Abstract: This paper is aimed to predict the fractal natures of genes of two different cancers in Homo sapiens. The cancers whose properties were studied include the cancer of Lungs and Skin. Two important genes that are responsible for each of these cancers were acquired to analyze their fractal nature. Wavelet transformation had been extensively used in the process of determining the fractal nature. The mRNA sequences were collected from the NCBI database and were then pre-processed efficiently. DNA walk plots were generated for each of these sequences using an in-house program written in MATLAB language. The data so obtained was checked for fractal property using wavelet transformations using an in-house program written in Matlab. All the sequences were checked for quadratic variations with the use of Daubechies-2(DB2) wavelets. The parameters were then plotted and the nature of each cancer was obtained from the plots. We have found that the gene sequences are exhibiting multifractal nature.

Keywords : Fractal, Matlab, DNA Walk-plot

INTRODUCTION

Cancer is a class of diseases in which a group of cells display uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood). Cancer caused about 13% of all human deaths in 2007 (7.6 million).

Cancers are caused by abnormalities in the genetic material of the transformed cells. These abnormalities may be due to the effects of carcinogens, such as tobacco smoke, radiation, chemicals, or infectious agents. Other cancer-promoting genetic abnormalities may randomly occur through errors in DNA replication, or are inherited, and thus present in all cells from birth. The heritability of cancers is usually affected by complex interactions between carcinogens and the host’s genome.

Many natural objects, including most objects studied in pathology, have complex structural characteristics and the complexity of their structures, for example the degree of branching of vessels or the irregularity of a tumor boundary, remains at a constant level over a wide range of magnifications. These structures also have patterns that repeat themselves at different magnifications.
property known as scaling self-similarity. This has important implications for measurement of parameters such as length and area, since Euclidean measurements of these may be invalid. The fractal system of geometry overcomes the limitations of the Euclidean geometry for such objects and measurement of the fractal dimension gives an index of their space-filling properties. The fractal dimension may be measured using image analysis systems and the box-counting, divider (perimeter-stepping) and pixel dilation methods have all been described in the published literature. Fractal analysis has found applications in the detection of coding regions in DNA and measurement of the space-filling properties of tumors, blood vessels and neurones. Fractal concepts have also been usefully incorporated into models of biological processes, including epithelial cell growth, blood vessel growth, periodontal disease and viral infections. (Cross SS, 1997)

Autonomous and uncoordinated proliferation of epithelia leads to various well-known growth patterns such as expansive (cauliflower-like), radiating infiltrative, polycyclic, and roundish-ovoid figures. All attempts to describe such natural growth patterns graphically by Euclidean geometry have failed and remain no more than works of art. However, fractal geometry is a new tool for the characterization of irregularly-shaped and complex figures. Moreover, behind a fractal structure there is a basic power-law which provides the opportunity to simulate these forms artificially. A prerequisite for achieving this goal of simulating tumour growth by computer is to establish whether typical tumour growth patterns are fractal. Hence, an investigation was undertaken of 20 tumors (malignant, metastases or benign) exhibiting the above-mentioned typical patterns. If tumor outlines are fractal they have to possess a fractal non-integer dimension which significantly exceeds the integer Euclidean dimension. The fractal dimension of tumor outlines was determined using the box-counting method. Almost all tumors presented a fractal dimension and virtual tumor images were created by utilizing available fractal software. In conclusion, the determination of the fractal dimension of solid neoplasms may be an additional morphometric parameter for growth assessment and it probably provides further opportunity to simulate cancer growth and infiltration by computer animation. (Sedivy R, 1996)

Fractal is an object or quantity that displays self-similarity, in a somewhat technical sense, on all scales. The object need not exhibit exactly the same structure at all scales, but the same "type" of structures must appear on all scales.

• It has a fine structure at arbitrarily small scales.
• It is too irregular to be easily described in traditional Euclidean geometry language.
• It is also self-similar (at least approximately or stochastically).
• It has a simple and recursive definition.

Fractal dimension is a measure of how "complicated" a self-similar figure is. It is a number that quantitatively describes how an object fills its space. The fractal dimension, D, is a statistical quantity that gives an indication of how completely a fractal appears to fill space, as one zooms down to finer and finer scales.

The fractal dimension measures are derived from fractals which are formally-defined. However, organisms and real-world phenomena exhibit fractal properties (see Fractals in nature), so it can often be useful to characterize the fractal dimension of a set of sampled data. The fractal dimension measures cannot be derived exactly but must be estimated. This is used in a variety of research areas including physics, image analysis, acoustics, Riemann zeta zeros and even (electro) chemical processes.

Multi Fractal System:

A multifractal system is a generalization of a fractal system in which a single exponent (the fractal dimension) is not enough to describe its dynamics; instead, a continuous spectrum of exponents (the so-called singularity spectrum) is needed.

