MATHEMATICAL MODELING OF P53 GENE BASED ON MATLAB CODE

Nadia Helal[1,*], Rizk Abdel Moneim[2] & Marwa Fathi[2]
[1] Nuclear& Radiological Regulatory Authority, Nasr City 11762, P. O. Box 7551, Cairo, Egypt.
[2]Physics Department, Faculty of Science, Helwan University, Ain Helwan 11795, Cairo, Egypt
*__Email__: Nadia_Helal2003@yahoo.com

ABSTRACT
Accurate prognosis of cancer can spare a significant number of cancer patients from receiving unnecessary radiotherapy/chemotherapy and its related expensive medical coasts. Recent studies have demonstrated the potential value of gene expression signatures in assessing the risk of post-surgical disease recurrence. However, these studies all attempt to develop genetic marker-based prognostic systems to replace the existing clinical criteria, while ignoring the rich information contained in established clinical markers. Given the complexity of cancer prognosis, a more practical strategy would be to utilize both clinical and genetic marker information that may be complementary. In this study, we introduce 2D graphical representation method based on the neighboring dual nucleotides of p53 gene. The introduced graph is applied to characterize and compare coding sequences of normal p53 gene and 10 different cancers. For this purpose, we have firstly, construct a model using Matlab program. Secondly, the constructed model is further validated by applying for the IARC library of p53 mutation database http://www.iarc.fr . The cancers under considerations are: ovary, skin, breast, bladder, esophagus, gastric, head& neck, leukemia, brain and lung. The obtained 2D graph may act as a "marker" that may aid in the correct initial staging diagnosis for cancers, such as mutations within the p53 gene.

Keywords: *P53 gene, genetic-marker, dual nucleotides and matlab.*

1. INTRODUCTION

The incidence of cancer in the world is increasing, particularly in relation to prolonged life expectancy from worldwide improvements in standards of living. One gene, called p53, is of major focus in the evaluation of gene therapy, as a significant fraction of cancers have been shown to have mutation (alteration) of this gene. Studies have indicated that there is a high rate of error in the initial determination of extent and/or aggressiveness (stage) of cancers, which leads to suboptimal therapeutic approaches for patients [1]. Quantitative methods represent an important technique to numerically characterize DNA sequences in the hope that such characterizations will have the way for rapid selection and identification of coding sequences. In recent years many schemes have been proposed to numerically characterize DNA sequences based on multiple dimension space such as 2D and 3D. These schemes allow visual inspection of data and can facilitate the analysis and comparison of DNA sequences. Examples of 2D graphical representation of DNA sequences have been suggested by Raychaudhury and Nandy [2] who employed quantitative techniques from 2D-graphical representation of DNA sequences to develop DNA descriptors and showed that the resulting numbers tallied well for the species considered. Examples of 3D graphical representation of DNA sequences include the methods [3-5]. The graphs obtained therefrom for individual genes have been found to vary slightly from one cancer to another, which has been interpreted to be a consequence of evolutionary divergence.

2. METHODOLOGY

In this study, we have presented P53 gene by a simple 2D graphical approach. For this purpose, a Matlab program has been constructed based on the dual pairs of the neighboring nucleotide in the DNA sequence. The program is further validated using the IARC library of P53 database http://www.iarc.fr.

2.1 Outline of the 2D graphical representation of P53 sequence based on neighboring dual nucleotide (DN)

This mathematical method is applied in order to provide a direct and simple graphical approach that can display the features based on the pairs of the neighboring nucleotides in a DNA sequence. The four DNA bases A, C, G, and T can be divided into classes, purine R = {A,G}, pyrimidine Y = {C,T}, amino M = {A,C}, keto K = {G,T}, and weak-H bond W = {A,T}, strong-H bond S = {C,G} according to their chemical properties. By considering the base order, i.e., AG is different from GA, twelve combinations can be obtained: AG, GA, CT, TC, AC, CA, GT, TG, AT, TA, CG, and GC. If we take two same bases at a time out of A, C, G, and T, we have four combinations: AA, CC,

