International Journal of New Technology and Research (IJNTR)
ISSN:2454-4116, Volume-1, Issue-6, October 2015 Pages 43-47

Statistical Analysis of Genomic Information: Correlations Exist in Potential Relevance to the Pathogenesis of Alzheimer Disease

Tong-han Lan,Zi Yang Lan , Xiao Li,Hao Cai

***Abstract*—In this paper the detrended fluctuation analysis (DFA) method is introduced to characterize the self-similarity of DNA sequences. Using this method, four whole genomes of the pathogenesis of Alzheimer Disease provided by NCBI are analyzed. The correlation properties in the nucleotide density distribution among these DNA sequences are explored. Calculating the sample DFA exponent based on seven different mapping rules corresponding to four different pathogenesis of Alzheimer Disease, DFA exponent α is significantly larger than 1.5. The main outcome of this study reveals existence of the correlation in the pathogenesis of Alzheimer Disease. This result points to need a new direction for analyzing and understanding the intrinsic structures of DNA sequences.**

***Index Terms*—DFA; Correlation; pathogenesis of Alzheimer Disease.**

I. INTRODUCTION

Alzheimer Disease (AD) is the most common form of dementia and the most frequent degenerative brain disorder that humans encounter in old ages, The risk of developing AD substantially increases after age of 65[1].AD is rapidly becoming one of the major universal healthcare problems. While the cause of this disease remains unknown, there is evidence for substantial genetic influence[1]-[5]. With the unclear pathogenesis, several hypotheses about the pathogenesis of AD has been proposed, such as the abnormal protein hypothesis, cholinergic hypothesis, oxidative stress theory and the estrogen hypothesis, etc. Mutations in three genes, a myloid precursorprote in (APP), presenilin 1 and 2(PSEN1 , PSEN2), result in an early onset, autosomal dominant form of the disease beginning in the third or fourth decade. The ε4 allele of apolipoprot ein E (APOE) increases the risk of both sporadic and familial AD occurring later in life around the sixth decade. Each of these genes is involved in the production or processing of the amyloid β peptide, which is deposited in the brain as dense plaques that are characteristic of the disease [3]. At present, however, there are neither precise diagnostic approaches nor effective therapeutic agents available for Alzheimer Disease.

Tong-Han Lan, Department of Biomedical Engineering,Huangzhong University of Science and Technology,Wuhan 430074 PR China.

Zi-Yang Lan, Department of Biomedical Engineering,Huangzhong University of Science and Technology,Wuhan 430074 PR China.

Xiao Li, Department of Biomedical Engineering,Huangzhong University of Science and Technology,Wuhan 430074 PR China.

Hao Cai, Department of Biomedical Engineering,Huangzhong University of Science and Technology,Wuhan 430074 PR China.

In the last decade or so there has been a ground swell of interest in unraveling the mysteries of DNA. In order to distinguish coding regions from non-coding ones, many approaches have been proposed. The correlation properties of nucleotides in DNA sequences were investigated as well[6]-[15]. C.K. Peng.et al found that there is long-range correlation in non-coding regions but not in coding regions by using the one-dimensional DNA walk model. At the same time, the nonlinear scaling method, such as complexity(Jun Xu,et al.,1994) and fractal analysis[13]-[17] were used.

Detrended fluctuation analysis is a scaling analysis method providing a simple quantitative parameterthe scaling exponent αto represent the correlation properties of a signal. The advantage of DFA over many other methods is that it permits the detection of long-range correlations embedded in seemingly nonstationarity time series, and also avoids the spurious detection of apparent long-range correlations that are an artifact of nonstationarity. The DFA method has been applied in many fields, such as cardiac dynamics, bioinformatics, economics, meteorology, material science and biology, etc[17]-[23].The correct interpretation of the scaling results obtained by the DFA method is crucial for understanding the intrinsic kinetics of the systems under our study. In fact, for all systems where the DFA method was applied, there are many problems that remain unexplained. Two of the common challenges are that the correlation exponent is not always a constant and crossover often exists. The main purpose of this paper is to analyze the scaling behavior of the fluctuations in the pathogenesis of Alzheimer Disease by means of detrended fluctuation analysis and to reveal the existence of correlation features in the pathogenesis of Alzheimer Disease. Moreover, we proposed a global measure of the DNA walker that has potential utility in analysis of the existence of correlation features in the pathogenesis of Alzheimer Disease. Although we did not discuss the possible mechanism of the pathogenesis of Alzheimer Disease, the results suggested existence of correlation in the pathogenesis of Alzheimer Disease. Thus the research has technically significance.

