Research Article

Complexity on Acute Myeloid Leukemia mRNA Transcript Variant

Carlo Cattani and Gaetano Pierro

Dipartimento di Matematica, Università di Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy

Correspondence should be addressed to Carlo Cattani, ccattani@unisa.it

This paper deals with the sequence analysis of acute myeloid leukemia mRNA. Six transcript variants of mlf1 mRNA, with more than 2000 bps, are analyzed by focusing on the autocorrelation of each distribution. Through the correlation matrix, some patches and similarities are singled out and commented, with respect to similar distributions. The comparison of Kolmogorov fractal dimension will be also given in order to classify the six variants. The existence of a fractal shape, patterns, and symmetries are discussed as well.

1. Introduction

In some recent papers [1–14], it has been shown that the basic structure of genome is based on fractal geometry. Indeed, the fractal dimension is defined according to the concept of information entropy [15, 16], so that a change in the DNA structure, that is, in the distribution of nucleic acids, implies a corresponding change in the information and, as a consequence, a variation in the entropy. In [2–4, 8–14], it has been suggested that a variation in the entropy can be interpreted as the symptom of malignant evolution of the cell activity, thus being an expedient test for cancer prognosis.

Despite the many still unsolved questions about distribution of base pairs (bps), it is generally understood that the cell activity is functionally dependent on the distribution of nucleotides, that is, the distribution of the 4 symbols A, C, G, and T along the DNA sequence [17, 18]. Some further attempts to understand better about this distribution and about the existence of large-scale structure or hidden rules (eventually fractals) were given in [1, 6, 7, 19–26]. Namely, the large-scale depends on the possibility to show the long range correlation among bps [23–25, 27–43]. The multifractality was also used in connection with the concept of entropy to analyze the complexity of the DNA sequence [1, 2, 26, 41]. The main tasks are

to find (if any) some kind of mathematical rules or meaningful statistics in the nucleotides distribution and to use the deviation from these patterns as a means to detect the existence of malignant evolution.

On the other hand, the existence of patchiness and correlation would imply some important understanding of DNA organization. It has been observed that the source for long-range correlation is linked with existence of patchiness in the DNA sequence. The identification of these patches could be the key point for understanding the large-scale structure of DNA.

Correlation in a DNA sequence is interesting because base pairs in a sequence of millions of pairs seem to have some statistical dependence. The existence of correlation in DNA has been explained with the so-called process of duplication mutation.

The power law for long-range correlations is a measure of the scaling law, showing the existence of self-similar structures similar to the physics of fractals. The long-range correlation, which can be detected by the autocorrelation function, implies the scale independence (scale invariance) which is typical of fractals.

Any statistical analysis on DNA is based on a digitalization of the symbolic sequence, so that one may benefit from the statistical analysis of the digitalized time series, and the genome can be characterized by the classical statistical parameters like variance, deviation, or nonclassical like complexity, fractal dimension, or long-range dependence.

The easiest mathematical model for DNA is based on the transformation of the symbolic string into a numerical string by the Voss indicator function [27, 28] which is a discrete binary function. In the following, a complex representation is proposed in order to single out a fractal law in the cumulative distribution of nucleotides. In some recent papers, the indicator (or correlation) matrix [1, 26] has been proposed as a suitable tool for detecting fractal patterns on the dot-plot representation of the indicator matrix. Then, the computation of fractal dimension and complexity can be easily performed on the dot plot.

In the following, we will compare the fractal dimension and complexity of six mRNA variants of leukemia, showing that these parameters can be used to classify variants. Moreover, the complexity is compared with random and quasi-random sequences (based on the same symbolic alphabet). There follows that the mRNA variants have the complexity close to the random sequence, thus rising some more inquiries on the existence of long-range correlation.

Acute myeloid leukemia consists on the interruption of growth for bone marrow cells at the earliest stages of development. The mechanism of this interruption is under further investigation and still unclear in some aspects; however, it is known that it involves the activation of abnormal genes through chromosomal translocations and other genetic anomalies [42]. If a cell does not reach a mature state, then we can speak about a leukemia that usually has a very abrupt onset and for this reason is called acute leukemia. When, instead, the rate of immature cell with respect to healthy cells is low and this number increases slowly, then we speak about chronic leukemia.

The outcome for adults with AML (acute myeloid leukemia) depends on a variety of factors, including age of the patient and biologic characteristics of the disease, the most important of which are the cytogenetics at presentation [42–44]. The karyotype of the leukemic cells can be roughly classified into 3 groups with either favourable, intermediate, or poor prognostic risk [42, 43].

In the following, we will analyze the sequence of data obtained for homo sapiens mlf1 transcript variant mRNA which has been downloaded from gene bank [45].

2. Preliminary Survey on Leukemia

Blood cells are normally produced by the bone marrow and when mature go into circle. Bone marrow is the principal organ for the production of blood cells. It consists of a complex of cells with high proliferative capacity. The bone includes a portion of adipose tissue that may become prevalent with age (yellow marrow) compared with the hematopoietic component (red marrow). The parenchyma, of the bone marrow, is supported by a stroma composed of irregularly distributed fibroblasts that produce thin beams of reticular fibers. These cells are responsible for the production of growth factors, which activate blood-forming elements. These elements provide a suitable microenvironment for the growth process. The bone marrow is an organ producing every day a population of cells larger than 2.5 billion red blood cells and platelets and 1 billion of white blood cells per kilogram of body weight. It is known that leukemia originates in the bone marrow and from there it spreads into the bloodstream.

Bone marrow can be considered the reservoir from which all blood cells are produced; it contains the precursors of red blood cells, white blood cells (lymphocytes, monocytes, and granulocytes), and platelets, so that blood cells are derived from a single progenitor cell or steam cell. Each blood cell belongs to a different branch starting from the same progenitor cell.