Multifractal systems are common in nature, especially geophysics. They include fully developed turbulence, stock market time series, real world scenes, the Sun’s magnetic field time series, heartbeat dynamics, human gait, and natural
luminosity time series. Embryogenesis is also multifractal system which represents a new type of physics, named fractal mechanics. In a multifractal system $s$, the behavior around any point is described by a local power law:

$$s(x + \delta) - s(\bar{x}) \sim a^{h(x)}.$$

The exponent is called the singularity exponent, as it describes the local degree of singularity or regularity around the point $x$.

Visual methods illustrate how DNA sequences are read along a single DNA strand from the 50 end to the 30 end and they provide the hopes of gaining an understanding of the underlying genomic language. By handling genomic sequence residues as elements of a discrete-time signal, digital signal processing techniques can be employed for the analysis of genomic information. Using these representations and applying frequency domain transformations, it is shown that structures, or seemingly nonrandom behavior, may be readily identified in nucleotide sequences. We review the basic method of DNA walks and we show how these representations can be used to extract useful knowledge from the genomic data; namely longrange correlation information, sequence periodicities, and other sequence characteristics. Further information is elucidated through wavelet transform analysis. This work finally relates a measure of sequence complexity to these visual findings and offers conclusions regarding quantifying DNA sequence behavior or structure. (John A. Berger, 2004)

**DNA Walk Plot:**

The plot is drawn between DNA nucleotide number and Walk plot values. The walk plot value is generated for each nucleotide by the following rule:

- If ‘A’ or ‘G’ is encountered -1 is added.
- If ‘C’ or ‘T’ is encountered +1 is added.

**Objective of this paper is to find out the fractal nature of cancer and plot the graph through the programming language (MATLAB) using the Daubechies Wavelet Transform.**

**METHODS**

**Detrended Fluctuation Analysis:**

Detrended fluctuation analysis (DFA) is a method for determining the statistical self-affinity of a signal. It is useful for analyzing time series that appear to be long-memory processes (diverging correlation time, e.g. power-law decaying autocorrelation function) or $1/f$ noise.

The method of detrended fluctuation analysis has proven useful in revealing the extent of long-range correlations in time series. Briefly, the time series to be analyzed (with $N$ samples) is first integrated. Next, the integrated time series is divided into boxes of equal length, $n$. In each box of length $n$, a least squares line is fit to the data (representing the trend in that box). The $y$ coordinate of the straight line segments is denoted by $y_n(k)$. Next, we detrend the integrated time series, $y(k)$, by subtracting the local trend, $y_n(k)$, in each box. The root-mean-square fluctuation of this integrated and detrended time series is calculated by

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} [y(k) - y_n(k)]^2}.$$
Figure: The integrated time series: where $B(i)$ is the interbeat interval. The vertical dotted lines indicate boxes of size $n=100$, and the solid straight line segments represent the "trend" estimated in each box by a linear least-squares fit.

This computation is repeated over all time scales (box sizes) to provide a relationship between $F(n)$ and the box size $n$. Typically, $F(n)$ will increase with box size $n$. A linear relationship on a double log graph indicates the presence of scaling (self-similarity)—the fluctuations in small boxes are related to the fluctuations in larger boxes in a power-law fashion. The slope of the line relating $\log F(n)$ to $\log n$ determines the scaling exponent (self-similarity parameter), $\alpha$ as discussed.

Wavelets:

A wavelet is a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero. It can typically be visualized as a "brief oscillation" like one might see recorded by a seismograph or heart monitor. Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. Wavelets can be combined, using a "shift, multiply and sum" technique called convolution, with portions of an unknown signal to extract information from the unknown signal.

Daubechies Wavelet:

Named after Ingrid Daubechies, the Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling function (also called father wavelet) which generates an orthogonal multi resolution analysis.

MATLAB:

MATLAB stands for "MATrix LABoratory" and is a numerical computing environment and fourth-generation programming language. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, and Fortran.

Procedure:

1. The sequences of interest were collected from NCBI REFSEQ database. (The sequences are provided under the sequences section)

2. They were pre-processed to homogenize the data set.

3. A program was written to generate the DNA Walk Plot for the sequences (Refer the figures section for further details)
4. The Plot was then processed initially by the detrended fluctuation analysis method to extract the log-log plot, which suggested the fractal nature of the sequences.

5. After this being established, the DNA Walk Plot Data was then reanalyzed with wavelet based Multifractal Detrended Fluctuation Analysis. This was done with the help of a program written in MATLAB and Daubechies 3 wavelet was used to remove the quadratic fluctuation in the curve.