This paper is divided into four parts: the first part is the introduction of the pathogenesis of Alzheimer Disease and DNA series analysis basic situation; in the second part, we briefly review materials and the detrended fluctuation analysis method; the third and four part are the results and discussion of our research.

II. Materials and Methods

A. Data Resources

We will use the tools of the World Wide Web to search the GenBank DNA sequence database (http://www.ncbi.nlm.nih.gov). Homo sapiens amyloid beta (A4) precursor protein (APP), RefSeqGene on chromosome 21(NCBI Reference Sequence: NG_007376.1, GI:166795291); Homo sapiens apolipoprotein E (APOE), RefSeqGene on chromosome 19[24]; Mus musculus presenilin-1 gene, alternatively spliced transcripts, complete cds,GenBank: AF007560.1, GI:2463667; Mus musculus presenilin-2 gene,strain 129X1/SvJchromosome 12 CRA_211000022007779, whole genome shotgun sequence,GenBank: AAHY01101600.1, GI:69874353.

B. Mapping rules

A DNA sequence $\{n_i\}$ (i=1,2,…,L)of length L is comprised a series of 4 types of bases as following: adenine(A); thymine(T); guanine(G); and cytosine(C). In order to apply numerical methods to nucleotide sequence, we first prepare seven numerical sequences $\{u_i\}$,corresponding to seven means of mapping the original nucleotide sequence onto a one-dimensional numerical sequences. Mapping rules[25],[26] used to convert DNA sequences into binary numerical sequences are displayed in table Ⅰ.

Table Ⅰ

Rule	Assignment	
RY	A or G=1	C or T=-1
A	A =1	T, C or G=-1
G	G =1	A, T or C=-1
T	T =1	A, C or G=-1
C	C =1	A, T or G=-1
SW	C or G=1	A or T=-1
KM	A or C=1	G or T=-1

The RY rule has been widely used, but the other rules have also been applied.

C. Detrended fluctuation analysis

The method of detrended fluctuation analysis has proven useful in revealing the extent of long-range correlations in time series. It provides a simple quantitative parameter-the scaling parameter á, which is a signature of the correlation properties of the signal. The DFA method is initiated with dividing a time series y(k) of length N into N/n nonoverlapping boxes (also called windows). 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 yn(k). We detrend the integrated time series y(k) by subtracting the local trend yn(k) in each box. The root-mean-square fluctuation of this integrated and detrended time series is calculated by Eq.(1).

$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k)-y_n(k)]^2} \quad (1)$$

This computation is repeated over all time scales (box sizes) to characterize the relationship between F(n) and n, namely $F(n) \sim n^{\alpha}$,the average fluctuation, as a function of box size. Typically, F(n) will increase with box size n. A linear relationship on a log-log plot indicates the presence of power law (fractal) scaling. Under such conditions, the fluctuations can be characterized by a scaling exponent α , the slope of the line relating log F(n) to log n[19]. The possible result that α =0.5 indicates that the changes in the values of a time series are random, namely uncorrelated with each other. If 1> α >0.5, the signal is positive persistency (correlation); while if 0< α <0.5, the signal is antipersistent. If α >1, correlations exist but cease to be of a power-law form; α =1.5 indicates brown noise.

III. Results

A useful way of analyzing patchiness which arised from the heterogeneous purine-pyrimidine content is DNA walk, defined as table 1 above. The displacement of walker after n steps, y(n) is defined as $y(n) = \sum_{i=1}^{n} u_i$ and will display on a graph of y(n) vs n .

A. The characteristic of DNA walk

We find apparent patchiness in real DNA sequences—both in the noncoding and coding regions. Figure1(a-g) shows a representative DNA walk for different-rule mappings of amyloid beta precursor protein (APP) sequences.