GG and TT. Consequently, we obtain dual nucleotides set (DN-set): {AG,GA,CT,TC,AC,CA,GT, TG,AT,TA,CG,GC,AA,CC,GG,TT}, which can represent all possible mutations based on the pairs of the neighboring nucleotides [6, 7]. According to Zhaohui Qi and Xiaoqin Qi [8], a DN-curve on the cartesian coordinate system has been constructed. They have assigned a positive integer i (i = 1,2,3, . . .,N-1, where N is the length of the DNA sequence being studied) to +x, and the element of DN-set to +y, while the corresponding curve extends in the first quadrant. In detail,

Let,

$S = s_1 s_2 \ldots s_n$ be an arbitrary DNA primary sequence. Then we define a map φ, which maps S into a plot set, as the following:

$$\phi(s_i s_{i+1}) = \begin{cases} (i,1) \ if \ \ s_i s_{i+1} = AG \\ (i,2) \ if \ \ s_i s_{i+1} = GA \\ (i,3) \ if \ \ s_i s_{i+1} = CT \\ (i,4) \ if \ \ s_i s_{i+1} = TC \\ (i,5) \ if \ \ s_i s_{i+1} = AC \\ (i,6) \ if \ \ s_i s_{i+1} = CA \\ (i,7) \ if \ \ s_i s_{i+1} = GT \\ (i,8) \ if \ \ s_i s_{i+1} = TG \\ (i,9) \ if \ \ s_i s_{i+1} = AT \\ (i,10) \ if \ \ s_i s_{i+1} = TA \\ (i,11) \ if \ \ s_i s_{i+1} = CG \\ (i,12) \ if \ \ s_i s_{i+1} = GC \\ (i,13) \ if \ \ s_i s_{i+1} = AA \\ (i,14) \ if \ \ s_i s_{i+1} = CC \\ (i,15) \ if \ \ s_i s_{i+1} = GG \\ (i,16) \ if \ \ s_i s_{i+1} = TT \end{cases}$$

2.2 MATLAB code

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation. Using the MATLAB product, one can solve technical computing problems faster than with traditional programming languages, such as C, C++, and Fortran. One can use MATLAB in a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modeling and analysis, and computational biology. Add-on toolboxes (collections of special-purpose MATLAB functions, available separately) extend the MATLAB environment to solve particular classes of problems in these application areas. One can integrate MATLAB code with other languages and applications, and distribute MATLAB algorithms and applications [9]. The following charts (1-3) showing the constructed Matlab program based on the neighboring dual nucleotide discussed above.