The myeloid progenitor cell lineage for the white blood cells is called granulocyte. The lymphoid progenitor cells give rise to another type of white blood cell. The progenitor for erythroid lineage produces red blood cells, and finally a megakaryocytic precursor gives rise to platelets.

Unlike other cells of the body, which rarely duplicate, the bone marrow cells are characterized by higher proliferative capacity, so that the blood constantly contains a large number of red blood cells, white blood cells, and platelets, which are quickly renewed at different rate. It is clear that the probability of the formation of a malignant tumor is roughly proportional to the number of cell divisions, so that it is possible that almost any cell, in the bone marrow, becomes malignant and gives rise to this type of cancer called leukemia. The different speed of diffusion enables us to classify leukemia into acute (fast growth) or chronic (slow growth) leukemia. In the following, we will focus only on the acute myeloid leukemia.

The underlying pathophysiology in acute myeloid leukemia consists on the interruption of growth for bone marrow cells at the earliest stages of development. According to the rate of growing, we can have either acute or chronic leukemia. The main characteristics of both kind of leukemia are different upon the cell strain from which is originated the disease, the most common being leukemia myeloid and lymphoid. In general, we have four main types:

(1) acute myeloid leukemia that strikes mostly in old people and adults,

(2) chronic myeloid leukemia that is characterized by a highly specific chromosomal abnormality known as Philadelphia chromosome,

(3) chronic lymphocytic leukemia and acute lymphocytic leukemia with a high incidence in children between 2 and 5 years of age.

Usually, the clinical symptoms of leukemia are initially underestimated and they can be easily confused with symptoms of minor diseases. This is due to the fact that initially leukemia cells are quickly replaced by new cells produced by bone marrow. Only when the growth speed of the number of leukemia cells increases rapidly, the symptoms become more easily detectable.

Table 1: The eight FAB subtypes (t, translocation; del, deletion).

	Name	Cytogenetics
M0	Minimally differentiated acute myeloblastic leukemia	
M1	Acute myeloblastic leukemia, without maturation	
M2	Acute myeloblastic leukemia, with granulocytic maturation	t(8;21)(q22;q22), t(6;9)
M3	Promyelocytic, or acute promyelocytic leukemia (APL)	t(15;17)
M4	Acute myelomonocytic leukemia	inv(16)(p13q22), del(16q)
M4eo	Myelomonocytic together with bone marrow eosinophilia	inv(16), t(16;16)
M5	Acute monoblastic leukemia (M5a) or acute monocytic leukemia (M5b)	del(11q), t(9;11), t(11;19)
M6	Acute erythroid leukemias, including erythroleukemia (M6a) and very rare pure erythroid leukemia (M6b)	
M7	Acute megakaryoblastic leukemia	t(1;22)

Chr 3
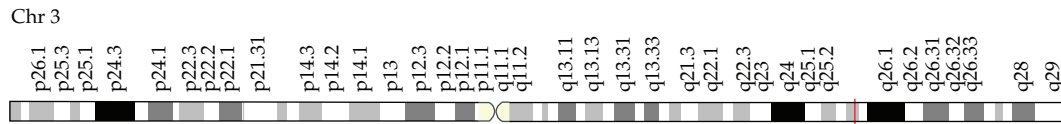
Figure 1: MLF1 in genomic location.

The most common symptoms are bleeding or bruising on the skin or mouth, due to lack of platelets or thrombocytopenia, neutropenia, due to a lack of white blood cells, or simply the pale and weakness, due to anemia. Namely, symptoms of leukemia and production rate of cells by marrow depend on the number of blasts.

The most popular classification of leukemia is the French-American-British (FAB) system that classifies AML into 8 subtypes, from M0 to M7, (see Table 1) based both on the type of cell from which the leukemia is developed and its degree of maturity. The classification is realized by analysing the appearance of the malignant cells under light microscopy and/or by using cytogenetics to characterize any underlying chromosomal anomalies. According to subtypes membership, leukemia owns different prognoses and responses to therapy.

The malignant cell in AML usually appears at the myeloblast level. During a normal hematopoiesis activity, the myeloblast is the immature precursor of myeloid white blood cells; a normal myeloblast usually gradually evolves into a mature white blood cell. Whereas, in AML, a single myeloblast accumulates some genetic changes which blocks the cell in its immature state and prevent differentiation. As seen in Table 1, there are six major features which have been recently added by two more systematic: MO and M7 megakaryoblastic.

In the following, we will focus on the MLF1 gene (myeloid leukemia factor 1). This gene MLF1 [45] encodes an oncoprotein which is thought to play a fundamental role in the phenotypic determination of hemopoetic cells. Translocations between this gene and nucleophosmin have been associated with myelodysplastic syndrome and acute myeloid leukemia. Multiple transcript variants encoding different isoforms have been found for this gene. In Figure 1 is represented the location of MLF1 at the genomic level [46, 47].
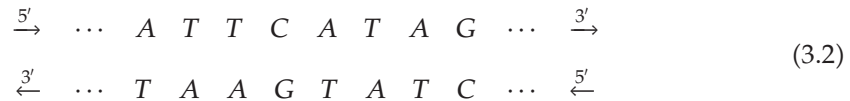
3. DNA Representation

3.1. Preliminary Remarks on DNA

The DNA, as well as the mRNA, of each organism of a given species is a sequence of a specific number of base pairs defined on the 4 elements alphabet of nucleotides:

$$A = \text{adenine}, \qquad C = \text{cytosine}, \qquad G = \text{guanine}, \qquad T = \text{thymine}. \qquad (3.1)$$

