International Journal of Global Engineering (IJGE) E- ISSN: 2456-3099

VOL 1 ISSUE 1 (2016) PAGES 36 - 41

Received: 21/07/2016. Published: 27/07/2016

MiRNA analysis in gene expression –Big data Analytics

[a] S. Geeitha*, [b] Dr. M. Thangamani

[a] Assistant Professor, Mahendra Engineering College for Women, Tamilnadu,India
[b] Assistant Professor, Kongu Engineering College, Tamilnadu, India

Abstract

The widely expanding big data analytics empowers the bioinformatics area where various medical information is gathered and analysis is carried out. It has made a great evolution in healthcare issues and provides solution in the medical realm. Regulating miRNA expression with hundreds of genes play a vital role in disease exploration. The integration of miRNA expression and gene expression is a huge research problem .The exploration of gene expression profiles leads to drug discovery. This paper provides the overview of big data analytics and enumerates various informatics and data mining tools and algorithms to integrate miRNA expression and gene expression.

Keywords: Bioinformatics, miRNA, big data, gene expression and Data mining.

I. Introduction

A gene is a small piece of genetic material written in a code called DNA. Gene expression is the process by which the information contained within a gene becomes a useful product and also the information from a gene is used in the synthesis of a functional gene product. By mining this gene expression data brings a great solution for predicting various diseases and implementing same for precise diagnosis.

1.1 miRNA

Any of a group of short (generally 21 to 3/24 nucleotides in length), non-coding RNA molecules which fold upon themselves ("hairpins") and are usually cleaved from large hairpin-containing RNA (itself often processed from some portion of mRNA. Any of a group of short MiRNA is conserved through evolution and plays a role in RNA interference, destroying mRNA made by specific genes. Suppressing gene expression and controlling translation of target mRNAs, thereby regulating critical aspects of plant and animal development. miRNAs (microRNAs) are short non-coding RNAs that regulate gene expression post-transcriptional

1.2. Big data analytics

Big data analytics have emerged to perform analysis on massive amount of data in descriptive and predictive manners, to provide intelligent informed decisions. Big data refers to a high volume of heterogeneous data formed by continuous or discontinuous information stream [1]. Volume of big data is growing fast in bioinformatics research. Big data is not limited to any particular physics experiments, data volume is not only raising everyday in bioinformatics research but it is also supported by decreasing computing cost and increasing analytics throughput with growing big data technologies [2].

II. Related Works

Big data integrates the data and uncover the hidden values from data which aims making the sense of data. New methodologies, modelling, visualization, machine learning, etc., are included in the big data analytics [3].

2.1. Human Genome

The human genome contains 21,000 genes approximately At any instant, each of our cells has some combination of these genes that can be turned on, and others are turned off. Scientists have to figure out which are on and which are off? Comparative analysis of the genes from a diseased and a normal cell will help the identification of the biochemical constitution of the proteins synthesized by the diseased genes. This information is used to synthesize drugs which combat with these proteins and reduce their effect.

2.2. Microarray technology

The microarray technology has the capacity to determine thousands of genes simultaneously. Analysing gene expression is very important in biological research. The data mining algorithms provide various tools to find the specific feature of gene expression in patients. Using this microarray technology, gene expression profiles acts as great medical diagnosis tool that provides the state of cell at molecular level [4] using microarray technology a sample that consists of gene both from normal and diseased persons is taken and spots are obtained for diseased genes if the gene is over expressed This expression pattern is then compared to the expression pattern of a gene [5] responsible for a disease.

2.3. miRNA with Gene Expression

miRNA regulates hundreds and thousands of gene expression to find out cause of numerous diseases. The integration of miRNA expression data sets with gene expression is major key research problem in bioinformatics field. A cloud environment named BioVLAB-MMIA is introduced for integrated analysis of miRNA and mRNA in gene expression [6].