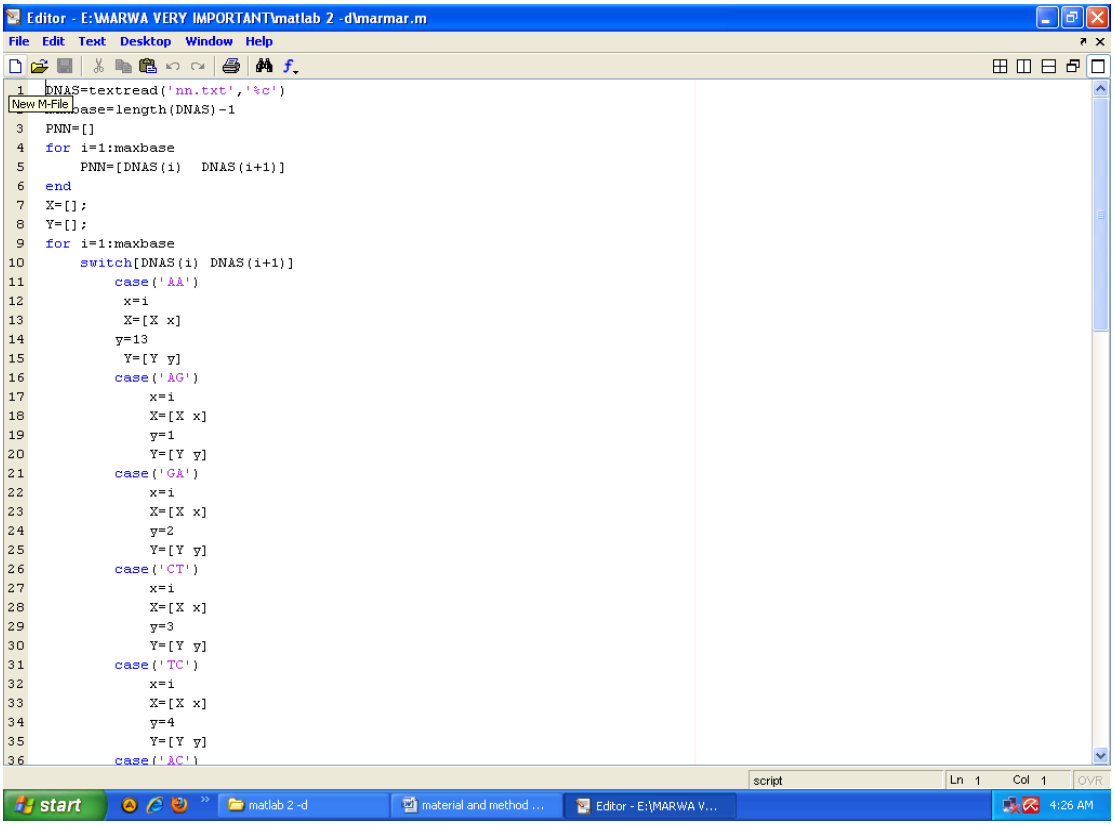

Figure 1: the constructed program of P53 gene based on neighboring dual nucleotide.

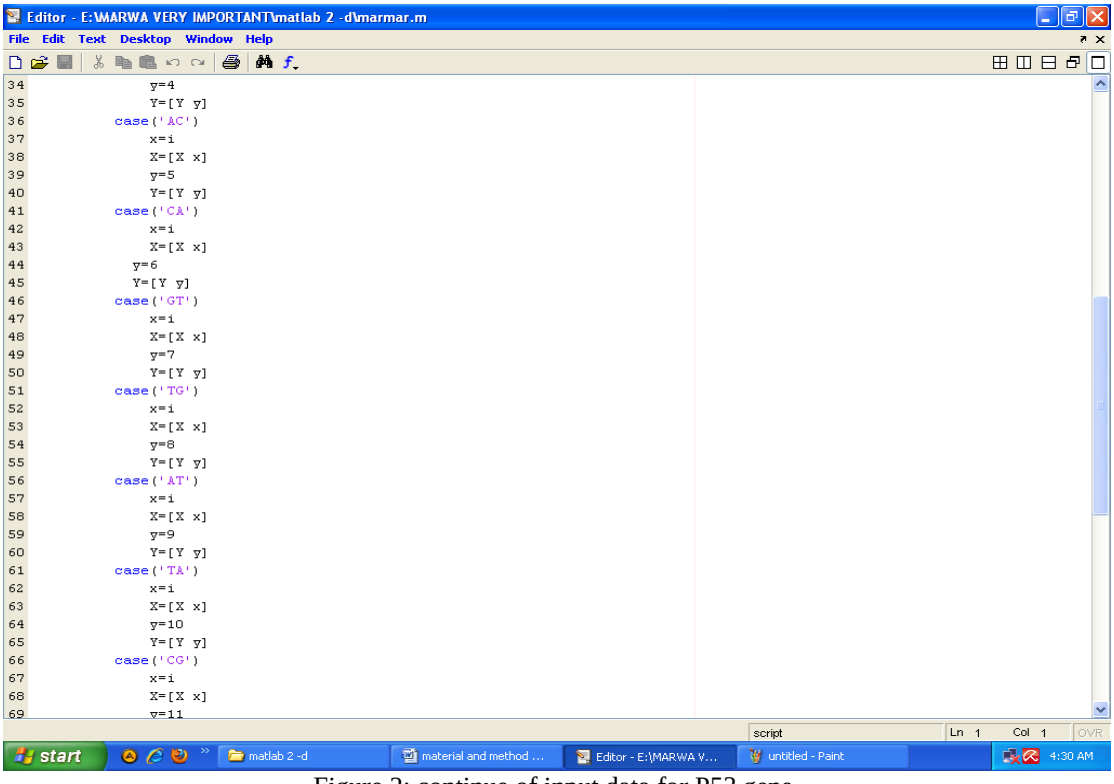

Figure 2: continue of input data for P53 gene.

