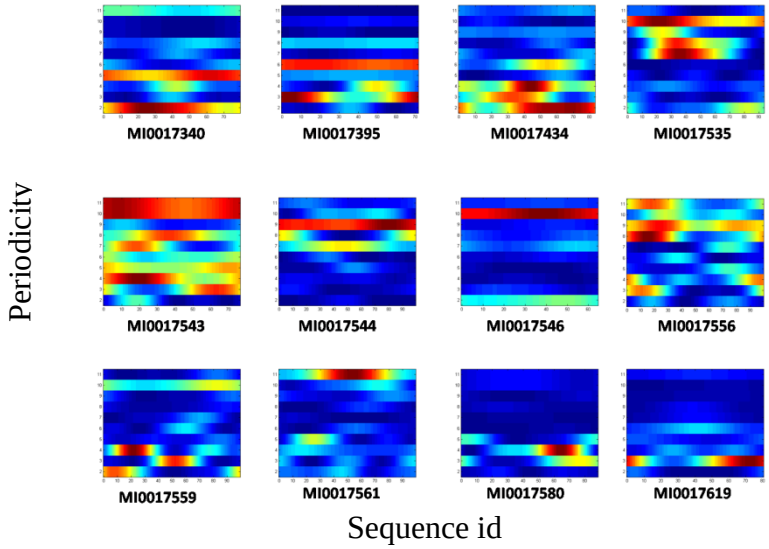


Figure 155: Spectrum corresponding to Periodicity 2-11 for sequences MI00017340 TO MI0017619

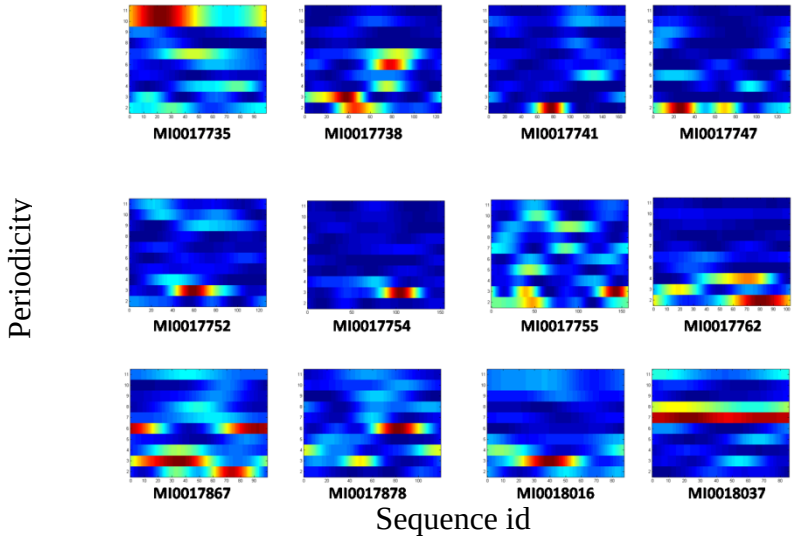


Figure 156: Spectrum corresponding to Periodicity 2-11 for sequences MI00017735 TO MI0018037

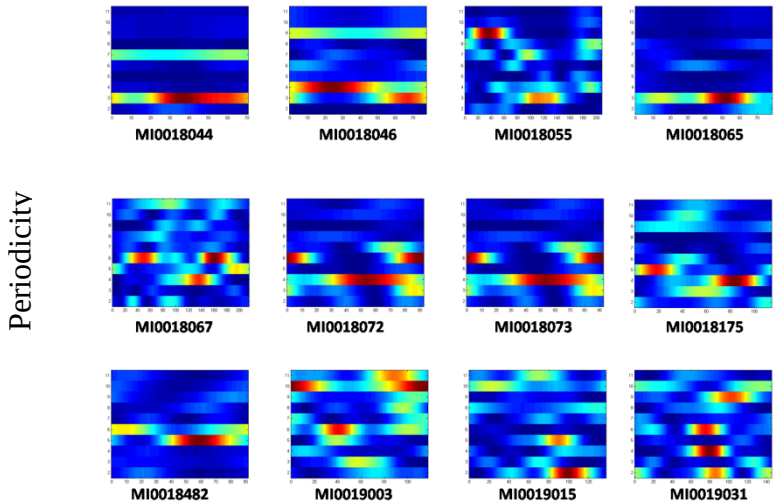


Figure 157: Spectrum corresponding to Periodicity 2-11 for sequences MI00018044 TO MI0019031

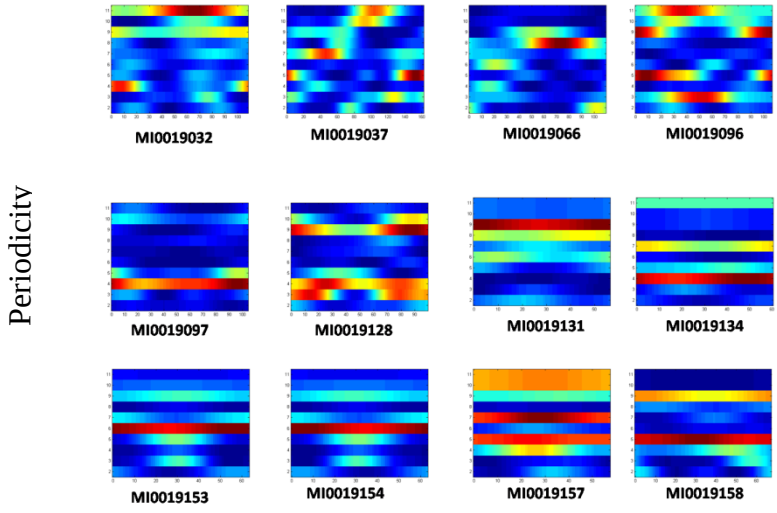


Figure 158: Spectrum corresponding to Periodicity 2-11 for sequences MI00019032 TO MI0019158

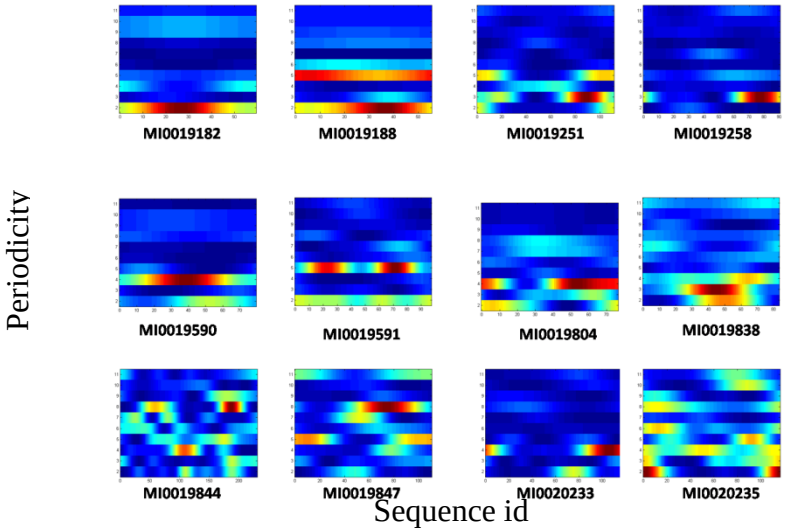


Figure 159: Spectrum corresponding to Periodicity 2-11 for sequences MI00019182 TO MI0020235