6. The resultant plot was obtained between f(q) and q.

7. This was further analyzed to prove the fractal nature of the gene sequence.

**MATLAB PROGRAMS:**

```
% fft fractal program
format long;
a=load('lc3_result.txt');
y=a(1,:);
x=[];
for i=0:length(a)-1
    x=[x i];
end
l=length(y);
meany=mean(y);
diffy=y-meany*ones(1,1);
cumy=cumsum(diffy);
fft_cumy=fft(cumy);
half_fft_cumy=fft_cumy(1:(l/2));
figure,scatter(log(1:(l/2)),log(half_fft_cumy));

% figure,scatter(log(1:l),log(fft_cumy));
fslope=alpharegline1(log(abs(fft_cumy)));
fslope

% qvalue program
clc;
clear;
q=input('enter the stop value of q');
qslp=[];
z=1;
for i=-q:q
    qslp(z)=wavemfd1a1(i);
tauq(z)=q*qslp(z)-1;
z=z+1;
end
lam=-q:q;
figure,plot(lam,qslp);

% figure,plot(lam,tauq);
% wavelet based multifractal detrended fluctuation analysis program
function[slope]=wavemfd1a1(q)
format long;

% name = input('enter the file name along the path','s');
a=load('lc3_result.txt');
y=a(1,:);
x=[];
for i=0:length(a)-1
    x=[x i];
end
l=length(y);
figure,plot(x,y);
b=[];
sumb=[];
for i=1:l
    b=[b y(i)];
average=mean(b);
lb=length(b);
u=b-(ones(1,lb)*average);
sumb(i)=sum(u);
end
m=1;
wavelet='db3';

%m=input('enter the level to which decomposition is to be done');
%wavelet=input('enter the wavelet to be used','s');
temp=sumb;
for count=1:m
    temp=dwt(temp,wavelet);
    temp=approx;
    approx=[];
end
reconst=temp;
for count=1:m
```

Octa Journal of Biosciences
reconstr=idwt(reconst,zeros(1,length(reconst)),wavetable);
end
profdiff=sumb-reconst;
reprofldiff=wrev(profdiff);
boxsrt=1;
box=1;
boxstp=1;
for i=boxsrt:box:boxstp
nofbox=l/i;
accumvar=[];
for j=1:nofbox
k0=((j-1)*i)+1;
end
%this gives starting of each box
varbox=var(profdiff(k0:(i+j)));
accumvar(j)=varbox;
end
k=j-1;
for j=1:nofbox
k0=((j-1)*i)+1;
end
%this gives starting of each box
varbox=var(reprofldiff(k0:(i+j)));
accumvar(k+j)=varbox;
end
fq(i)=qfluc(accumvar,q);
end
[slope boo]=alpharegline(log(fq(4:l)));
booo;

%scatter(log(1:1),log(fq));
end

RESULTS AND DISCUSSION

The DNA walk plots were analyzed and the following conclusions had been obtained:

• All the genes analyzed have been found to be exhibiting fractal nature from the log-log plots.
• The genes of same cancer have a similar plot of F(q)-q
• The fractal dimensions of each gene has been tabulated at q=2, which is the same as obtained in a Detrended Fluctuation Analysis of a monofractal data.

It is concluded that cancer exhibit multifractal nature, which help us understand the variation of the genes from a statistical point of view. Experimental results show that DNA is multifractal, and that the multifractality changes depending upon the location (coding or noncoding region) in the sequence.

There is a significant difference between coding region and noncoding region with DNA sequence power spectrum. For instance, studies have found that mutations in certain parts of the noncoding regions lead to cancer. Physicists backed the suspicions a few years ago, when those studying fractals noticed certain patterns in junk DNA. They found that noncoding sequences display what are termed long-range correlations. Those signs suggested that junk DNA might contain some kind of organized information.

Chaos game representation (CGR) is proposed as a scale-independent representation for DNA sequences and provides information about the statistical distribution of oligonucleotides in a DNA sequence. Multifractal method is effective for describing the fractal feature of CGR graph of DNA sequences. Multi fractal techniques have proven successful in the examination of a wide range of signals such as computer images, radio signals, human fingerprints and speech. By using them to analyze DNA, the prospect of finding some higher Level information is very good.

WAVELET BASED MULTIFRACTAL DETRENDED FLUCTUATION ANALYSIS PLOTS:

X-AXIS : F(q)
Y-AXIS : q

Octa Journal of Biosciences
Skin cancer sequence 2 : NM_003070.3

REFERENCES


5. Izquierdo-Kulich E, de Quesada MA, Pérez-Amor CM, Texeira ML, Nieto-Villar JM. (2009); The dynamics of tumor growth and cells pattern morphology, Mathematical Biosciences and Engineering, Pages: 547 - 559, Volume 6, Issue 3,