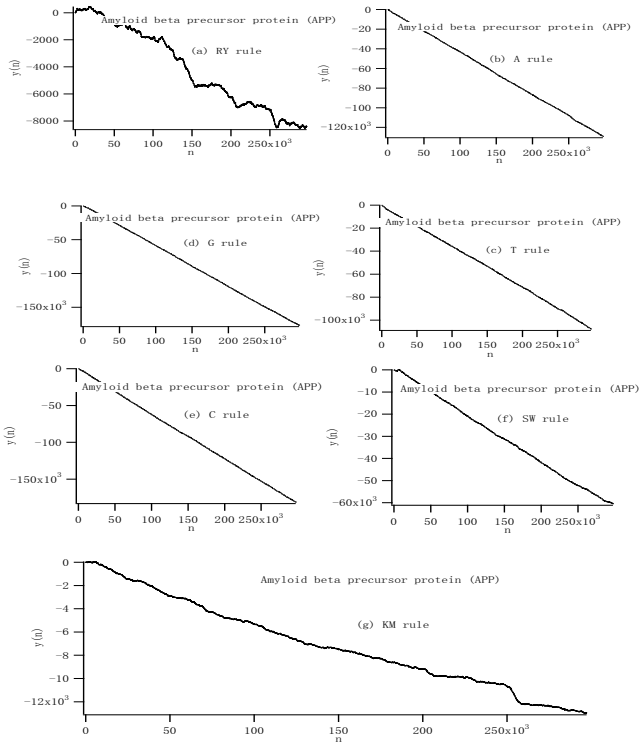

Fig1.(a-g) shows a representative DNA walk for different-rule mappings of amyloid beta precursor protein (APP) sequences.

B. The characteristic of brown noise and white noise

We use simulation method and produce two series, namely brown noise and white noise.The displacement of walker after n steps y(n) is used, Fig2.shown brown noise and white noise walker.

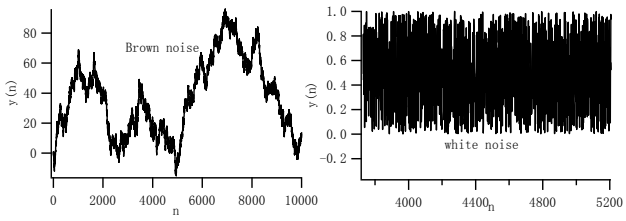

Fig2.shown brown noise and white noise

C. Detrend fluctuation analysis results

A series of double-logarithmic plots of the detrended fluctuation analysis exponent of DNA walk was demonstrated for a different rule mapping of amyloid beta precursor protein (APP) sequences.it is shown in Fig.3. Similarly, the exponent was obtained from the slope of the line fitted through the data. Values of á calculated for four different amyloid beta (A4) precursor protein (APP)、apolipoprotein E、presenilin-1 gene and presenilin-2 gene of different rule are shown in Table Ⅱ-Ⅴ.

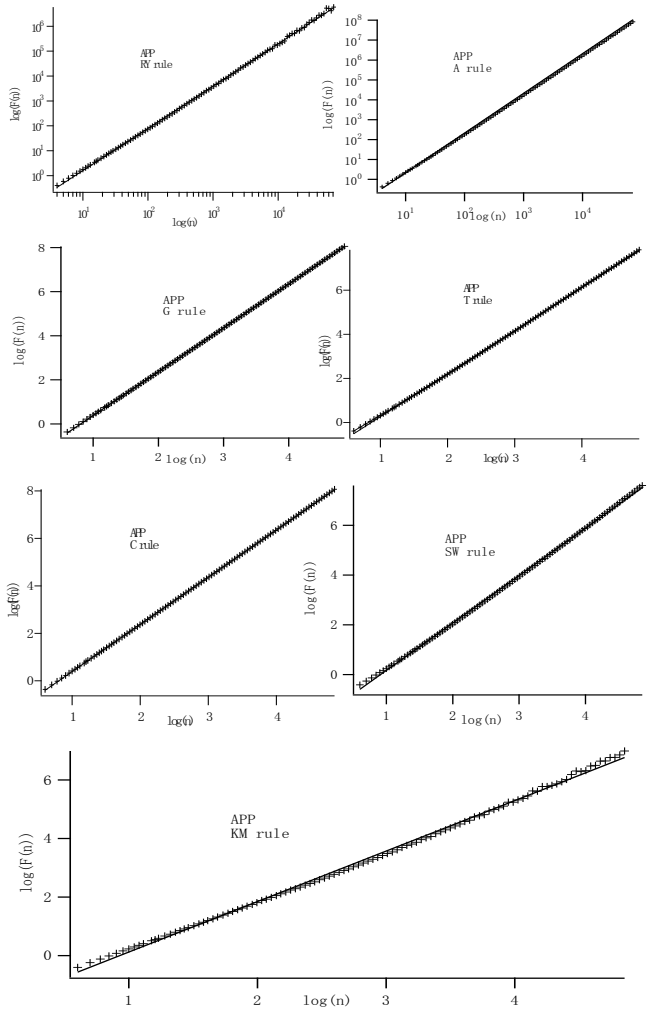

Fig.3. A series of double-logarithmic plots of the detrended fluctuation analysis exponent of DNA walk for a different rule mapping of amyloid beta precursor protein (APP) sequences are shown.