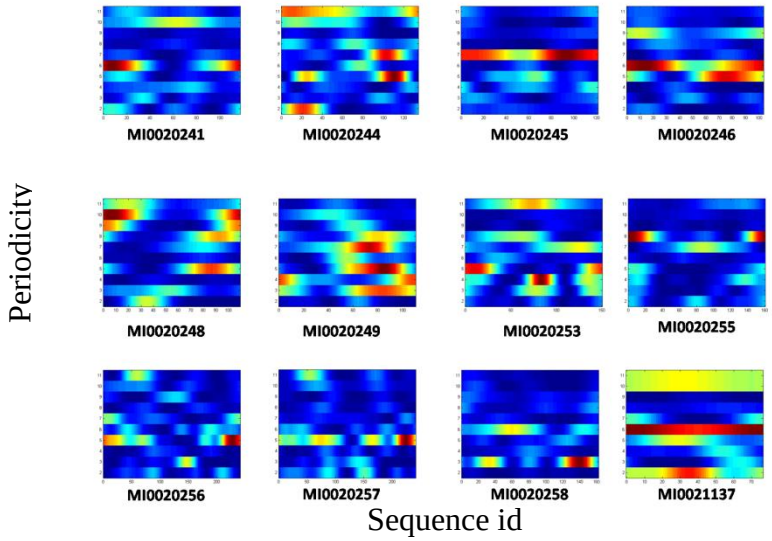


Figure 160: Spectrum corresponding to Periodicity 2-11 for sequences MI00020241 TO MI0021137

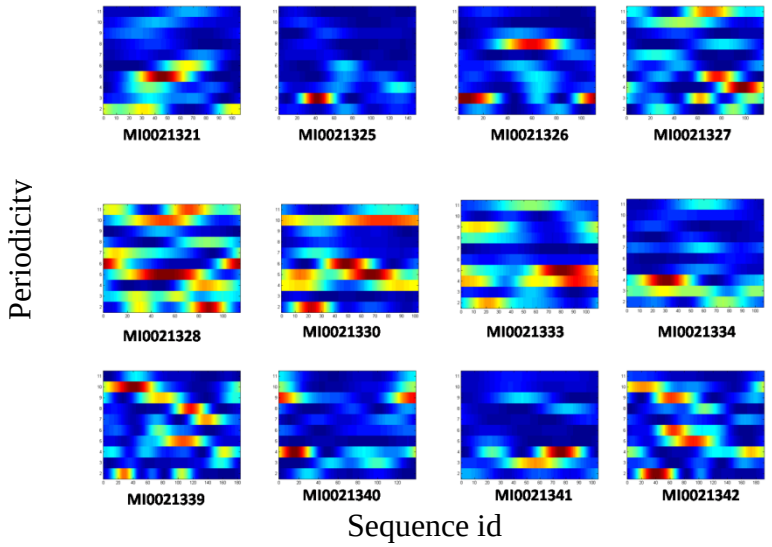


Figure 161: Spectrum corresponding to Periodicity 2-11 for sequences MI00021321 TO MI0021342

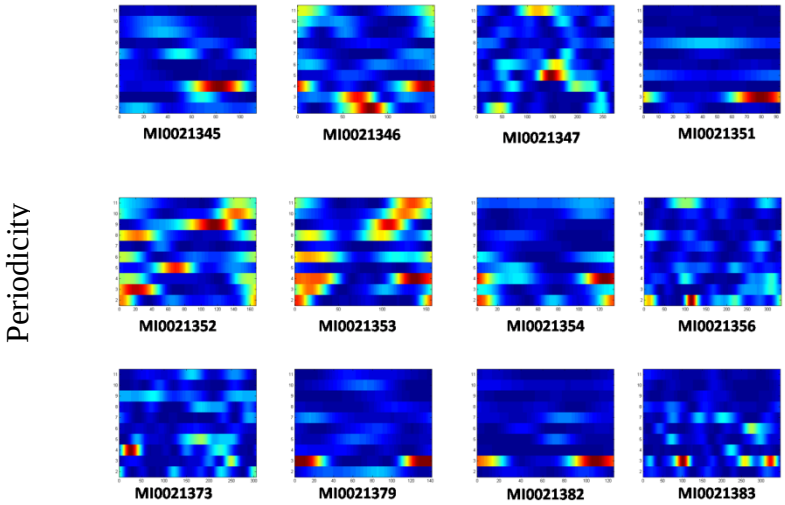


Figure 162: Spectrum corresponding to Periodicity 2-11 for sequences MI00021345 TO MI0021383

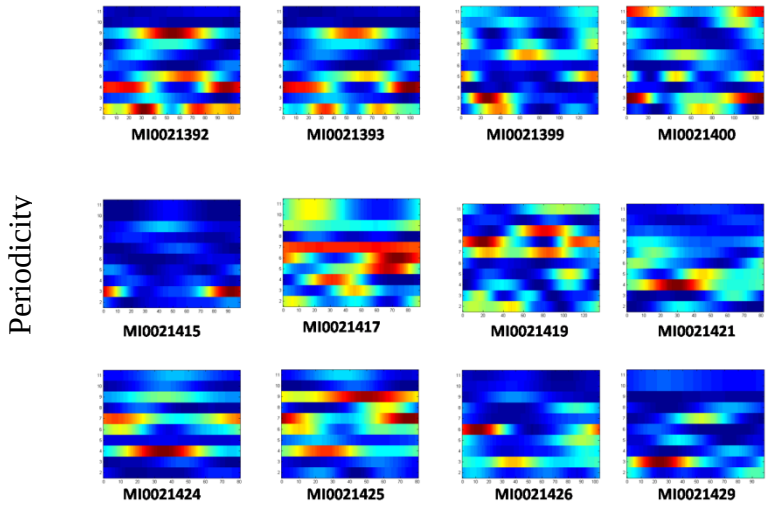


Figure 163: Spectrum corresponding to Periodicity 2-11 for sequences MI00021392 TO MI0021429

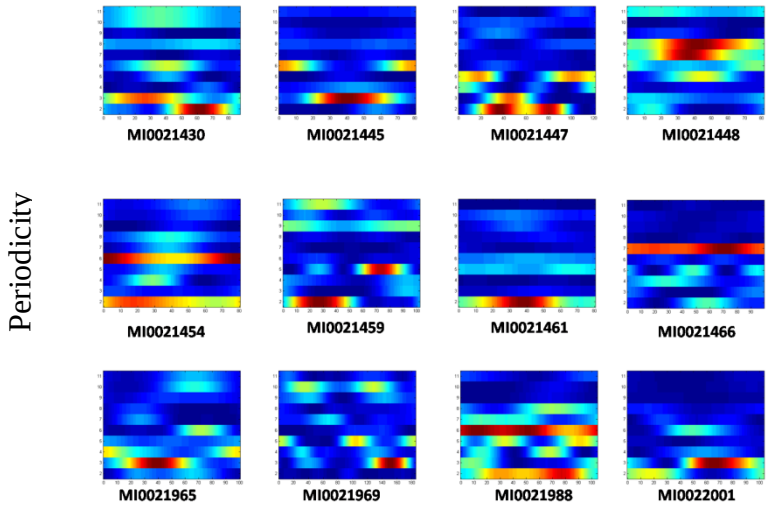


Figure 164: Spectrum corresponding to Periodicity 2-11 for sequences MI00021430 TO MI0022001