Since the base pairs are distributed along a double helix, when straightened, the helix appears as a double-strand system,

$$
\begin{aligned}
&\xrightarrow{5'} \quad \cdots \quad A \quad T \quad T \quad C \quad A \quad T \quad A \quad G \quad \cdots \quad \xrightarrow{3'} \\
&\xleftarrow{3'} \quad \cdots \quad T \quad A \quad A \quad G \quad T \quad A \quad T \quad C \quad \cdots \quad \xleftarrow{5'}
\end{aligned}
\qquad (3.2)
$$

The two sequences on opposite strands are complementary in the sense that opposite nucleotides must fulfill the ligand rules of base pairs, between purines A and G and pyrimidines T and C,

$$
\begin{array}{c|c|c}
\text{purines} & & \text{pyrimidines} \\
\hline
A & \longleftrightarrow & T \\
G & \longleftrightarrow & C
\end{array}
\qquad (3.3)
$$

In a DNA sequence, there are some subsequences, coding and noncoding regions, having special meaning. In particular, genes (coding regions) are characteristic sequences of base pairs, and the genes in turn are made by some alternating subsequences of exons and introns (except Procaryotes where the introns are missing),

$$
ATCG \cdots \overbrace{AT \cdots GAG}^{\text{gene}} \underbrace{}_{\text{...exon...intron...exon...}} \cdots TCG \cdots \overbrace{CA \cdots TAG}^{\text{gene}} \underbrace{}_{\text{...exon...intron...exon dots}} \cdots \qquad (3.4)
$$

After the transcription, each exon region is made of triplets of adjacent bases called codon. Since the bases are 4, there are 64 possible codons. Each codon synthesizes a specific amino acid in the translation process, so that a sequence of codons defines a protein. There are only 20 amino acids; therefore, the correspondence codons to amino acids are many to one. The exons region is also called the coding region.

3.2. Indicator Function in a 4-Symbol Alphabet

In this section, we will define the indicator matrix [1, 26] which can be used to visualize the existence of some patterns in the nucleotide distribution.

Let

$$\aleph_4 = \{A, C, G, T\} \qquad (3.5)$$

be the finite set (alphabet) of nucleotides and $x \in \aleph_4$ any member of the 4-symbol alphabet.

A DNA sequence is the finite symbolic sequence

$$\mathfrak{D}(N) = \mathbb{N} \times \aleph_4, \tag{3.6}$$

so that

$$\mathfrak{D}(N) = \{x_h\}_{h=1,\dots,N}, \quad N < \infty, \tag{3.7}$$

being

$$x_h = (h, x) = x(h), \quad (h = 1, 2, \dots, N; x \in \aleph_4) \tag{3.8}$$

the acid nucleic x at the position h.

The indicator function [1, 26] is the map

$$u : \mathfrak{D}(N) \times \mathfrak{D}(N) \longrightarrow \{0, 1\}, \tag{3.9}$$

such that,

$$u_{hk} = u(x_h, x_k) = \begin{cases} 1 & \text{if } x_h = x_k \\ 0 & \text{if } x_h \neq x_k \end{cases} \quad (x_h \in \mathfrak{D}(N), x_k \in \mathfrak{D}(N)). \tag{3.10}$$

According to (3.9), the indicator map of the N-length sequence can be easily represented by the $N \times N$ sparse matrix of binary values $\{0, 1\}$, and this matrix can be visualized by the dot plot obtained by putting a black dot where $u_{hk} = 1$ and white spot when $u_{hk} = 0$ (see correlation matrix (3.11) for the 4-symbol alphabet)

$$
\begin{array}{c|cccccccccc}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
G & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\
C & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & \dots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & \dots \\
T & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \dots \\
C & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \\
T & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\
G & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots \\
A & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \dots \\
\hline
u_{hk} & A & G & T & C & A & T & A & A & C & G & \dots
\end{array}
\tag{3.11}
$$

Indeed, the definition (3.9) is expedient to study the autocorrelation of each sequence. In general for two sequences $\mathfrak{D}_1(N), \mathfrak{D}_2(N)$, definition (3.9) can be extended to

$$u : \mathfrak{D}_1(N) \times \mathfrak{D}_2(N) \longrightarrow \{0, 1\}, \tag{3.12}$$

with

$$u_{hk} = u(x_h, x_k) = \begin{cases} 1 & \text{if } x_h = x_k \\ 0 & \text{if } x_h \neq x_k \end{cases} \quad (x_h \in \mathfrak{D}_1(N),\ x_k \in \mathfrak{D}_2(N)). \tag{3.13}$$

In order to understand the acid nucleic distribution in the 6 variants of AML mRNA, we will compare them with some artificial sequence based on the same alphabet $\aleph_4$. In particular, we will consider the N-length *random sequence*

$$\mathfrak{R}(N) = \mathbb{N} \times \aleph_4, \tag{3.14}$$

so that

$$\mathfrak{R}(N) = \{x_h\}_{h=1,\dots,N}, \quad N < \infty, \tag{3.15}$$

being

$$x_h = (h, x) = x(h), \quad (h = 1, 2, \dots, N; x \in \aleph_4) \tag{3.16}$$

the acid nucleic x at the position h, randomly chosen.