Table. Ⅱ amyloid beta precursor protein (APP) sequences values of á

Rule	Genome	α
RY	APP	1.6969
A	APP	1.9968
G	APP	1.9868
T	APP	1.9529
C	APP	1.9868
SW	APP	1.9074
KM	APP	1.7236

Table.Ⅲ apolipoprotein E sequences values of á

Rule	Genome	α
RY	APOE	1.7225
A	APOE	1.9741
G	APOE	1.9692
T	APOE	1.9362
C	APOE	1.9319
SW	APOE	1.8249
KM	APOE	1.5618

Table.Ⅳ presenilin-1 gene sequences values of á

Rule	Genome	α
RY	Presenilin-1	1.6322
A	Presenilin-1	1.9641
G	Presenilin-1	1.9496
T	Presenilin-1	1.9749
C	Presenilin-1	1.9766
SW	Presenilin-1	1.7949
KM	Presenilin-1	1.6638

Table.Ⅴ presenilin-2 gene sequences values of á

Rule	Genome	α
RY	Presenilin-2	1.7202
A	Presenilin-2	1.9499
G	Presenilin-2	1.945
T	Presenilin-2	1.9874
C	Presenilin-2	1.9867
SW	Presenilin-2	1.9312
KM	Presenilin-2	1.7224

Results above shows that Correlations Exist in potential relevance to the pathogenesis of Alzheimer Disease. Simulation rand noise and brown noise, is according to computing the detrended fluctuation analysis exponent α. The results show: Rand noise α =0.5; Brown noise α =1.5; it does not exist self-similar. Double-logarithmic plots of the detrended fluctuation analysis exponent of rand noise and brown noise RY rule mapping sequences were shown in Fig.4. The exponent was obtained from the slope of the line fitted through the data. A different rule mapping of the pathogenesis of Alzheimer Disease of detrend fluctuation analysis exponent were shown in Fig 5.

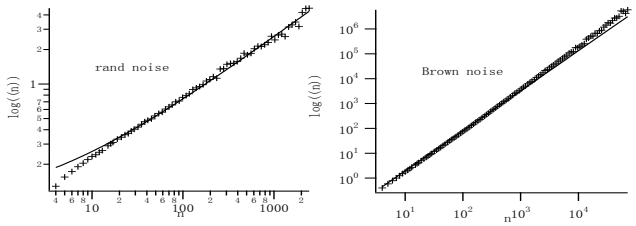

Fig.4.Double-logarithmic plots of the detrended fluctuation analysis exponent of rand noise and brown noise RY rule mapping sequences were shown.

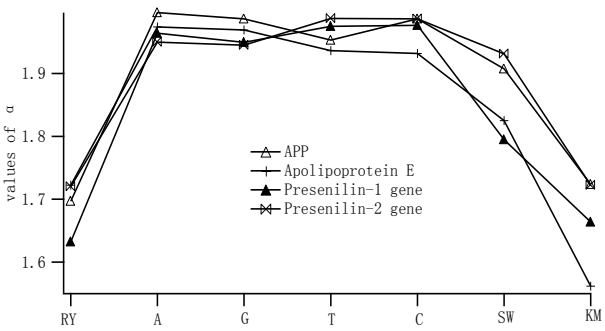

Fig 5.A different rule mapping of the pathogenesis of Alzheimer Disease of detrend fluctuation analysis exponent was shown.

IV. Discussion

We have examined the detrended fluctuation analysis exponent of DNA sequences in four difference mutution genes. The detrended fluctuation analysis exponent with different power law decays has been clearly distinguished. The result indicated that the existence of correlation in the system. The time series result was subjected to the detrended fluctuation analysis methods. The technique shows that the correlation properties of the DNA sequences. The detrended fluctuation analysis exponent is above 1.5. Although for the detrended fluctuation analysis, A power-law relationship between F(n) and n indicates scaling with an exponent á, F(n) ~ ná. Notice that such a process exists a power-law.