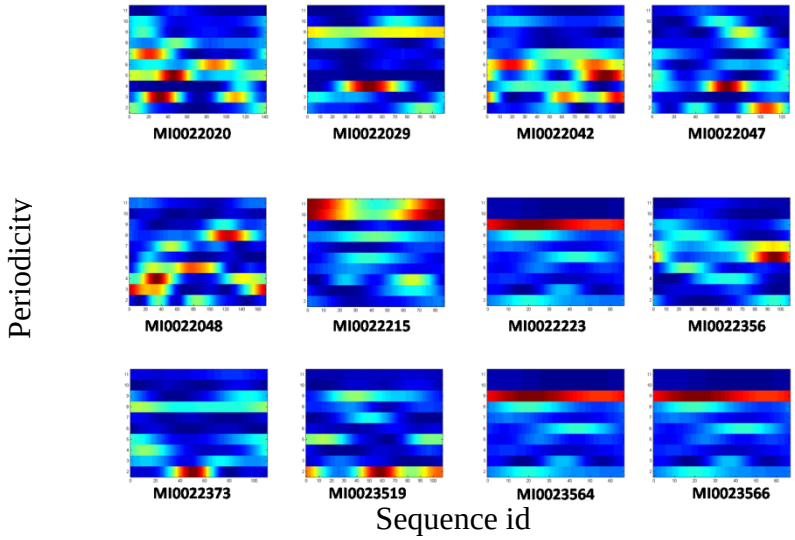


Figure 165: Spectrum corresponding to Periodicity 2-11 for sequences MI00022020 TO MI0023566

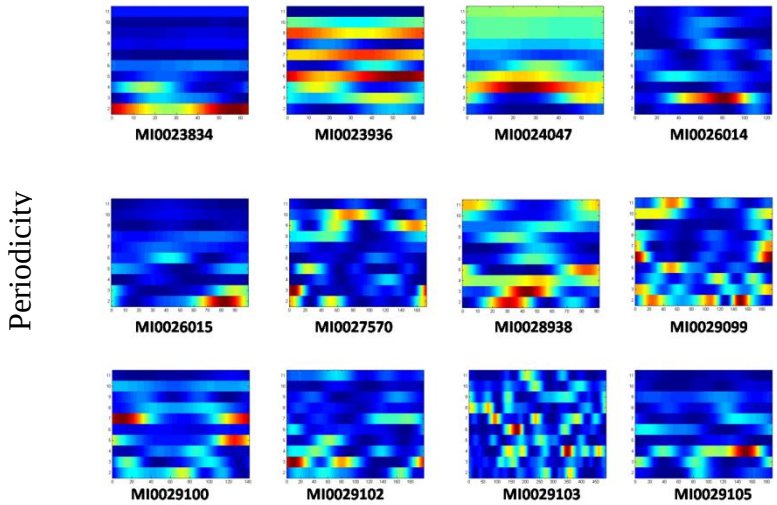


Figure 166: Spectrum corresponding to Periodicity 2-11 for sequences MI00023834 TO MI0029105

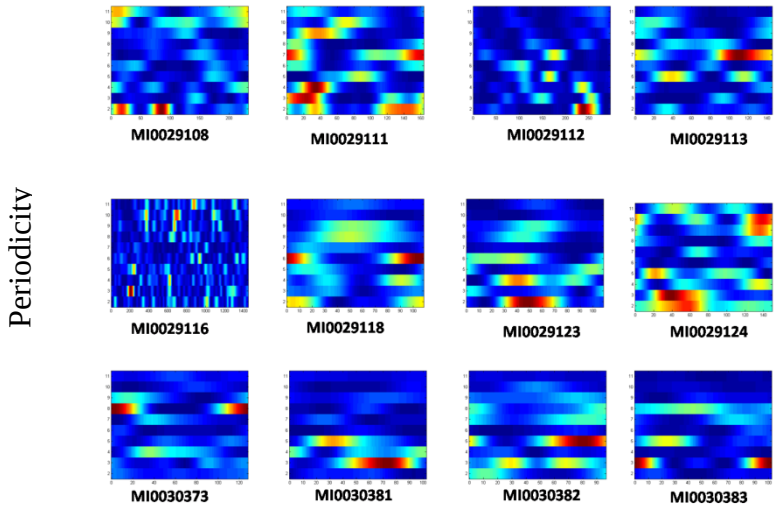


Figure 167: Spectrum corresponding to Periodicity 2-11 for sequences MI00029108 TO MI0030373

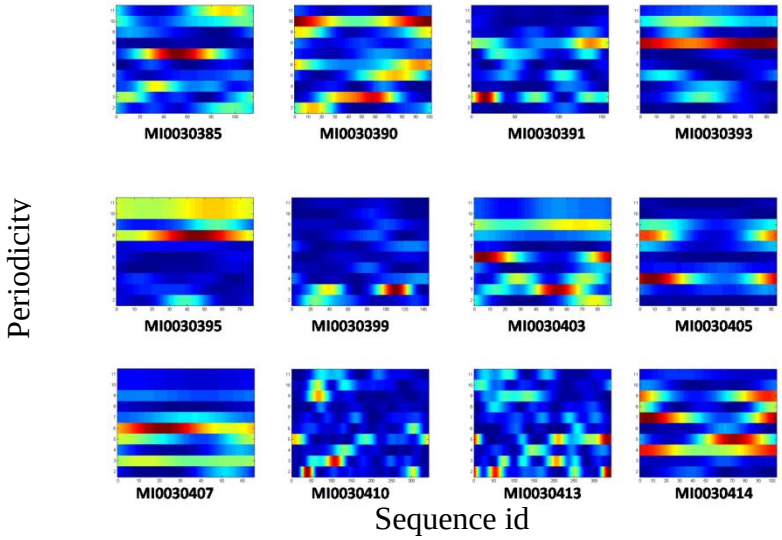


Figure 168: Spectrum corresponding to Periodicity 2-11 for sequences MI00030385 TO MI0030414

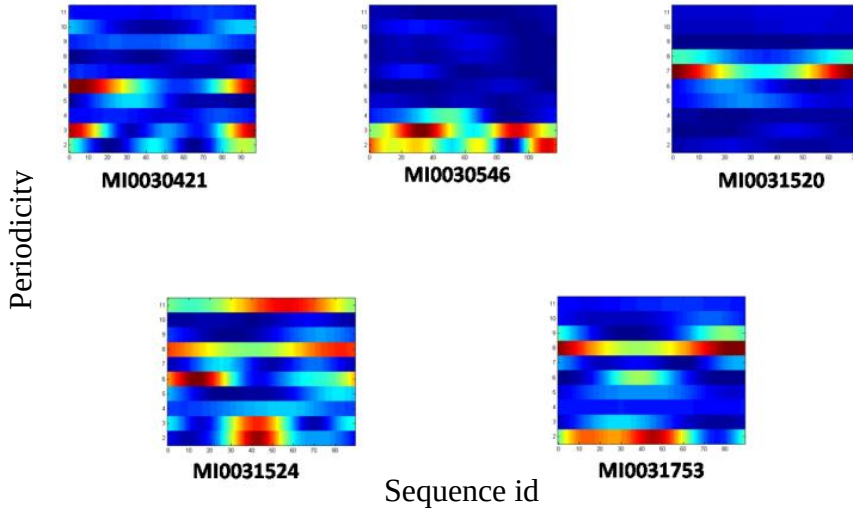


Figure 169: Spectrum corresponding to Periodicity 2-11 for sequences MI00030421 TO MI0031753

It has been previously reported that the periodicities 10 and 11 are related to helical function in DNA. The reason is that the average length of a helical turn in B-DNA is about 10.5 nucleotides [13, 14]. Extrapolating this, we hypothesize that periodicities 10 and 11 are also possibly responsible for the helical function in RNA. Analyzing the MST results we have only 17 sequences that show peak at periodicities 10 and 11.