2.4. Mining algorithms and bioinformatics tools

Classic Hungarian algorithm and Greedy algorithm is introduced for global alignment of bimolecular networks which is done in acceptable running time [7]. Regression based algorithm named bLARS that out performs the state of art algorithm that performs the regulatory interactions from a predefined target genes [8]. Khmer [9] is a software tool mainly used for pre-processing large amounts of genomic sequence data prior to analysis with conventional bioinformatics tools shown in Fig.1 and Fig.2

Received: 21/07/2016. Published: 27/07/2016

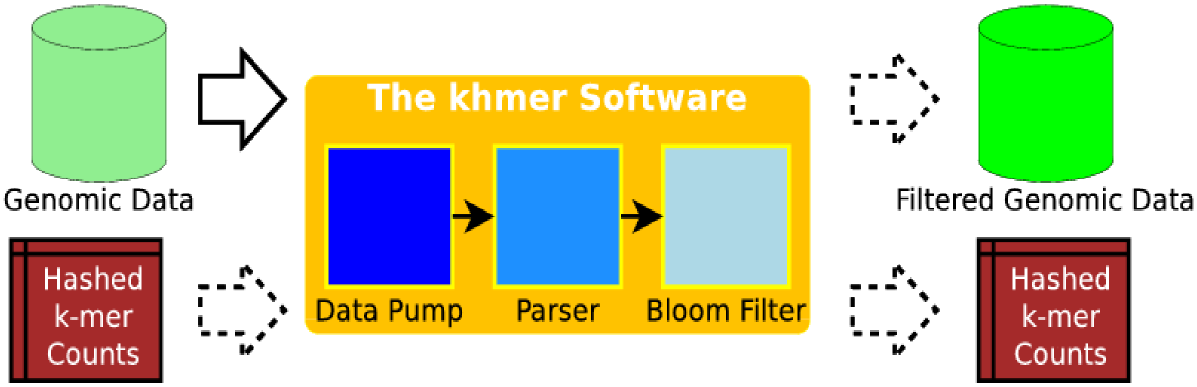

Fig.1 Data flow through Khmer software

Algorithms such as Hidden Markov model (HMM) [10] offer pattern recognition among data sets, in particular nucleotide sequences, Naïve Bayes and SVM (Support Vector Machine) are classification model that is obtained by applying a relatively simple method to a training data set and categorizes objects based on a set of features for each object respectively. ProMir, a HMM based tool **i**s a probabilistic co-learning model that is applied to predict human miRNA genes.

MiRFinder is an SVM-based tool that compares genome-wide and pair-wise sequences between related species. It identifies hairpin structures from a set of miRNA candidates and excludes non-robust structures by SVM analysis of 18 different parameters. A BPP (Bit parallel Processing) algorithm [11] is implemented to quantify miRNAs that produces 91% accuracy and Bowtie sequence mapping tool is used to increase the speed of operation.

2.5 Mining Algorithm in Gene Expression

Clustering techniques [12] reveals the natural structures and identify the patterns in the data and clustering algorithms are directly applied to the gene expression data. K-Means algorithm, partitions the data set into K disjoint subsets. Association rules provide relevant associations between different genes and gene expression. Other algorithms such as Apriori, FP growth and LPMiner can be implemented for pattern extraction.

2.6. Big data Analytics in Bioinformatics

Big data plays a significant role in medical field since big volume of data arises from huge amount of records stored for patients. Big velocity is arising when data is coming at high speed and big variety comprises to data sets with varying types [13]. The complexity of big data requires one to invent new softwares tools which help to analyse, store and visualize the data. Image processing, Signal processing and genomics are the three main areas where big data analytics concentrates. Some image techniques such as Computed tomography (CT, Photo acoustic imaging, mammography etc., are applied. Medical signals also possess high volume and velocity. Analysing genome [14] scale for diagnosing the diseases play a pivotal role in the bioinformatics field.