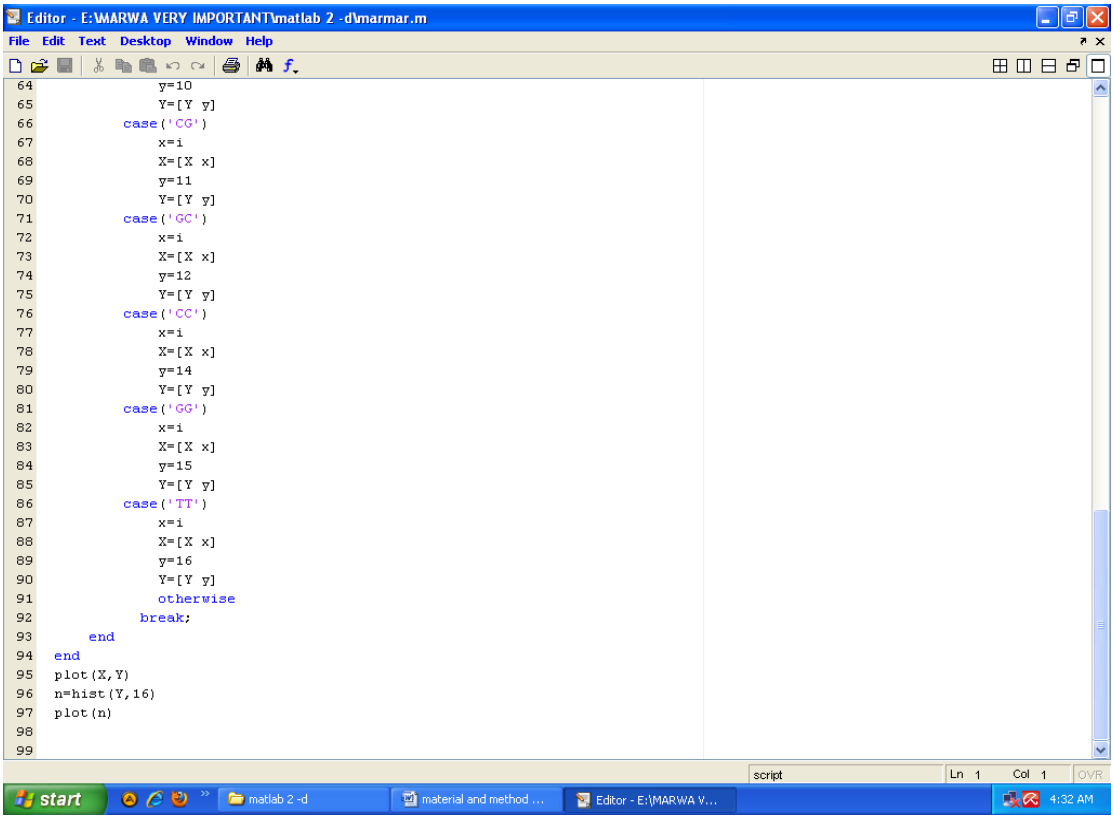

Figure 3: the last chart of the input data of P53 gene.

3. RESULTS AND DISCUSSIONS

The IARC TP53 Database compiles TP53 gene variations identified in human populations and tumor samples. Data are compiled from the peer- reviewed literature and from generalist databases. The following datasets are available:

- TP53 **somatic mutations** in sporadic cancers
- TP53 **germline mutation** in familial cancers
- Common TP53 **polymorphisms** identified in human populations
- **Functional and structural properties** of p53 mutant proteins
- TP53 gene status in human **cell-lines**
- **Mouse-models** with engineered TP53

A computational study is performed on the IARC database from the website http://www.iarc.fr. The corresponding curve is called dual nucleotides curve (DN curve). It represents P53 (1164 bases) in two dimensions. Figures (4-6) illustrates the DN curves of normal sequence of P53 compared to 10 different mutant sequences corresponding to 10 different cancers.