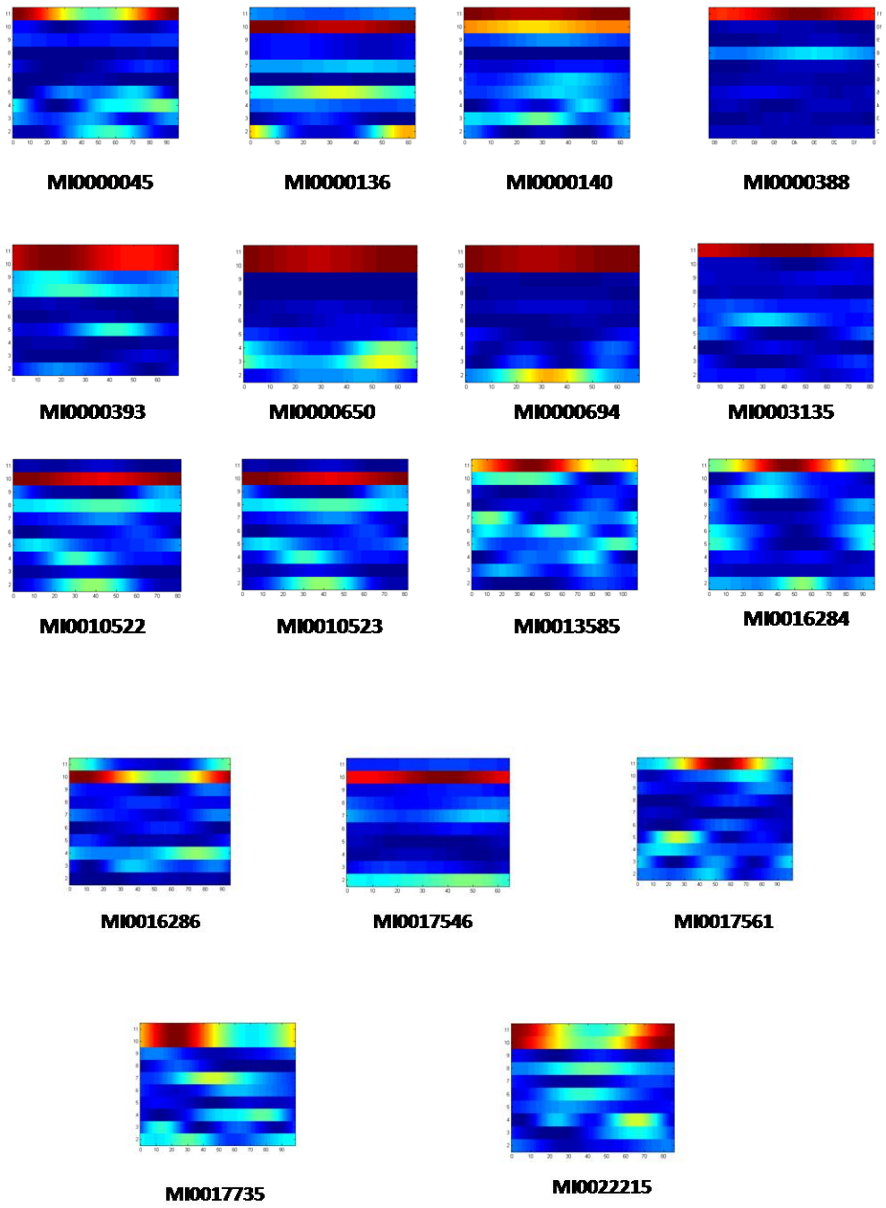
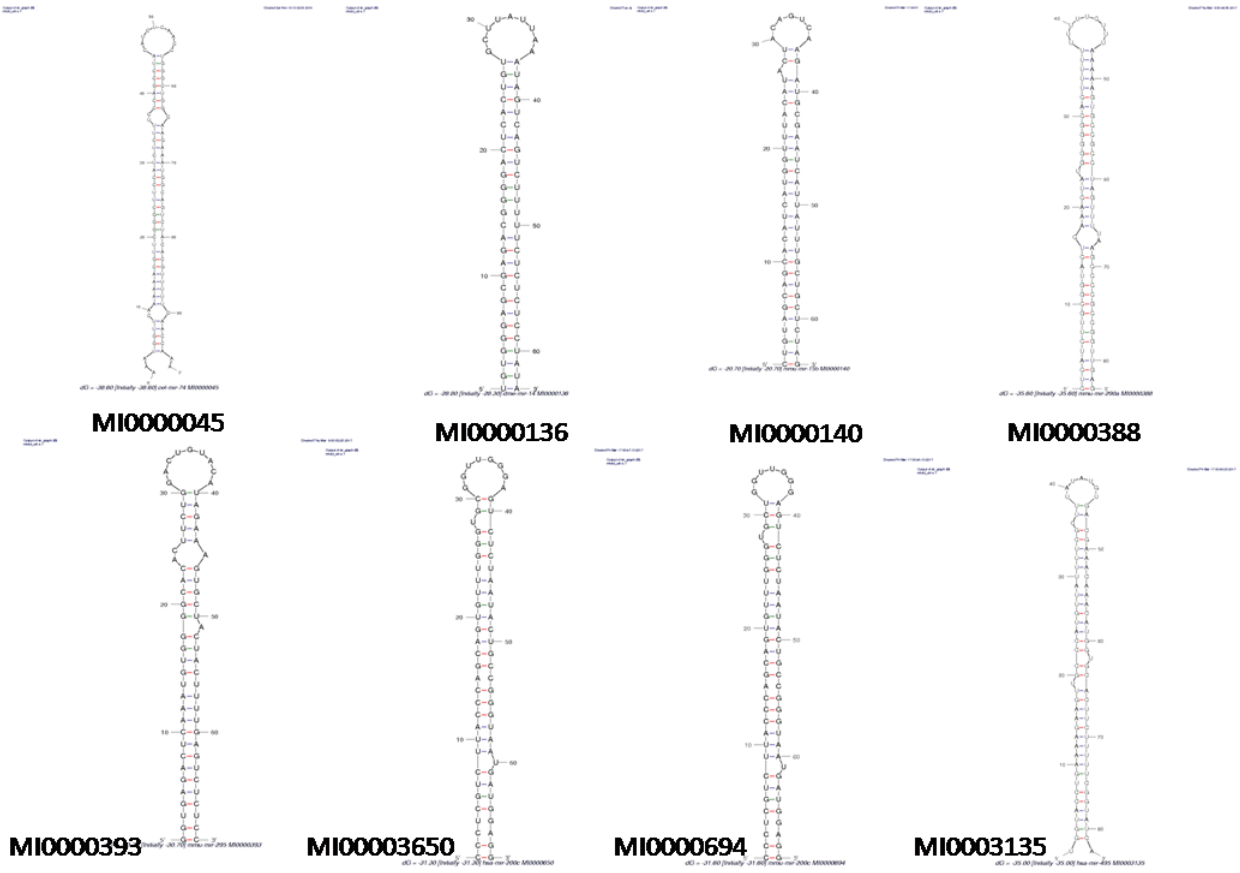


Figure 170: MST results of sequences which only contain periodicity 10 and 11

3.2 Mfold results

Mfold predicts the secondary structure for the high confidence data. M-fold uses the free energy minimization algorithm. The seventeen sequences that showed periodicity at 10 and 11 exclusively are shown with the predicted secondary structure in Figure 171:



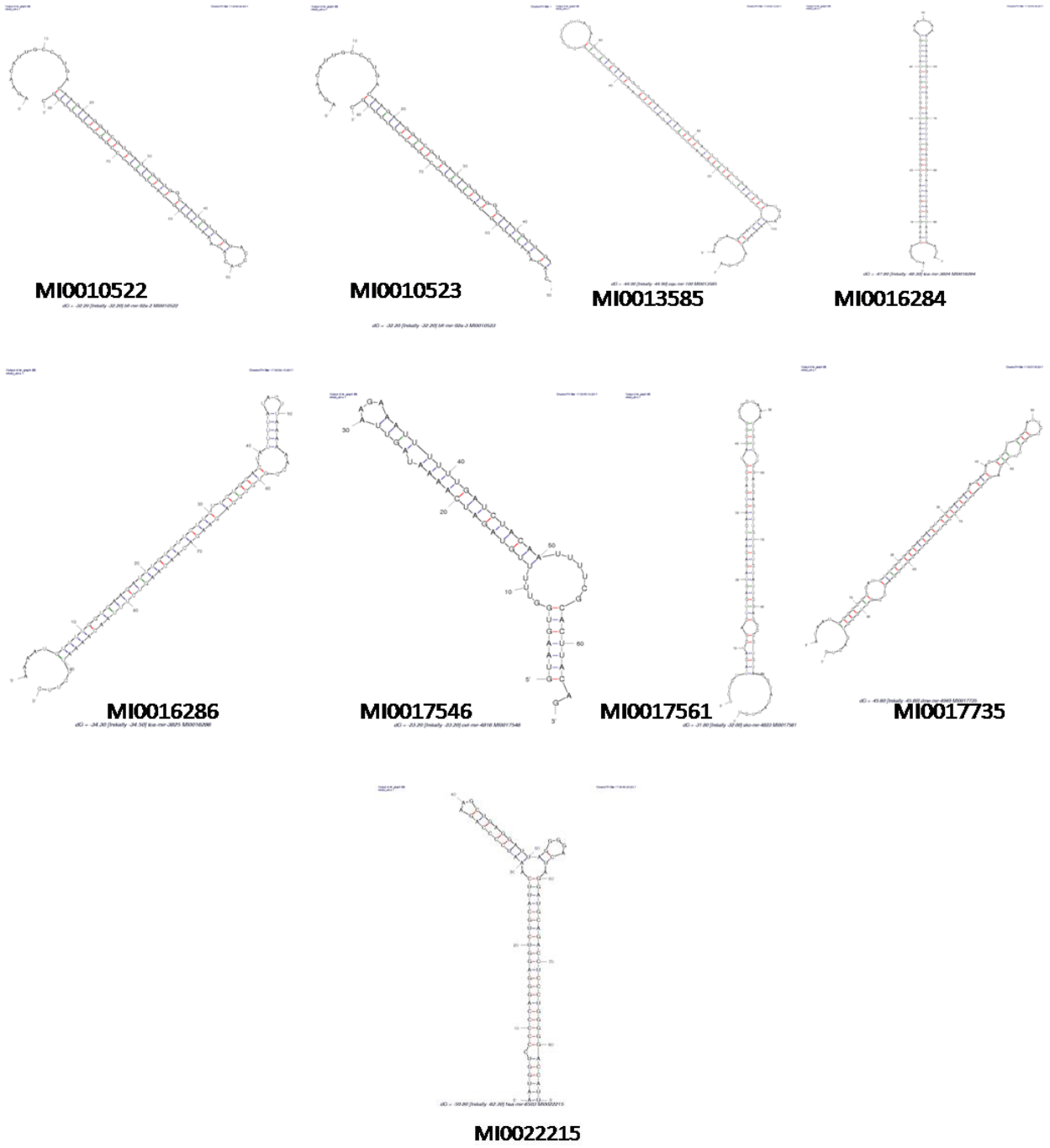


Figure 171: Mfold results of sequences which only contain Periodicity 10 and 11

Chapter 4 - Discussion

MicroRNAs (miRNA) are very small pieces of RNA, which have a strong position in the cell. They can bind to RNA those codes for a protein, to repress this protein. miRNAs are involved in a variety of disease processes, playing a role in different types of cancer, for instance, breast cancer is often caused by an error in the production of RNA. This regulatory role makes miRNA very interesting as a drug target. We used signal processing methods for the secondary structure identification of miRNAs [15].

Signal processing methods are easy and fast which will help us to easily detect the secondary structures of the miRNAs. Moreover the signal processing methods are data independent so there is no need for prior training. During the identification of the secondary structure of miRNAs we find out that some periodicities are related to the stem region of the RNA region because the most of the periodicities which had shown the peaks in regions of sequences when matched with the results of the Mfold database then these periodicities are mostly related to the stem regions. There are total 1996 sequences on which we have implemented the MST algorithm. As there is more than one periodicity observed for each sequence, we analyzed the periodicities obtained and plotted as a histogram (Figure 172).

This histogram shows the number of sequences found in each periodicity. E.g 750 sequences are found in period 2 and 824 sequences are found in period. Among all the peridocities (2-11), periodicity 3 was found to be the highest in 824 sequences. And there are 160 sequences, where the periodicity 11 was found.

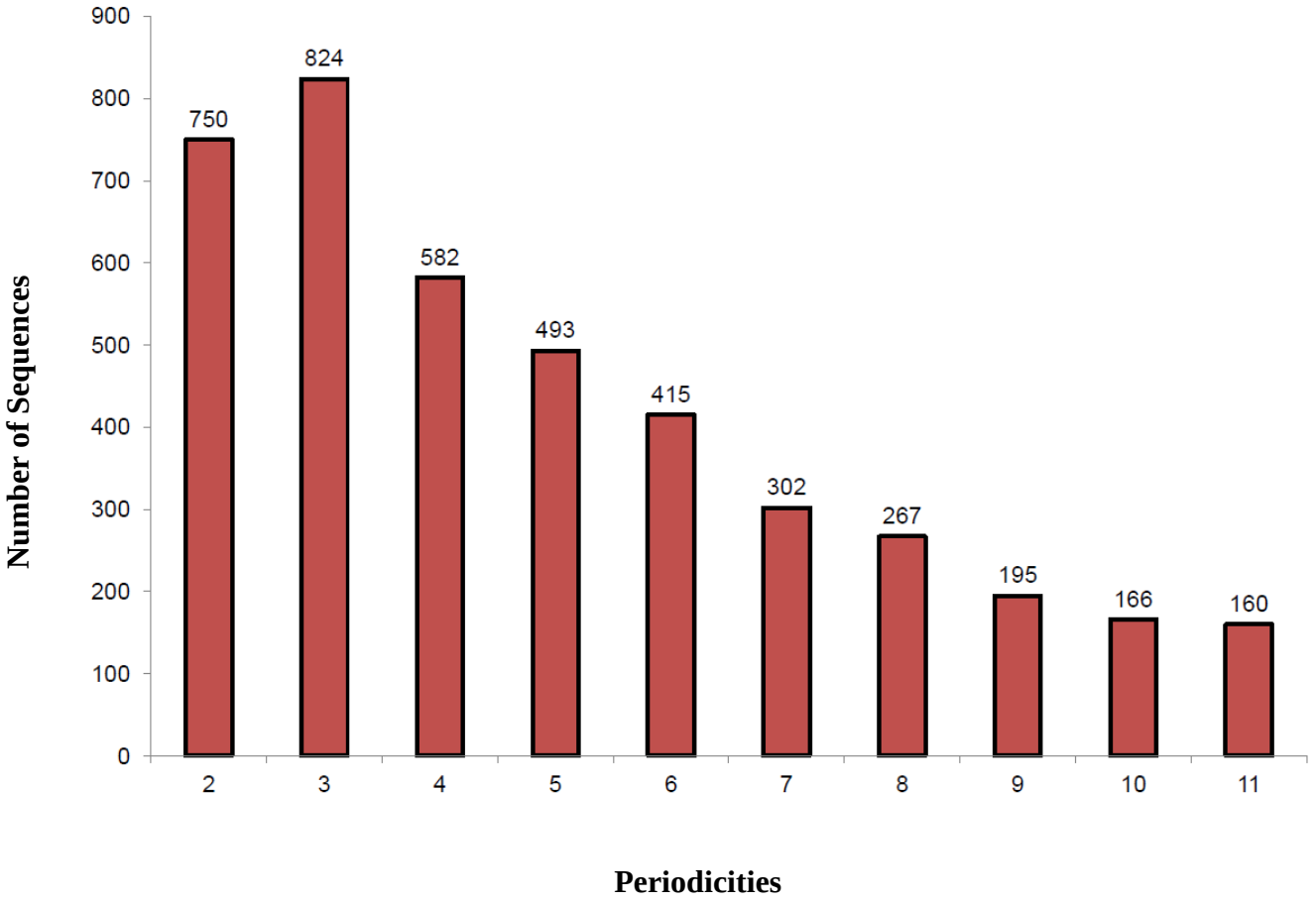


Figure 172: Histogram indicating the periodicity identified in each sequence.