The *periodic sequence* $\pi(N, p)$ is defined by repeating the same p-length random sequence $\pi(p)$, so that

$$\pi(N, p) = \oplus_{i=1,\dots,N} \pi(p),$$
$$\pi(N, p) = \{x_h\}_{h=1,\dots,N}, \quad p < N < \infty, \tag{3.17}$$

with

$$x(h + i) = x(i), \quad (h = 1, 2, \dots, N; i = 1, \dots, p; x \in \aleph_4). \tag{3.18}$$

The *quasiperiodic sequence* is obtained by alternating n-length periodic sequence of period p with random r-length sequences

$$\mathfrak{P}(N, n, p, r) = \oplus_{k=1,\dots,N/k} [\pi(n, p) \oplus \mathfrak{R}(r)]_k. \tag{3.19}$$

For example, it is

$$\mathfrak{P}(24, 8, 4, 4) = [\pi(8, 4) \oplus \mathfrak{R}(4)]_1 \oplus [\pi(8, 4) \oplus \mathfrak{R}(4)]_2$$

$$= \left\{ \underbrace{\underbrace{G, C, C, G,}_{\pi(4)}\ \underbrace{G, C, C, G,}_{\pi(4)}}_{\pi(8,4)}\ \underbrace{C, G, T, C,}_{[\mathfrak{R}(4)]_1}\ \underbrace{G, C, C, G, G, C, C, G,}_{\pi(8,4)}\ \underbrace{T, T, C, T}_{[\mathfrak{R}(4)]_2} \right\}. \tag{3.20}$$
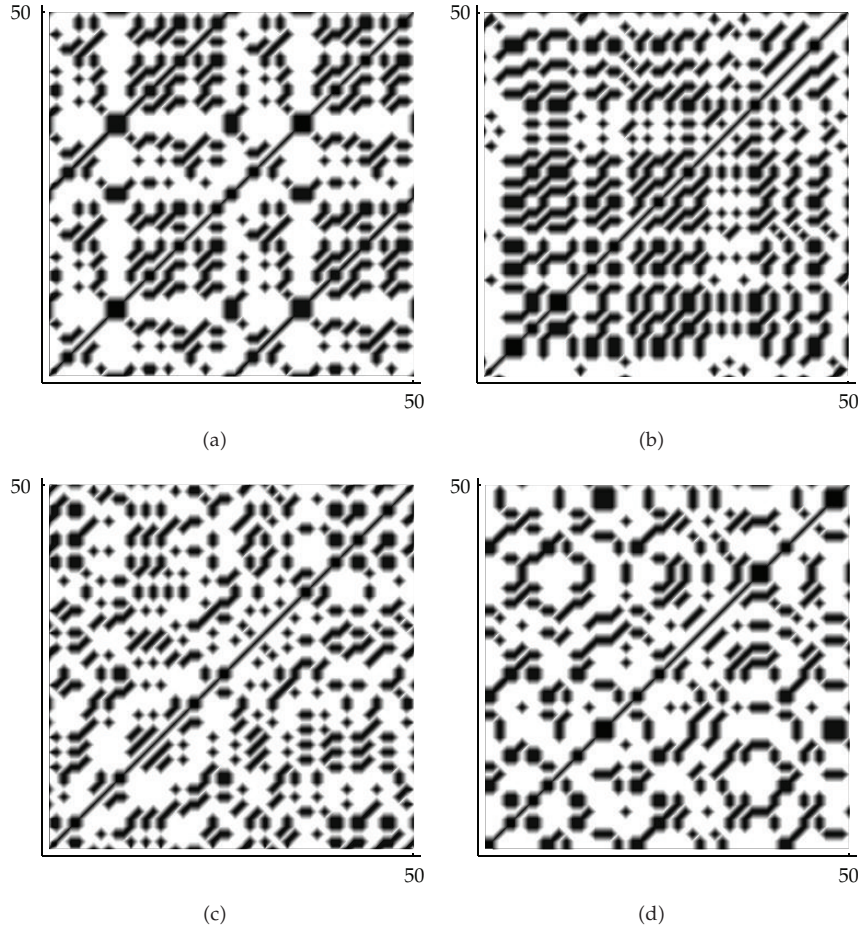
(a)

(b)

(c)

(d)

Figure 2: Dot plot on a 50-length chain for the (a) periodic sequence $\pi(50, 25)$; (b) mRNA variant H1; (c) full random sequence $\mathfrak{R}(50)$; (d) quasiperiodic sequence $\mathfrak{P}(50, 4, 4, 16)$.

If we compare the dot plot of the AML mRNA with the artificial sequences, we can see (Figure 2) that mRNA is much more alike the random or quasirandom sequence.

Although the mRNA sequence cannot be easily recognized by the dot plot, we can better characterize the distribution of nucleic acids by computing some parameters which are strictly related to the complexity. These parameters are based on the computation of "1" in the indicator matrix (see following section).

3.3. Fractal Estimate by the Indicator Matrix

From the indicator matrix, we can have an idea of the "fractal-like" distribution of nucleotides as follows: let
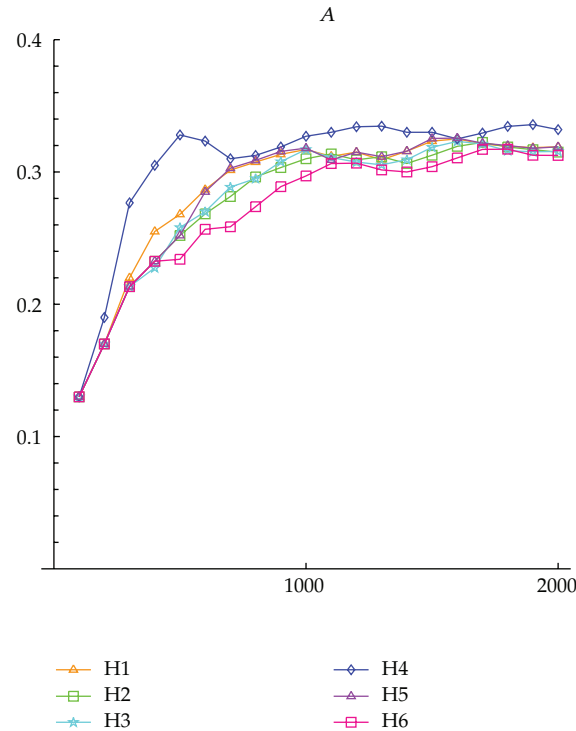
$$p_X(\ell), \quad X \in \{A, C, G, T\} \tag{3.21}$$

Figure 3: Probability density distribution of *A*.

be the probability that the acid nucleic *X* can be found at the position ℓ. This value can be approximated by the frequency count. So that for the transcript variant, we have the probability density distribution of Figure 3.