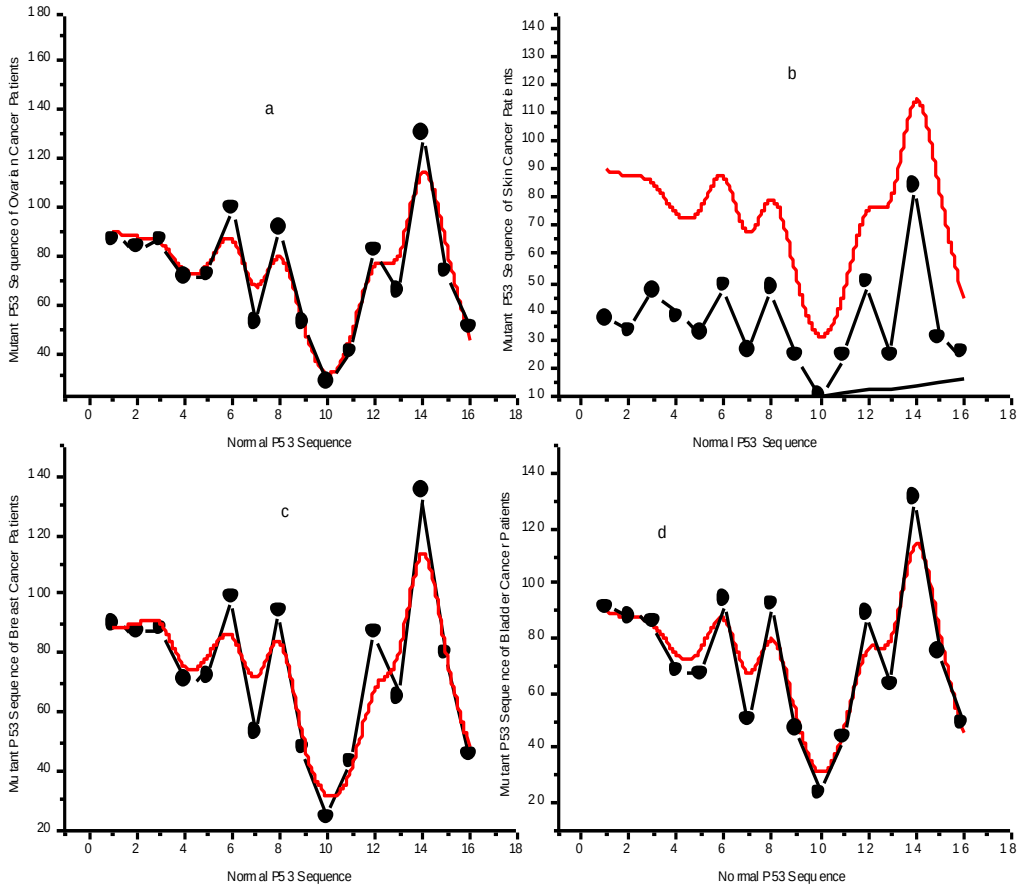

Figure 4: 2D dual nucleotide method used to represent normal P53 gene sequence against: a) ovarian cancer, b) skin cancer, c) breast cancer and d) bladder cancer.

From figure 4, by comparing the normal and mutant curves of the cancer cases we can find that: the obtained DN curves of ovary, skin, breast and bladder reflects the distribution of mutational changes. These differences possibly reveal biological function of different classification of bases. In addition we can still find out some interesting characteristics. For example, the graph areas of ovarian (fig. 4-a) cancer from the 1st pair of bases to the 10th pair of bases are almost the same as the corresponding graph areas of breast (fig. 4-c) and bladder (fig. 4-d) cancers. There graph areas at the 14th pair of bases show the highest mutation. The graph areas of skin cancer (fig. 4-b) from the 1st pair of bases to the end of the graph is so much dissimilar to the other cancers.