The table of the sequences which we validated with the M-fold results is given in Table 2.

Table 2: MST Results validation With Mfold Results

Accession ID	Peaks in Spectrum	Mfold		
		Stem	**Bulge**	**Loop**
MI0000001	periodicity 7 (10-40)	5-88,86-89,12-41,57-84,40-41,52-54	9-11,54-55-56	42-51
MI0000002	periodicity 2(20-55)	2-6,87-91,7-14,76-83,17-35,73,54,36-40,52-48	84-86,15-16,76-74,53	41-47
MI0000003	periodicity 6(35-70)	7-35,65-93,38-42,59-63,43-46,57-54	36,37,64,58	47-53
MI0000004	periodicity 4(15-45)	4-8,94-98,11-19,82-90,20-28,72-80,29-31,68-70,35-41,59-65,45-46,52-53	9,10,81,71,32-34,66,67,42-44,54-58	47-51
MI0000005	periodicity 6(10-65)	2-5,88-85,7-13,81-75,17-23,7064,25-27,59-61,29-33,58-54,35-37,53-51,38-	6,82-84,14-16,71-74,24,63,62,28,34,50	40-47

Accession ID	Peaks in Spectrum	Mfold		
		Stem	Bulge	Loop
		39,48-49		
MI0000006	periodicity 2(45-85), periodicity 4(55-75), periodicity 6(30-40), periodicity (1--40)	5-35,65-97,39-48,55-64	36,37,38,65,55	49-54
MI0000007	periodicity 2(40-65), periodicity 7(0-25,75-100), periodicity 8(20-40), periodicity 5(30-40)	3-6,93-96,8-34,66-92,37-45,56-64	7,35,36,65	46-55
MI0000008	periodicity 2(5-45), periodicity 4(1-25,80-100), periodicity 6(25-50), periodicity 9(10-55)	5-11,92-98,15-33,69-87,37-46,57-66	12-14,88-91,34-36,68,67	47-56
MI0000009	periodicity 7(1-30,75-100)	1-5,91-95,7-21,73-87,22-29,64-71.31-38,56-63,41-42,51-52	6,72,30,39-40,53-55	43-50
MI0000010	periodicity 6(1-30,75-90)	2-6,82-86,8-23,66-81,24-25,63-64,27-42,47-62	7,65,26	43-46
MI0000011	periodicity 4(10-45)	2-6,87-91,7-29,62-84,31-42,50-61	85,86,30	43-49
MI0000012	periodicity 3(1-5,70-100)	1-6,88-93,10-27,87-70,28-31,65-68,34-40,58-64	7-9,69,32,33	41-57
MI0000013	periodicity 2(20-40), periodicity 3(45-60), periodicity 5(10-45), periodicity 11(1-40,75-100)	6-26,66-86,31-38,58-67,41-42,51-52,	27-30,39-40,57-53	43-50
MI0000014	periodicity 6(40-65)	3-9,88-95,14-29,88-73,30-34,67-71,36-40,62-66,42-44,59-61,45-46,56-57	10-13,72,35,41,58	47-55
MI0000015	periodicity 2(1-25,75-100), periodicity 3(15-25), periodicity 7(5-25)	1-5,90-94,9-24,86-71,25-27,67-69,30-37,59-66,44-47,53-56	6-8,87-89,70,28,29,38-43,57,58	48-52
MI0000016	periodicity 2(10-30), periodicity 5(50-100), periodicity 9(1-45,85-100)	7-23,71-87,24-26,67-69,29-36,59-66,40-43,54-57	70,27,28,37,38,39,58	44-53
MI0000017	periodicity 4(45-80)	5,6,89,90,12-22,73-83,25-29,65-69,31-43,52-64	7-11,84-88,23,24,70-72,30	44-51
MI0000018	periodicity 3(55-85), periodicity 4(45-70)	5-7,93-95,8-10,89-91,16-26,73-83,28-31,69-72,34-40,59-65,42-45,52-55	92,11-15,84-88,27,32,33,66-68,41,57,58	43-51
MI0000019	periodicity 2(50-85), periodicity 3(10-40), periodicity 4(70-90), periodicity 5(30-55), periodicity 8(80-100)	1-6,82-87,9-26,63-80,29-42,48-61	7,8,81,27,28,62	43-47

Accession ID	Peaks in Spectrum	Mfold		
		Stem	Bulge	Loop
MI0000020	periodicity 3(50-80)	11-34,67-90,37-47,56-66	35,36	48-55
MI0000021	periodicity 3(60-95)	1-3,91-93,4-30,63-89,34-40,55-61,42-44,52-54	90,62,31,32,33,41	45-51
MI0000022	periodicity 3(65-95)	2-4,88-90,10,82,12-30,63-81,32-38,56-62,39-42,51-54	11,31,61,55	43-50
MI0000023	periodicity 5(45-75)	1-6,91-96,7-8,88-89,9-12,82-85,18-21,73,76,25-27,67-69,29-31,64-66,35-43,53-61,46-47,51-52	90,86,87,13-17,77-81,22-24,70-72,28,32-34,62-63,44,45	48-50
MI0000024	periodicity 4(15-45)	1-6,89-94,8-11,81-84,18-23,69-75,27-30,63-66,33-42,52-61,43-44,49-50	7,85-88,12-17,76-80,24-26,68,67,31-32,62,51	45-48
MI0000025	periodicity 3(60-85), periodicity 4(1-10,40-100)	1-6,94-99,10-22,79-91,23-26,74-77,27-34,65-72,38-40,59-61,44-47,52-55	7-9,92,93,78,73,35-37,62-64,41-43,56-58	48-51
MI0000026	periodicity 6(20-60)	7-10,91-94,13-19,84-90,21-36,68-83,37-44,59-66	11-12,20,67	45-58
MI0000027	periodicity 2(1-50), periodicity 3(15-50), periodicity 8(55-85), periodicity 10(5-50)	3-4,93,94,5-6,90-91,9-36,62-89,40-42,59-61,43-45,55-57	92,7-8,37-39,58	46-54
MI0000028	periodicity 2(20-40)	2-4,86-88,8-25,65-82,27-33,58-64,35-43,49-57	5-7,83-85,26,34	44-48
MI0000030	periodicity 5(20-70)	1-7,94-100,14-28,75-89,29-41,61-73,49-52,57-60	8-13,90-93,74,42-48	53-56
MI0000031	periodicity 4(25-40), periodicity 5(1-20,75-100), periodicity 9(35-85)	17-31,72-86,32-35,67-70,37-47,56-66	71,36	48-55
MI0000032	periodicity 2(50-90)	8-14,91-98,15-38,66-89,40-46,59-65,47-48,57-56	90,39,50	49-55
MI0000034	periodicity 2(25-55)	5-12,92-99,13-29,73-89,30-33,66-69,35-39,62-65,41-45,54-58	90-91,71-72,34,40,59-61	46-53
MI0000035	periodicity 4(60-95)	7-10,100-103,13-32,77-96,35-37,73-75,38-44,64-70,45-47,60-62	11-12,97-99,76,33-34,71,72,63	48-59
MI0000036	periodicity 4(65-100), periodicity 7(1-35,85-100)	C	9-10,85-88,67,28-29,35-37,59-61,41-43,52-55	46-50
MI0000037	periodicity 3(20-40), periodicity 8(1-25,60-100)	1-5,89-93,9-15,81-87,19-26,70-77,28-40,57-69	6-8,88,16-18,80-78,27	41-56