It can be seen from Figures 3, 4, 5, and 6 that, for higher values of ℓ, the probabilities tend to assume some constant values, thus showing that nucleotides are heterogeneously distributed. However, there are some significant differences among the variants; H4 shows a higher distribution of *A* while H6 is a minor content of *A*.

The frequency distribution implies a corresponding frequency of correlation in the correlation matrix. By using the indicator matrix, it is possible to give a simple formula which enables us to estimate the fractal dimension as the average of the number $p(n)$ of 1 in the randomly taken $n \times n$ minors of the $N \times N$ correlation matrix u_{hk}

$$D = \frac{1}{N} \sum_{n=2}^{N} \frac{\log p(n)}{\log n}. \tag{3.22}$$

If we compare the fractal dimension of DNA with the random sequence and the periodic sequence, we can see (Figure 7) that the mRNA dimension is closer to the dimension of random sequence.

Moreover, the fractal dimension of the indicator matrix can be used to characterize the different distribution of acid nucleic in the mRNA 6 variants. In fact, it can be seen that some variants have lower values of the dimension (Figure 8), like H5 (and for a short interval H2).
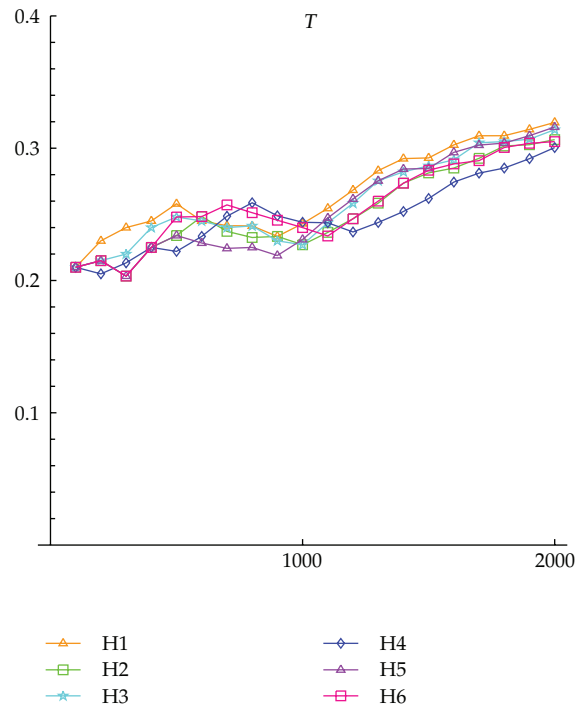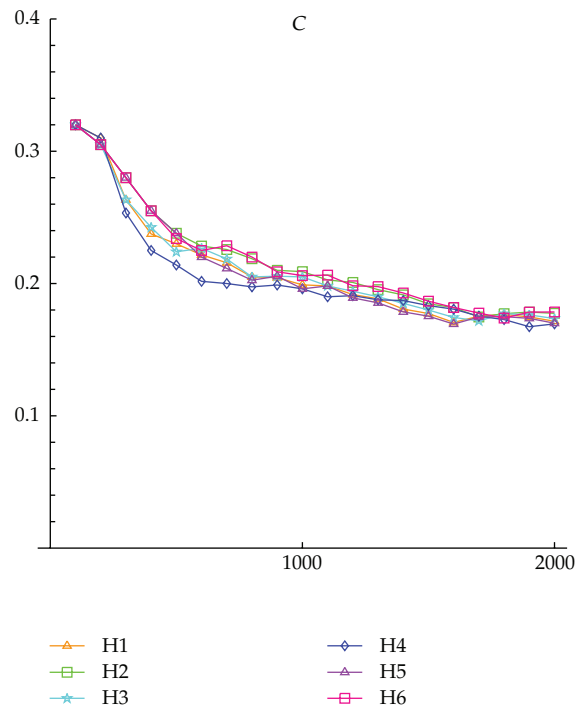
Figure 4: Probability density distribution of *T*.

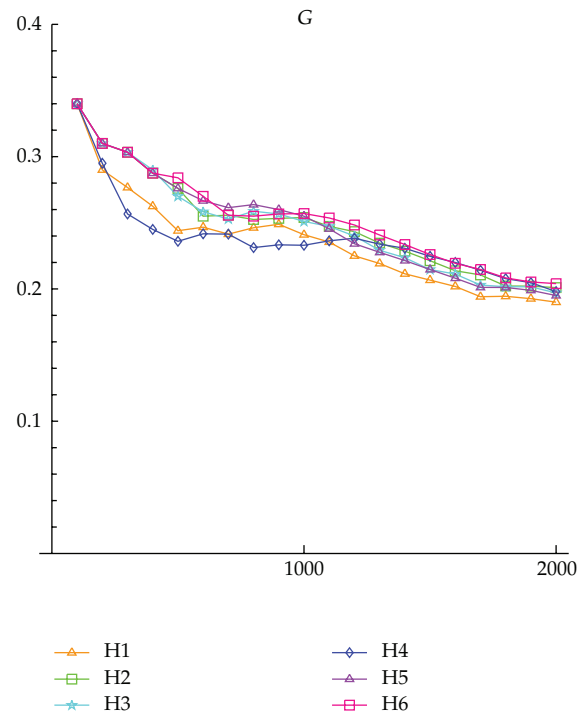**Figure 5:** Probability density distribution of *C*.

