

A Comparative Study of CpG Islands Detected by Different Tools: Newcpgreport and CpGCluster*

Eng: Majd Al-ssabbagh**

Dr. Eng. Rasha Massoud***

Dr. Ali Saker****

Abstract

CpG islands play an important role in genes transcription regulation, due to the fact that these islands overlap with the genes' promoter regions, and the methylation of those CpG islands may repress the transcription of the associated genes. Previous studies reported that methylation of CpG islands is an important indicator of the presence and possibility of developing cancers. There are mainly two types of algorithms to identify CpG islands in the nucleotides sequences: distance-based and sliding-window algorithms. The outputs of these algorithms are different for the same nucleotide sequence. The aim of this study is to compare the performance of the above mentioned algorithms by using two web tools named CpGCluster and newCpGReport.

CpG islands in human chromosome 22 were identified by applying the two algorithms on this chromosome, and the variation in the number and length of the identified islands was clear. The results also show that about 60% of both tools' output is crossed. Moreover, the effect of the traditional parameters of CpG islands (length, C+G content and Observed/expected ratio) on the number of the identified islands was studied. The results show that the length parameter has a great effect on the number of islands identified by newCpGReport, while it does not affect CpGCluster's performance. The effect of making CpG islands identified by newCpGReport start and end with CpG was also studied, due to this operation C+G content and Observed/expected ratio increased for most islands, taking into account that 25% of the islands became shorter than 200 nucleotides.

Keywords: CpG islands, Methylation, Genes, Chromosomes, Cancer, C+G content, ewCpGReport, CpGCluster.

* For The paper in Arabic see pages (207-195)

** Biomedical Engineering Department, Faculty of Electrical & Mechanical Engineering, Damascus University.

*** Biomedical Engineering Department, Faculty of Electrical & Mechanical Engineering, Damascus University.

**** Department of Histology, Anatomy, Embryology and genetics, Faculty of Medicine, Damascus University.

References:

- [1] Illingworth R.S., Bird A.P., (2009), "CpG islands – 'A rough guide'", FEBS letters, Vol. 583, p.p. 1713-1720.
- [2] Hou P., Ji M., Liu Z., Shen J., Cheng L., He N., Lu Z., (2003), "A microarray to analyze methylation patterns of p16Ink4a gene 5'-CpG islands", Clinical Biochemistry, Vol. 36, p.p. 197-202.
- [3] Bastian P.J., Yegnasubramanian S., Palapattu G.S., Rogers C.G., Lin X., De Marzo A.M., Nelson W.G., (2004), "Molecular Biomarker in Prostate Cancer: The Role of CpG Island Hypermethylation", European Urology, Vol. 46, p.p. 698-708.
- [4] Claverie J.M., Notredame C., (2007), "Bioinformatics for Dummies 2nd Edition" Wiley Publishing, Inc., USA.
- [5] Wang Y., Leung F.C.C., (2004), "An evaluation of new criteria for CpG islands in the human genome as gene markers", Bioinformatics, Vol. 20(7), p.p. 1170-1177.
- [6] Ponger L., Mouchiroud D., (2002), "CpGProd: Identifying CpG islands associated with transcription start sites in large genomic mammalian sequences", Bioinformatics, Vol. 18(4), p.p. 631-633.
- [7] Hackenberg M., Previti C., Luque-Escamilla P.L., Carpena P., Martínez-Aroza J., Oliver J.L., (2006), "CpGcluster: a distance-based algorithm for CpG-island detection" BMC Bioinformatics, Vol.7, p.p. 446-459.
- [8] Ye S., Asaithambi A., Liu Y., (2008), "CpGIF: an algorithm for the identification of CpG islands", Bioinformatics, Vol.22(8), p.p. 335-338.
- [9] Lan M., Xu Y., Li L., Wang F., Zuo Y., Chen Y., Tan C.L., Su J., (2009), "CpG-Discover: A Machine Learning Approach for CpG Islands Identification from Human DNA Sequence", Proceedings of international joint conference, p.p. 1702-1707.
- [10] Chuang L., Chen Y., Yang C., (2009), "Designing of A Novel GA based on Fuzzy System for Prediction of CpG Islands in the Human Genome", Proceedings of FUZZ-IEEE international conference, p.p. 1009-1014.
- [11] Han L., and Zhao Z., (2009), "CpG islands or CpG clusters: how to identify functional GC-rich regions in a genome?", BMC Bioinformatics, Vol.10, p.p. 65-71.
- [12] Hackenberg M., Barturen G., Carpena P., Luque-Escamilla P., Previti C. and Oliver J., (2010), "Prediction of CpG-island function: CpG clustering vs. sliding-window methods", BMC Genomics, Vol.11, p.p. 327-340.
- [13] Takai D., Jones P., (2002), "Comprehensive analysis of CpG islands in human chromosomes 21 and 22", Pubmed, Vol.99, p.p. 2740-2745.

Internet websites:

- [14] Lopez R., (1999), "newCpGreport", emboss, <http://emboss.bioinformatics.nl/cgi-bin/emboss/newcpgreport>, (2011).
- [15] Takai D., Jones P.A., (2003), "The CpG Island Searcher: A New WWW Resource", <http://www.bioinfo.de/isb/2003030021/>, (2011).