Accession ID	Peaks in Spectrum	Mfold		
		Stem	Bulge	Loop
MI0000038	periodicity 3(25-65)	13-30,72-89,33-34,57-68	31,32,69,70,71	45-46
MI0000041	periodicity 4(45-90), periodicity 6(70-95)	9-50,57-98		51-56
MI0000042	periodicity 2(15-45), periodicity 3(1-25,95-100)	1-3,86-88,4-24,64-84,26-42,47-63	85,25	43-46
MI0000043	periodicity 7(40-60). periodicity 10(0-30,65-100), periodicity 11(0-30,65-100)	1-6.94-99,7-27,70-90,29-39,59-69	28,91-93	40-58
MI0000044	periodicity 2(30-50),periodicity 10(30-100),periodicity 11(30-100)	10-12,96-98,14-31,91-74,32-42,62-72,45-46,6061	13,92-95,73,43	47-59
MI0000045	periodicity 11(1-25,65-100)	4-8,91-95,11-35,65-89,38-45,56-63	9,10,90,36,37,64	46-55
MI0000046	periodicity 4(1-20,97-100)	8-13,91-96,16-43,60-87,44-47,55-58	14-15,88-89,59	48-54
MI0000048	periodicity 6(9-16,85-100), periodicity 9(60-90)	9-13,96-100,15-45,60-90,48-50,56-58	,14,91-95,46,47,59	51-55
MI0000050	periodicity 2(65-80), periodicity 3(20-45)	3-7,94-98,8-9,91-92,14-35,64-85,38-41,59-62,44-46,52-54	95.10-13,86-90,36,37,63,42,43,55-58	47-51
MI0000051	periodicity 3(25-50)	8-10,91-93,13-28,72-87,29-33,65-69,35-43,56-64	11-12,88-90,70-71,34	44-55
MI0000052	periodicity 17(1-25,95-110)	7-9,101-103,12-15,97-100,17-46,65-94,47-49,59-61	10,11,16,96,95,62-64	50-58
MI0000053	periodicity 2(30-50)	7-11,92-96,16-29,73-86,33-39,63-69,41-44,59-62,46-48,56-58	12-15,87-91,30-32,70-72,40,45	49-55
MI0000054	perod2(20-35), periodicity 3(35-70), periodicity 10(1-100), periodicity 9(40-90), periodicity 4(10-55)	1-3,96-98,7-30,67-90,32-39,59-66,40-41,56-57,42-43,52-53	4-6,91-95,31,58,54,55	44-51
MI0000055	periodicity 7(10-55),periodicity 9(35-80)	3-17,59-73,19-33,44-58	18	34-43
MI0000056	periodicity 5(65-95), periodicity 8(10-30), periodicity 9(30-65), periodicity 10(45-80)	4-10,98-104,16-32,77-93,33-47,61-75,48-50,55-57	11-15,94-97,76,58-60	51-54
MI0000057	periodicity 4(35-65), periodicity 5(1-35,80-100)	1-6,90-95,8-12,83-87,13-24,70-81,26-32,63-69,34-41,55-62,43-44,53-54	7,89,88,82,25,33,42	45-52

Accession ID	Peaks in Spectrum	Mfold		
		Stem	Bulge	Loop
MI0000058	periodicity 2(25-55)	1-3,91-93,5-7,88-90,11-16,81-86,17-29,67-79,31,66,33-36,62-65,38-40,59-61,41-42,56-5743-45,52-54	4,8-10,87,80,30,32,37,58,55	46-51
MI0000059	periodicity 4(1-15,85-100), periodicity 5(75-90),	2-5,94-97,6-23,75-92,24-29,68-73,31-38,60-67,39-40,57-58,42-44,54-56	93,74,30,59,41	45-53
MI0000060	periodicity 5(1-5,55-90)	1-5,76-80,7-27,55-75,31-33,39-41,43-44,49-50	6,28-30,42,51-54	45-48,51-54
MI0000061	periodicity 2(1-75)	1-3,70-72,6-26,48-68,36-38,45-47	4-5, 69	27-35, 39-44
MI0000062	periodicity 3(1-25,65-70), periodicity 4(10-40), periodicity 5(25-45), periodicity 8(15-55)	1-25,50-74,26-31,40-35	32-34	41-49
MI0000063	periodicity 3(1-15,65-80), periodicity 4925-40)	1-5,79-83,7-27,58-78,31-36,42-47,49-50,55-56	6,28-30,48,57	37-41,51-54
MI0000065	periodicity 5(0-25,80-90), periodicity (45-65), periodicity 10(20-65)	1-6,82-87,8-29,60-81,33-38,44-49	7,30-32	39-43,50-59
MI000070	periodicity 9(15-60)	1-22,68-84,23-33,56-66,39-42,52-55	67	34-38,43-51
MI0000072	periodicity 7(1-20,50-70) Periodicity 8(30-50)	2-24,48-70,26-28,45-47, 33-34, 41-42	25, 29-32, 43-44,	25-40
MI000079	periodicity 5(1-70)	1-10,64-73, 11-31, 42-62	63	32-41
MI0000086	periodicity 4(1-10,70-90), periodicity 11(1-90)	1-30, 57-86, 33-38, 45-50,	31,32,51, 52	39-44

For periodicities 10 and 11 the consensus sequences has been found which shows the applicability of the MST. The consensus sequences have been found by breaking the sequences at 10 and the 11 positions for both the periodicities respectively. Then, we calculated the nucleotide which was found maximum in the row of sequences. These nucleotide sequences for each row give us the consensus sequence for both the periodicities. The nucleotide sequences for both the periodicities have given in the below table:

Table 3: List of the Consensus sequences found in Period 10 and 11

miRNA	**Consensus pattern for periodicity 10**	**Consensus pattern for periodicity 11**
MI0000045	–	ACAUCCGUCXU
MI0000136	UCACUGUGCU	–
MI0000140	CUAUAUAGUU	CUGAAUUAXUA
MI0000388	–	GUXCUCGUUCU
MI0000393	AAUAXUUCUG	UGUGCAXCGAC
MI0000650	GCCUGGGAGU	CUGAXGXCUGU
MI0000694	UCCUGGGUAU	UXGXGCGGUGG
MI0003135	–	UAXUAUAUCAX
MI0010522	UAUXCUUUUA	–
MI0010523	UAUXAGUUGU	–
MI0013585	–	AGXUAUAUCAX
MI0016284	–	UUGAACUUGAG
MI0016286	UXAUAAACUU	–
MI0017546	UAUUAAAXUU	–
MI0017561	–	AUUAUUAUUAG
MI0017735	CCXAUGCGCG	UXAUCUXGCGG
MI0022215	ACUGGGGGAX	AAXAGCAGUCG

Chapter 5 - Conclusion

The result obtained from the m-fold web server are used it to correlate with the spectrums. We find out that the most of the periodicities which are present in the spectrum are mostly related with the stem structure of the miRNAs. The position which shows the peaks in the spectrum when we validate them with the mfold results we get that these periodicities are mostly related to the stem structure.