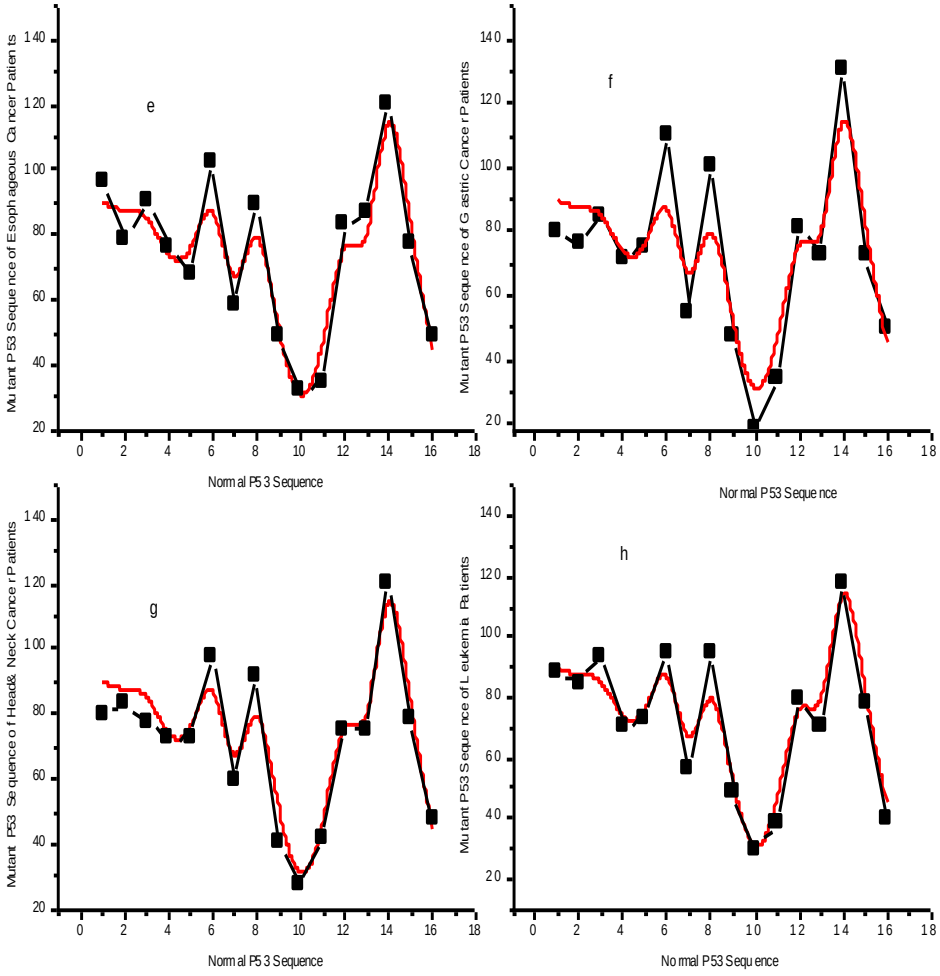

Figure 5: 2D dual nucleotide method used to represent normal P53 gene sequence against: e) Esophagus cancer, f) Gastric cancer, g) Head& Neck cancer and h) Blood cancer (Leukemia).

In figure 5 above, we can find that: the graph areas of the four cancers from the 1st pair of bases to the 5th pair of bases, also at 6th, 7th and 8th pair of bases are much dissimilar. Except gastric cancer (fig. 5f), the esophagus cancer (fig. 5e), head& neck cancer (fig. 5g) and blood cancer (fig. 5h) have regions of similarity at the 14th and 16th pair of bases.