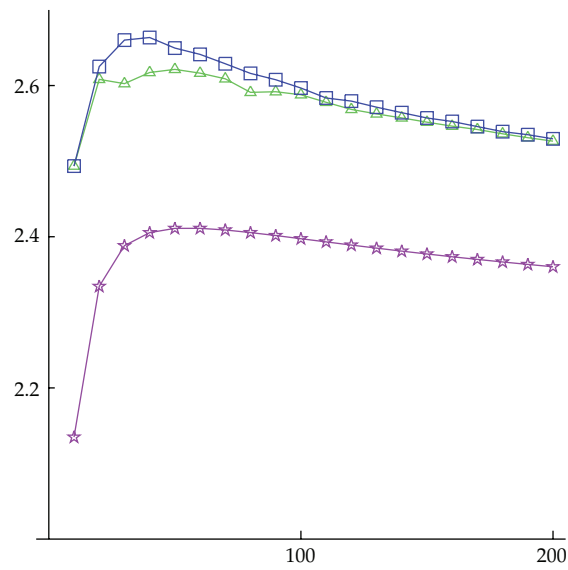**Figure 6:** Probability density distribution of *G*.

Figure 7: Fractal dimension as function of the length, for random $\mathfrak{R}(i)$, $i = 10, \ldots, 200$ (blue), H1 mRNA (green), periodic sequence $\pi(i, 8)$, $i = 10, \ldots, 200$ (purple).
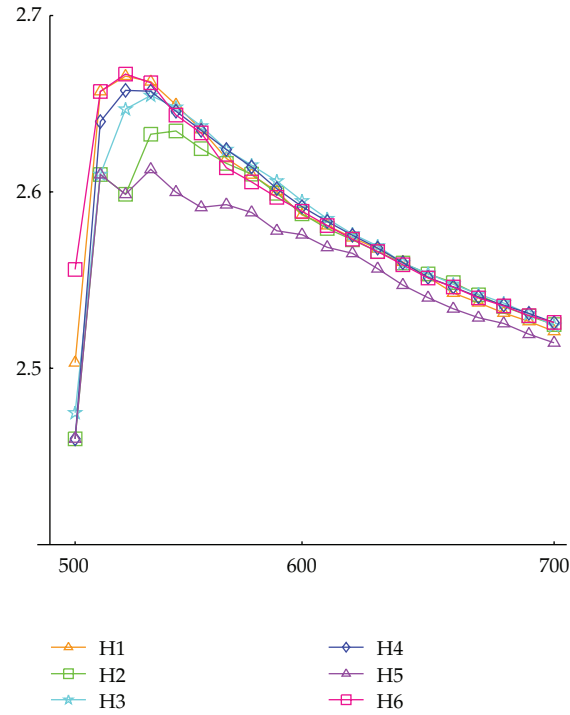
Figure 8: Fractal dimension as function of the length, from 500 bps to 700 bps, for the 6 variants of mRNA.

3.4. Complexity

The existence of repeating motifs, periodicity, and patchiness can be considered as a simple behavior of sequence, while nonrepetitiveness or singularity is taken as a characteristic feature of complexity. In order to have a measure of complexity, for an n-length sequence, we use the definition

$$K = \log \Omega^{1/n}, \tag{3.23}$$

with

$$\Omega = \frac{n!}{a_n! c_n! g_n! t_n!},$$

$$a_n = \sum_{h=1,\dots,n} u(A, x_h), \qquad c_n = \sum_{h=1,\dots,n} u(C, x_h), \qquad g_n = \sum_{h=1,\dots,n} u(G, x_h), \qquad t_n = \sum_{h=1,\dots,n} u(T, x_h). \tag{3.24}$$

By using a sliding n-length window over the full DNA sequence, one can visualize the distribution of complexity on partial fragment of the sequence.

It can be seen (Figure 9) that the complexity line of the mRNA-H1 sequence is bounded by the upper line of random data and the lower line of periodic sequence $\pi(4)$.

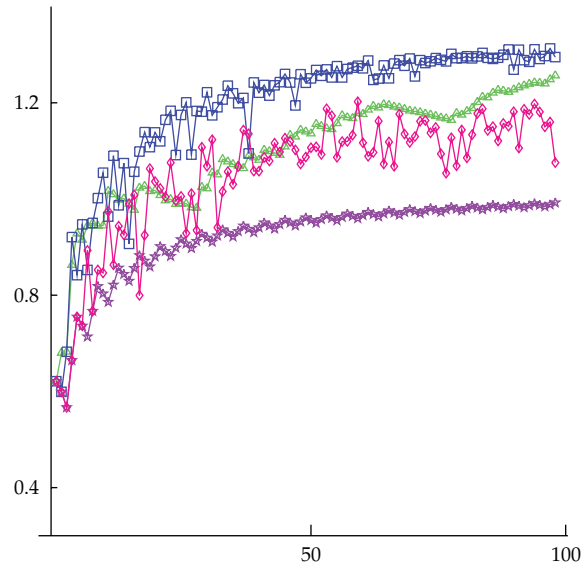**Figure 9:** Complexity of the 100-length H1 sequence (green), full random $\mathfrak{R}(i)$, $i = 4,\ldots,100$ (blue), periodic $\pi(i,4)$, $i = 4,\ldots,100$ (purple), and quasiperiodic $\mathfrak{P}(i,8,4,3)$, $i = 4,\ldots,100$ (red).