We have shown as a proof of concept that MST algorithm can be successfully implemented in for identifying periodicities in RNA sequences and have correlated the periodicities with secondary structure elements. Thus, this signal processing method can be used for predicting the secondary structures of the miRNAs.

Appendix I

```
%This is the code currently in use for Signal processing of Genomic
clc;
clear all;
clf;
tic
x=textread('C:\Users\131505\Documents\MATLAB\MI0000101.dat','%s');
a=char(x);
[r2 c2]=size(a);
c=0;
for i1=1:r2
for j1=1:c2
c=[c a(i1,j1)];
end
end
L_1=sqrt(-1);
%2 prog. for Generating Numeric Code for text-data.....................
D_length=length(c);
for i=1:D_length-1
code=c(i+1);
switch (code)
case 'A'
I(:,i)=[1+L_1]';
case 'G'
I(:,i)=[-1-L_1]';
case 'C'
I(:,i)=[-1+L_1]';
case 'U'
I(:,i)=[1-L_1]';
end
end
%step_3_prog for separate code*********************************************
Ma = I(1,:);
%remove the effect of dc value
MA=(Ma-mean(Ma));
% Compute the ST*****************************************************
Fs=1;
[M,N]=size(Ma);
 N2=fix(N/2); j = 1;
if N2*2==N;j = 0;end
f=[0:N2 -N2+1-j:-1]/N;
%f=[0:N-1];
%f=k2/N;
%x1=zeros(N2+1,N);
SIGa=fft(MA,N);
 gA=15*f+0.8;%1.6*exp(2*f.^3)-1.5*(2*f.^2)+42*f-
0.58%3.2*f+0.89;%5.6*f+0.084;%4.5*f+0.41;%3.2*f+0.89;
% gA=(1/N)+15*var(MA)*f;
%  gA=1*ones(1,length(f));
for i_1=1:1:10
x_1(i_1,:)=N/(i_1+1);%[N/2,N/3,N/4,N/5,N/6,];
end
%*******************************************************
for i1=1:10
    i=floor(x_1(i1));
%i=f(i_1);
SIGa1=circshift(SIGa,[0,-(i-1)]);
WA=(gA(i)/f(i))*2*pi*f;
%GA=sqrt(2*pi)*(f(i)/gA(i))*exp((-WA.^2)/2);
```

```
GA=exp(-(WA.^2)/2);
xA(i1,:)=abs(ifft(SIGa1.*GA));%/max(abs(ifft(SIGa1.*GA)));%   Normalization   performed
added extra operation
n8=abs(i*Fs);
s_d1(i)=(gA(i)*(N/n8))*Fs;
end
%*****************************************************************
x1=abs(xA).^2;%+abs(xG).^2+abs(xC).^2+abs(xT).^2;
x1=abs(x1);
%f=[2:N2]/N;
f=2:11;
p1=0:N-1;
 figure(1)
%t_1=load('F:\MG_WT_TR\Target\synthetic_1.dat')
%plot(p1,t_1,'g')
%axis([0 958 0 1])
%hold on

imagesc(p1,f,abs(x1));
set(gca,'Ydir','Normal'); %default Yaxis in imagesc is inverted
```

Appendix II

This code is for the extraction of the sequences from the Dat file.

```
#!/usr/bin/perl
use strict;
use warnings;

my ($line,$i,$a);
my $inputFile="output.txt";
my $outputFile=0;
my $numberToCopy=0;
#my $sequenceEntry;
my $sequenceTitle="";
my (@ac,@ad);
#Reading the high_conf.dat file
open f1, '<', "high_conf.dat";
#writing in to output.txt
open f2, '>', "output.txt";
while ($line = <f1>)
{     #matching with accession number
   if($line =~ m/^AC/)
     {

       print f2 "$line";
                }
            if($line =~ m/^ /)
{     #spliting with space and then changing in to uppercase
@ac=split(/ /,uc($line));
        print f2 @ac;
         }
}
```

Appendix III

This code takes the sequence file as input and then create a new text file for each sequence and save them with according to their accession id.

```
#!/usr/bin/perl
#open the input file
open (F1, "out.txt");
while (<F1>) {
$line = $_;
chomp $line;
#matching with MI
if ($line =~ /MI/)
{
#return the string which matches with MI
$new_file = substr($line,2);
#file extension
$new_file .= ".dat";
#opening the new file with their accesion id
open (F2, ">$new_file");
}
print F2 "$line\n";
}
close F2;
```

References

1. J. Moura, "What is signal processing? [President's Message", *IEEE Signal Processing Magazine*, vol. 26, no. 6, pp. 6-6, 2009.

2. B. Liu, L. Fang, F. Liu, X. Wang, J. Chen and K. Chou, "Identification of Real MicroRNA Precursors with a Pseudo Structure Status Composition Approach", *PLOS ONE*, vol. 10, no. 3, p. e0121501, 2015.

3. B. Liu, J. Childs-Disney, B. Znosko, D. Wang, M. Fallahi, S. Gallo and M. Disney, "Analysis of secondary structural elements in human microRNA hairpin precursors", *BMC Bioinformatics*, vol. 17, no. 1, 2016.

4. B. Yoon and P. Vaidyanathan, *Computational Identification and Analysis of noncoding RNAs*, 1st ed. IEE Signal Processisng Magzine, 2007, pp. 1-11.

5. P. Svoboda and A. Cara, "Hairpin RNA: a secondary structure of primary importance", *Cellular and Molecular Life Sciences*, vol. 63, no. 7-8, pp. 901-908, 2006.

6. "RNA Structure (Molecular Biology)", *What-when-how.com*, 2016. [Online]. Available: http://what-when-how.com/molecular-biology/rna-structure-molecular-biology/. [Accessed: 16- Dec- 2016].

7. W. Moss, "Computational Prediction of RNA Secondary Structure", Laboratory Methods in Enzymology: RNA, pp. 3-65, 2013.

8. S. Sharma, R. Saxena, S. Sharma and A. Singh, *Short Tandem Repats Detection in DNA Sequences using Modified S-Transform*, 1st ed. International Journal of Advances IN Engineering&Technology, 2015, pp. 1-13.

9. A. Kozomara and S. Griffiths-Jones, "miRBase: annotating high confidence microRNAs using deep sequencing data", *Nucleic Acids Research*, vol. 42, no. 1, pp. D68-D73, 2013.

10. "The Mfold Web Server | mfold.rit.albany.edu", *Unafold.rna.albany.edu*, 2016. [Online]. Available: http://unafold.rna.albany.edu/?q=mfold. [Accessed: 16- Dec- 2016].

11. M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction", *Nucleic Acids Research*, vol. 31, no. 13, pp. 3406-3415, 2003.

12. R. Stockwell, L. Mansinha and R. Lowe, *Localization of the Complex Spectrum: The S Transform*, 1st ed. IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 44, NO. 4, 1996, pp. 1-4.

13. J. Mrazek, "Comparative Analysis of Sequence Periodicity among Prokaryotic Genomes Points to Differences in Nucleoid Structure and a Relationship to Gene Expression", Journal of Bacteriology, vol. 192, no. 14, pp. 3763-3772, 2010.

14. E. Trifonov and J. Sussman, "The pitch of chromatin DNA is reflected in its nucleotide sequence.", Proceedings of the National Academy of Sciences, vol. 77, no. 7, pp. 3816-3820, 1980.

15. Z. Li and T. Rana, "Therapeutic targeting of microRNAs: current status and future challenges", Nature Reviews Drug Discovery, vol. 13, no. 8, pp. 622-638, 2014.