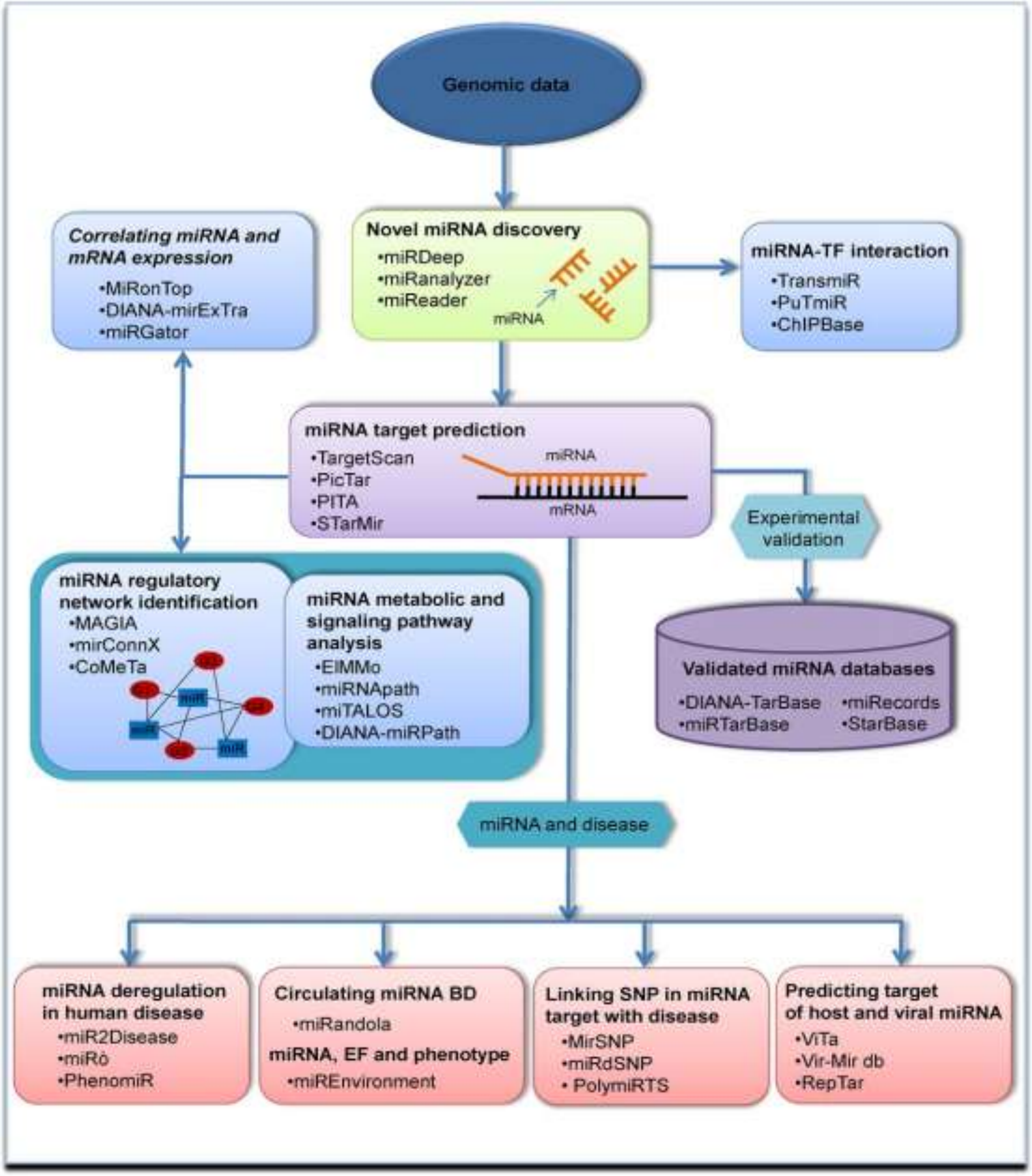

Fig.2 Overview of Bioinformatics tool

III. Conclusion

By comparing the gene profiles using various bioinformatics tools and data mining algorithms, it is possible to extract pattern of gene expression that is related to disease mechanisms. An insight at the individual genes still remains a drawback where it fails to account the connections between the genes. This paper concludes that by developing tools implementing algorithms with mathematical strategy and specific software clusters of correlated genes can be highlighted which helps to diagnosis the disease of a patient.

Received: 21/07/2016. Published: 27/07/2016

References:

1. Hirak Kashyap, Hasin Afzal Ahmed, Nazrul Hoque, Swarup Roy, and Dhruba Kumar Bhattacharyya, Big Data Analytics in Bioinformatics: A Machine Learning Perspective, Journal of Latex Class Files, Vol. 13, No. 9, Pp.1.20, 2014.
2. R. J. Robison, "How big is the human genome Precision Medicine, January 2014.
3. Qiang Yang, Introduction to the IEEE Transactions on Big Data, IEEE transactions on Big Data ,Vol. 1, No. 1, 2015.
4. Rabindra kumar Singh, M. Sivabalakrishnan, , Feature Selection of Gene Expression Data for cancer Classification: A Review -Big Data, Cloud and Computing Challenges,Procedia Computer Science, Elsvier ,Vol. 50, Pp. 52-57, 2015.
5. Premier Biosoft, Accelerating Research in Life sciences, http://www.premierbiosoft.com/tech_notes/microarray.html.
6. Hyungro Lee , Youngik Yang ; Heejoon Chae ; Seungyoon Nam ; Donghoon Choi ; Patanachai Tangchaisin ; Chathura Herath ; Suresh Marru ; Kenneth P. Nephew ; Sun Kim, BioVLAB-MMIA: A Cloud Environment for microRNA and mRNA Integrated Analysis (MMIA) on Amazon EC2, IEEE Transactions on NanoBioscience, Vol.11, No.3, Pp.266 – 272, 2012.
7. Jiang Xie , Chaojuan Xiang ; Jin Ma ; Jun Tan ; Tieqiao Wen ; Jinzhi Lei ; Qing Nie, An Adaptive Hybrid Algorithm for Global Network Alignment, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol.13 , No.3, Pp. 483-493, 2016.
8. Nitin Singh , Mathukumalli Vidyasagar, bLARS: An Algorithm to Infer Gene Regulatory Networks, IEEE/ACM Transactions on Computational Biology and Bioinformatics,Vol. 13 , No. 2, pp. 1295 - 1303, 2015.
9. C.T. Brown and et al., "khmer: genomic data filtering and partitioning software." http://github.com/ged-lab/khmer.
10. Mauluda Akhtar, Luigina Micolucci1, Md Soriful Islam, Fabiola Olivieri, and Antonio Domenico Procopio, Bioinformatic tools for microRNA dissection, Nucleic Acid Research,Vol. 44 No. 11 , 2016.
11. Saleem.A, Amjesh.R, Vinoth Chandra S.S,An Improved Algorithm for miRNA profiling from Next Generation Sequencing Data, Springer International Publishing Switzerland,Vol.3, No.10, pp.38-47, 2016.
12. Denis A. Sarigiannis, Data infrastructure and data mining model of internal exposome , WP 7 Novel bioinformatics for predictive biomarker discovery, HEALs, Version 1, http://www/heals-eu.eu/, 2015.
13. Matthew Herland, Taghi M Khoshgoftaar , Randall Wald, **A** review of data mining using big data in health informatics**,** Journal Of Big Data,Springer OpenVol.1,No.2, 2014, DOI: 10.1186/2196-1115-1-2
14. Ashwin Belle, Raghuram Thiagarajan,S. M. Reza Soroushmehr, Fatemeh Navidi,4 Daniel A. Beard, Kayvan Najarian, Big Data Analytics in Healthcare, BioMed Research International,vol.2015, No.10, 2015.

Received: 21/07/2016. Published: 27/07/2016

Authors Biography

Ms. S. Geeitha has completed Master of Engineering in Computer Science and Engineering in Anna University Application. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing and ontology. She has presented 12 papers in national and international conferences in the above fields. She is currently working as Assistant Professor in Mahendra Engineering College for Women.

Dr. M. Thangamani completed her B.E., from Government College of Technology, Coimbatore, India. She completed her M.E in Computer Science and Engineering from Anna University and PhD in Information and Communication Engineering from the renowned Anna University, Chennai, India in the year 2013. Dr. M. Thangamani possesses nearly 23 years of experience in research, teaching, consulting and practical application development to solve real-world business problems using analytics. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing, ontology development, web services and open source software. She has published nearly 70 articles in refereed and indexed journals, books and book chapters and presented over 67 papers in national and international conferences in above field. She has delivered more than 60 Guest Lectures in reputed engineering colleges and reputed industries on various topics. She has got best paper awards from various education related social activities in India and Abroad. She has organized many self-supporting and government sponsored national conference and Workshop in the field of data mining, big data and cloud computing. She has received the International Award for the “Women of Distinction” from Venus International Foundation on 5th March, 2016 and “Senior Women Educator and Scholar Award” from the National Foundation for Entrepreneurship Development on 8th March 2016. She continues to actively serve the academic and research communities and presently guiding 8 Ph.D Scholars under Anna University. She is on the editorial board and reviewing committee of leading research journals and on the program committee of top international data mining and soft computing conferences in various countries. She is also seasonal reviewer in IEEE Transaction on Fuzzy System, international journal of advances in Fuzzy System and Applied mathematics and information journals. She has organizing chair and keynote speaker in international conferences in India and countries like California, Dubai, Malaysia, Singapore, Thailand and China. She has Associate Editor in Canadian Arena of Applied Scientific Research, Canada. She has Life Membership in ISTE, Member in CSI, International Association of Engineers and Computer Scientists in China, IAENG, IRES, Athens Institute for Education and Research and Life member in Analytical Society of India. She is currently working as Assistant Professor at Kongu Engineering College at Perundurai, Erode District.