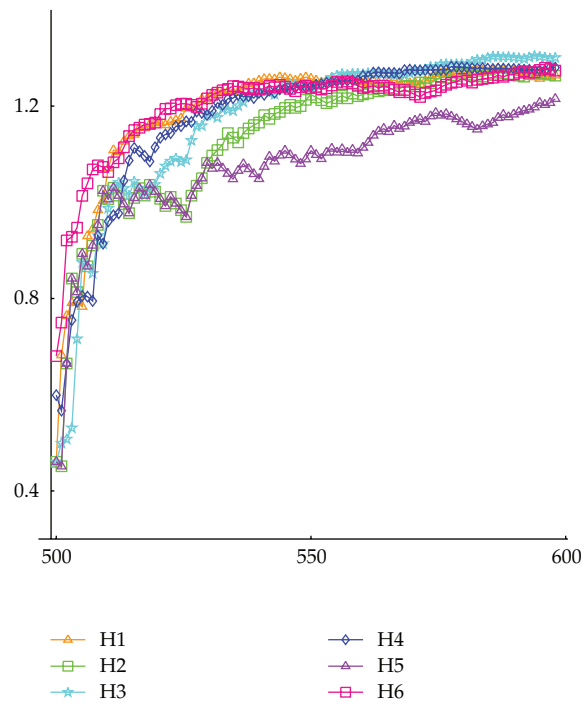|	H1		H4
	H2		H5
	H3		H6

Figure 10: Complexity of the 100 length from 500 bps to 600 bps for the 6 variants of mRNA.

The quasiperiodic $\pi(8,4) \oplus r(3)$, instead, shows high oscillations between the two bonds. It should be also noticed that the complexity of DNA tends to the asymptotic value of the random sequence.

The explicit computation of the complexity line for the remaining mRNA variants (Figure 10) shows some different behavior, and for this reason, it can be used as a parameter to characterize the differences among variants. Like for the fractal dimension, H5 shows some lower values of complexity.

4. Conclusions

In this paper, six variants of acute myeloid leukemia mlf1 mRNA have been analyzed through the correlation matrix. In particular, some parameters like fractal dimension and complexity have been computed and compared. It has been shown that some variants have many similarities, and practically they can be considered as belonging to the same class of mRNA. Some others instead have a very typical distribution of bps very different from the remaining variants. Some variants look like pseudorandom sequence.

References

[1] C. Cattani, "Wavelet algorithms for DNA analysis," in *Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications*, M. Elloumi and A. Y. Zomaya, Eds., Wiley Series in Bioinformatics, chapter 35, pp. 799–842, John Wiley & Sons, New York, NY, USA, 2010.

[2] K. Metze, "Fractal dimension of chromatin and cancer prognosis," *Epigenomics*, vol. 2, no. 5, pp. 601–604, 2010.

[3] R. L. Adam, R. C. Silva, F. G. Pereira, N. J. Leite, I. Lorand-Metze, and K. Metze, "The fractal dimension of nuclear chromatin as a prognostic factor in acute precursor B lymphoblastic leukemia," *Cellular Oncology*, vol. 28, no. 1-2, pp. 55–59, 2006.

[4] K. Metze, I. Lorand-Metze, N. J. Leite, and R. L. Adam, "Goodness-of-fit of the fractal dimension as a prognostic factor," *Cellular Oncology*, vol. 31, no. 6, pp. 503–504, 2009.

[5] D. V. Lebedev, M. V. Filatov, A. I. Kuklin et al., "Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties," *FEBS Letters*, vol. 579, no. 6, pp. 1465–1468, 2005.

[6] J. G. McNally and D. Mazza, "Fractal geometry in the nucleus," *The EMBO Journal*, vol. 29, no. 1, pp. 2–3, 2010.

[7] M. Takahashi, "A fractal model of chromosomes and chromosomal DNA replication," *Journal of Theoretical Biology*, vol. 141, no. 1, pp. 117–136, 1989.

[8] A. Delides, I. Panayiotides, A. Alegakis et al., "Fractal dimension as a prognostic factor for laryngeal carcinoma," *Anticancer Research*, vol. 25, no. 3 B, pp. 2141–2144, 2005.

[9] R. C. Ferreira, P. S. De Matos, R. L. Adam, N. J. Leite, and K. Metze, "Application of the Minkowski-Bouligand fractal dimension for the differential diagnosis of thyroid follicular neoplasias," *Cellular Oncology*, vol. 28, no. 5-6, pp. 331–333, 2006.

[10] M. R. B. Mello, K. Metze, R. L. Adam et al., "Phenotypic subtypes of acute lymphoblastic leukemia associated with different nuclear chromatin texture," *Analytical and Quantitative Cytology and Histology*, vol. 30, no. 2, pp. 92–98, 2008.

[11] L. Goutzanis, N. Papadogeorgakis, P. M. Pavlopoulos et al., "Nuclear fractal dimension as a prognostic factor in oral squamous cell carcinoma," *Oral Oncology*, vol. 44, no. 4, pp. 345–353, 2008.

[12] A. Mashiah, O. Wolach, J. Sandbank, O. Uziel, P. Raanani, and M. Lahav, "Lymphoma and leukemia cells possess fractal dimensions that correlate with their biological features," *Acta Haematologica*, vol. 119, no. 3, pp. 142–150, 2008.

[13] K. Metze, D. P. Ferro, M. A. Falconi et al., "Fractal characteristics of nuclear chromatin in routinely stained cytology are independent prognostic factors in patients with multiple myeloma," *Virchows Archiv*, vol. 445, supplement 1, pp. 7–21, 2009.

[14] V. Bedin, R. L. Adam, B. C. S. de Sá, G. Landman, and K. Metze, "Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma," *BMC Cancer*, pp. 260–265, 2010.

[15] L. Pontrjagin and L. Schnirelmann, "Sur une propriété métrique de la dimension," *Annals of Mathematics*, vol. 33, pp. 156–162, 1932.

[16] A. N. Kolmogorov and V. M. Tihomiroff, "ε-entropy and ε-capacity of sets in functional spaces," *Uspehi Matematicheskih Nauk*, vol. 14, no. 2, pp. 3–86, 1961.

[17] J. P. Fitch and B. Sokhansanj, "Genomic engineering: moving beyond DNA sequence to function," *Proceedings of the IEEE*, vol. 88, no. 12, pp. 1949–1971, 2000.

[18] H. Gee, "A journey into the genome: what's there," *Nature*, 2001.

[19] P. D. Cristea, "Large scale features in DNA genomic signals," *Signal Processing*, vol. 83, no. 4, pp. 871–888, 2003.