From the results we find that different nucleotide sequence representations will get different results. The detrended fluctuation analysis exponent ordered by representing of amyloid beta precursor protein 、 apolipoprotein E 、 presenilin-1 and presenilin-2 the sequences as A rule,G rule, C rule,T rule is almost 2. The detrended fluctuation analysis exponent value are computed above 1.5 by the other rule. In general, the detrended fluctuation analysis exponent value exhibit for 0.5 and 1.5 from “random noise” sequence and “brown noise” sequence.

In this paper the correlation properties in the whole genomes of amyloid beta (A4) precursor protein (APP)、apolipoprotein E、presenilin-1 gene、presenilin-2 gene are verified. The results of this paper suggest that the asymptotic correlation property is one of the natural properties that DNA sequences possess, and is directly related to the structure and function of the whole DNA molecule. There are still many discussions concerning the biological meanings and the origins of the correlation properties in the DNA sequences. Such as, APP gene encodes a cell surface receptor and transmembrane precursor protein that is cleaved by secretases to form a number of peptides. Some of these peptides are secreted and can bind to the acetyltransferase complex APBB1/TIP60 to promote transcriptional activation, while others form the protein basis of the amyloid plaques which are found in the brains of patients with Alzheimer disease. Mutations in this gene have been implicated in autosomal dominant Alzheimer disease and cerebroarterial amyloidosis. Multiple transcript variants encoding several different isoforms have been found for this gene[27].Chylomicron remnants and very low density lipoprotein (VLDL) remnants are rapidly removed from the circulation by receptor-mediated endocytosis in the liver. Apolipoprotein E, a main apoprotein of the chylomicron, binds to a specific receptor on liver cells and peripheral cells. ApoE is essential for normal catabolism of triglyceride-rich lipoprotein constituents[24].Transcriptional regulation of the mouse presenilin-1 gene[28],A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome[29], These properties are considered to be related to the construction of the higher order structure of the DNA molecule.

This paper is preliminary with respect to Correlations Exist in potential relevance to the pathogenesis of Alzheimer Disease. Its aim is to provide the means for the analysis of fluctuation properties of DNA sequences. Although the research results show Correlations Exist in the pathogenesis of Alzheimer Disease, This may lead to a series of questions. What is the physical phenomena and possible mechanism of correlation in the pathogenesis of Alzheimer Disease? How to categorize different types of the pathogenesis of Alzheimer Disease? How to understand the function of the pathogenesis of Alzheimer Disease? How to build the pathogenesis of Alzheimer Disease kinetics model systematically? Although the data tell us that correlation does exist in the pathogenesis of Alzheimer Disease, we will still study the physical properties of the pathogenesis of Alzheimer Disease. What kind of distribution form do these the pathogenesis of Alzheimer Disease have? All these studies should be the future direction of research.

Acknowledgment

This project was supported by China National Nature Science Foundation No:30470413, No:31160200.Hubei province Nature Science Foundation No:2004ABA220 and China Postdoctoral Science Foundation. We are also indebted to sir lan zi yang for his valuable support and helpful discussions.

References

1. Joseph H. Lee& Sa ndra Bar ral& Rong Cheng.etc.Age-at-ons et linkage analysis in Caribbea n Hispanics with familial late-onset Alzheimer Disease. Neurogenetics (2008) 9:51 –60
2. Nussbaum RL, Ellis CE (2003) Alzheimer ’s disease and Parkinson’sdisease. N Engl J Med 348:1356 –1364
3. St George-Hyslop PH, Petit A (2005) Molecular biology andgenetics of Alzheimer ’s diseas e. C R Biol 328:119 –130.
4. Tanzi RE, Bertram L (2005) Twenty years of the Alzheimer ’s disease amyloid hypothesis: a genetic perspective. Cell 120:545 –555.
5. Pastor P, Goate AM (2004) Molecular genetics of Alzheimer’s diseaseAlzheimer Disease. CurrPsychiatry Rep 6:125– 133.
6. Li W, Kaneko K.Long-range correlation and partial 1/ f spectrum in a non- coding DNA sequence.Europhys Lett,1992 17 :655～660.
7. Peng C K, Buldyrev S, Goldberg A L, et al. Long- range correlation in nucleotide sequence.Nature,1992 (356) :168～170.
8. Stanley H E, Buldyrev S V, Goldberg A L, et al. Analysis of DNA sequences Using methods of statistical physics. J.PhysicaA,1998.(249) :430～438.
9. Lou L F , Tsai Li, Zhou Y M. Informational parameters of nucleicacid and molecular evolution,J TheorBiol,1988(130) :351～361.
10. Luo L F, Tsai Li . Fractal dimension of nucleicacid sequences and its relation to evolutionary level.ChinPhysLett,1988(5) :421～424.