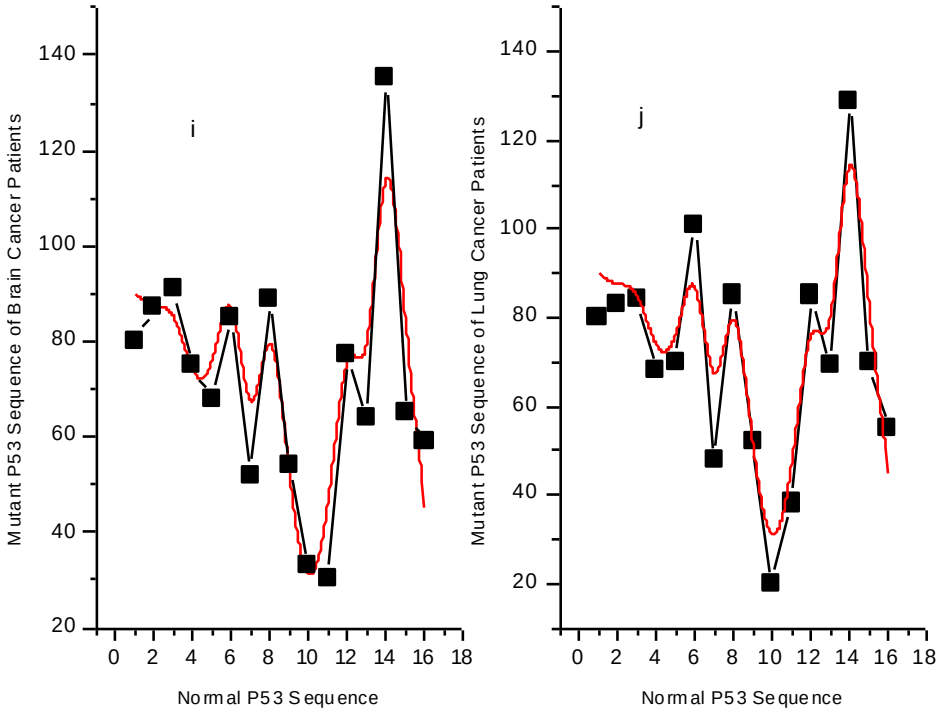

Figure 6: 2D dual nucleotide method used to represent normal P53 gene sequence against: i) brain cancer, and j) lung cancer.

In brain cancer (fig. 6i), the highest mutation points are found in 7^{th}, 13^{th} and 15^{th} pair of bases region. While the hot points of lung cancer (fig. 6j) are found at the 5^{th}, 6^{th}, 10^{th} and 15^{th} pair of bases along the dual nucleotide curves.

4. CONCLUSION

The graphical representation method outlined here has provided us with a powerful tool for characterization and comparison of P53 sequences. In this study, based on the pairs of the neighboring nucleotides, we outlined a 2D graphical representation of P53 gene sequences: the DN curve, can offer the properties of mutations in a DNA sequence. It is useful for both computational scientists and molecular biologists to visualize the local and global features of any DNA sequences and facilitate the visual discovery of interesting features in a DNA sequence. The advantage of this scheme is that it allows visual inspection of data based on DNs, helping in recognizing major similarities of mutations among different DNA sequences.

Based on the study on a 2-D graphical representation models, the present work indicate the rationality of the graphical representation methods that compare different mutated p53 gene and normal gene sequences.

5. REFERENCES

[1]. IAEA-TECDOC-1040, "Design and Implementation of a Radiotherapy Program": Clinica, Medical Physics, Radiation Protection and Safety aspects.
[2]. Barry Lacopetta, "TP53 Mutation in Colorectal Cancer", Human Mutation 21:271-276 (2003)
[3]. Massimo Tommasino, Rosita Accardi, Sandra Caldeira, Wen Dong, Ilaria Malanchi, Anouk Smet, and Ingeborg Zehbe "The Role of TP53 in Cervical Carcinogenesis" Human Mutation 21: 307-312 (2003)
[4]. B.Liao and T M Wang, "on property of the invariant of graphical representation" J.Biomol.Str.Dyn.21(2004)827.
[5]. F.L.Bai and T.M.Wang, " on property of the invariant of graphical representation" J.Biomol.Str.Dyn ,23(2006)537.
[6]. C.Li and J. Wang, Comb.Chem.HighT.Scr.7(2004)23.
[7]. Zhaohui Qi and Xiaoqin Qi, "Novel 2D graphical representation of DNA sequence based on dual nucleotides", Chemical Physics Letters, 2007,249,681-690.
[8]. M. Randic, M. Vracko, N. Lers, D. Plavsic " Analysis of Similarity/Dissimilarity of DNA Sequences based on Novel 2-D Graphical Representation" Chem. Phys. Letter. 202, 371(2003)
[9]. Ping-an He, and Jun wang "Characteristic Sequences for DNA Primary Sequence" J. Chem. Comput. Sci. 2002, 42, 1080-1085
[10]. Xiao-Qin Qi , Jie Wen, and Zhao-Hui Qi, "New 3D graphical representation of DNA sequence based on dual nucleotides", Journal of Theoretical Biology 249 (2007) 681–690