[20] K. B. Murray, D. Gorse, and J. M. Thornton, "Wavelet transforms for the characterization and detection of repeating motifs," *Journal of Molecular Biology*, vol. 316, no. 2, pp. 341–363, 2002.

[21] A. Arneodo, Y. D'Aubenton-Carafa, E. Bacry, P. V. Graves, J. F. Muzy, and C. Thermes, "Wavelet based fractal analysis of DNA sequences," *Physica D*, vol. 96, no. 1–4, pp. 291–320, 1996.

[22] B. Borstnik, D. Pumpernik, and D. Lukman, "Analysis of apparent spectrum in DNA sequences," *Europhysics Letters*, vol. 23, pp. 389–394, 1993.

[23] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides," *Physical Review E*, vol. 49, no. 2, pp. 1685–1689, 1994.

[24] S. Karlin and V. Brendel, "Patchiness and correlations in DNA sequences," *Science*, vol. 259, no. 5095, pp. 677–680, 1993.

[25] E. Schrödinger, *What is Life? Physical Aspects of Living Cell*, Cambridge University Press, Cambridge, UK, 1948.

[26] C. Cattani, "Fractals and hidden symmetries in DNA," *Mathematical Problems in Engineering*, vol. 2010, Article ID 507056, 31 pages, 2010.

[27] R. F. Voss, "Evolution of long-range fractal correlations and $1/f$ noise in DNA base sequences," *Physical Review Letters*, vol. 68, no. 25, pp. 3805–3808, 1992.

[28] R. F. Voss, "Long-range fractal correlations in DNA introns and exons," *Fractals*, vol. 2, pp. 1–6, 1992.

[29] M. Zhang, "Exploratory analysis of long genomic DNA sequences using the wavelet transform: examples using polyomavirus genomes," in *Genome Sequencing and Analysis Conference VI*, pp. 72–85, Hilton Head, NC, USA, 1995.

[30] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, "Characterizing long-range correlations in DNA sequences from wavelet analysis," *Physical Review Letters*, vol. 74, no. 16, pp. 3293–3296, 1995.

[31] B. Audit, C. Vaillant, A. Arneodo, Y. D'Aubenton-Carafa, and C. Thermes, "Long-range correlations between DNA bending sites: relation to the structure and dynamics of nucleosomes," *Journal of Molecular Biology*, vol. 316, no. 4, pp. 903–918, 2002.

[32] S. V. Buldyrev, A. L. Goldberger, A. L. Havlin et al., "Long-range fractal correlations in DNA," *Physical Review E*, vol. 51, no. 5, pp. 5084–5091, 1995.

[33] H. Herzel, E. N. Trifonov, O. Weiss, and I. Große, "Interpreting correlations in biosequences," *Physica A*, vol. 249, no. 1–4, pp. 449–459, 1998.

[34] W. Li, "The study of correlation structures of DNA sequences: a critical review," *Computers and Chemistry*, vol. 21, no. 4, pp. 257–271, 1997.

[35] W. Li and K. Kaneko, "Long-range correlations and partial spectrum in a noncoding DNA sequence," *Europhysics Letters*, vol. 17, pp. 655–660, 1992.

[36] C. K. Peng, S. V. Buldyrev, A. L. Goldberger et al., "Long-range correlations in nucleotide sequences," *Nature*, vol. 356, no. 6365, pp. 168–170, 1992.

[37] O. Weiss and H. Herzel, "Correlations in protein sequences and property codes," *Journal of Theoretical Biology*, vol. 190, no. 4, pp. 341–353, 1998.

[38] Z. G. Yu, V. V. Anh, and B. Wang, "Correlation property of length sequences based on global structure of the complete genome," *Physical Review E*, vol. 63, no. 1, Article ID 011903, 8 pages, 2001.

[39] P. P. Vaidyanathan and B. J. Yoon, "The role of signal-processing concepts in genomics and proteomics," *Journal of the Franklin Institute*, vol. 341, no. 1-2, pp. 111–135, 2004.

[40] P. Bernaola-Galván, R. Román-Roldán, and J. L. Oliver, "Compositional segmentation and long-range fractal correlations in DNA sequences," *Physical Review E*, vol. 53, no. 5, pp. 5181–5189, 1996.

[41] W. Li, "The complexity of DNA: the measure of compositional heterogenity in DNA sequence and measures of complexity," *Complexity*, vol. 3, pp. 33–37, 1997.

[42] J. M. Bennett, M. L. Young, J. W. Andersen et al., "Long-term survival in acute myeloid leukemia: the Eastern Cooperative Oncology Group experience," *Cancer*, vol. 80, no. 11, pp. 2205–2209, 1997.

[43] D. Grimwade, H. Walker, F. Oliver et al., "The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial," *Blood*, vol. 92, no. 7, pp. 2322–2333, 1998.

[44] M. L. Slovak, K. J. Kopecky, P. A. Cassileth et al., "Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest oncology group/Eastern cooperative oncology group study," *Blood*, vol. 96, no. 13, pp. 4075–4083, 2000.

[45] National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/.

[46] "Gene Location," Weizmann Institute of Science, http://genecards.weizmann.ac.il/geneloc/index.shtml.

[47] "e!Ensemble, Ensembl project," EMBL-EBI & Wellcome Trust Sanger Institute, http://www.ensemble.org/.

Advances in
Operations Research

Advances in
Decision Sciences

Mathematical Problems
in Engineering

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Combinatorics

International Journal of
Differential Equations

Advances in
Mathematical Physics

Submit your manuscripts at
http://www.hindawi.com

Hindawi

Journal of
Complex Analysis

Journal of
Mathematics

International Journal of
Stochastic Analysis

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

Journal of
Applied Mathematics

Journal of
Optimization