11. Herzel H, Grobe I. Correlations in DNA sequences – the role of protein coding segments J.PhysRevE,1997(55) :800～810.
12. Jun Xu, Yang Chao, Rensheng Chen. Fractal geometry study of DNA binding proteins.J theorBiol,1994(171) :239～249
13. Zhang Y P,Tong-han Lan,etal,Existence of memory in membrane channels:analysis of ion current through a voltage-dependent potassium single channel.cell biolint.2012(36).171-175.
14. Bai-linHao,Hoong-Chien Lee,and Shu-yu Zhang, Fractals related to long DNA sequences and complete genomes,Chaos,Solitons and Fractals, 2000.825-836.
15. Bassingth waighte, J.B. and Raymond, G.M. Evaluting rescaled range analysis for time series, Ann. Biomed. Eng. 1994;22:432-444.
16. Churilla,A.M.,Gottschalk,W.A.,Liebovitch,L.S.,Selector,L.Y.,Todoro v,A..L&Yeandle,S. Membrane potential fluctuations of human T-lymphocytes have fractal characteristics of fractional Brownian motion. Ann. Biomed. Eng.1996;24:99-108.
17. Caballero R, Jewson S, Brix A .Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation. Clim Res.2002; 21:127-140.
18. Kun Hu, Plamen Ch, Ivanov, Zhi Chen, Pedro Carpena, and H. Eugene Stanley. Effect of trends on detrended fluctuation analysis. Phys. Rev.2001;E 57, 011114.
19. Chen Z, Ivanov PC, Hu K, Stanley HE. Effect of nonstationarities on detrended fluctuation analysis. Phys Rev.2002; E 65:041107.
20. V. N. Kazachenko, M. E. Astashev and A. A. Grinevich. Multifractal analysis of K+ channel activity.Biologicheskie Membrany. 2007; Vol. 24, No. 2: 175–182.
21. Tong-Han Lan,Zhi-yong Gao,Ahmed N. Abdalla,Bo Cheng,Shu Wang.Detrended fluctuation analysis as a statistical method to study ion single channel signal.Cell Biology International 2008;32:247-252.
22. Yazawa, T., Tanaka, K., Kato, A., & Katsuyama, T.. The scaling exponent calculated by the detrended fluctuation analysis, distinguishes the injured sick hearts against normal healthy hearts. Proceeding IAING (WCECS08) International Conference on Computational Biology (ICCB), 2008;V(2):7-12. 22-24 October, San Francisco, USA.
23. Yazawa, T., Tanaka, K., & Katsuyama, T. Alternans lowers the scaling exponent of heartbeat fluctuation dynamics: A detrended fluctuation analysis in animal models and humans". Proceeding CSIE2009, World Congress on Computer Science and Information Engineering. Computer Soc.2009;April, 221-225, Los Angeles, CA, USA.IEEE DOI 10.1109/CSIE.2009.784.
24. NCBI Reference Sequence: NG_007084.2.
25. P.Bernaola-Galvan ,P.Carpena ,R.Roman-Roldan ,J.L.Oliver. Study of statistical correlations in DNA sequences. Gene2002 (300)105-115.
26. Buldyrev,S.V.,Goldberger,A.L.,Havlin,S.,Mantegna,R.S.,Matsa,M.E., Peng,C.-K.,Simons,M.,Stanley,H.E..Long-range correlations properties of coding and noncoding DNA sequences:GenBankanalysis.Phys.Rev,1995.E51,5084-5091.
27. Van Ulsen,P., van Alphen,L., Hopman,C.T., van der Ende,A. and Tommassen,J.In vivo expression of Neisseria meningitidis proteins homologous to the Haemophilus influenzae Hap and Hia autotransporters.FEMS Immunol. Med. Microbiol. 32 (1), 53-64 (2001).
28. Mitsuda,N., Roses,A.D. and Vitek,M.P.Transcriptional regulation of the mouse presenilin-1 gene.J. Biol. Chem. 272 (38), 23489-23497 (1997).
29. Mural,R.J., Adams,M.D., Myers,E.W.,A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome.Science 296 (5573), 1661-1671 (2002).