conferenceseries.com
JOINT EVENT
12th International Conference and Expo on
Proteomics and Molecular Medicine &
12th International Conference on
Advancements in Bioinformatics and Drug Discovery
November 26-28, 2018 | Dublin, Ireland

Jing Doo Wang, J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C5-124

Reducing the gap between phenotypes and genotypes via comparing tagged whole genomic sequences

Jing Doo Wang
Asia University, Taiwan
China Medical University Hospital – CMU, Taiwan

With the progress of next generation sequences (NGS) nowadays, it is possible to have whole (complete) genomic sequences of instances of distinct organisms available. It is interesting and attractive to partition selected instances of organisms into classes according to features (phenotypes) defined or observed by domain experts precisely and then to extract and identify some distinctive genomic subsequences as biomarkers by comparing their whole genomic sequences. To overcome the computational bottleneck of whole genomic sequences comparison across those instances of organisms, a scalable approach based on previous work is applied to extract the maximal repeats from these tagged whole genomic sequences and meanwhile compute class frequency distributions of these maximal repeats extracted. These repeats with extremely biased class frequency distribution or just appearing in all instance of one class, if existing, may provide valuable hints or clues for biologists to further analyze or inspect whether the relationship between defined features (phenotypes) and extracted repeats (genotypes) is significant or not. Most of all, above computation of maximal repeat extraction could be achieved via cloud computing that can provide scalable computing environment if necessary. The method as described above opens a novel direction of researches to explore the connections between phenotypes and genotypes by comparing tagged whole genomic sequences.

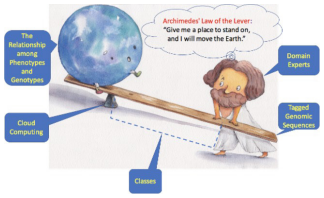

Figure 1: "Mining for biomarkers via Observing Class frequency Distributions of Maximal Repeats from Tagged Genomic Sequences" is somewhat like "Archimedes' Law of the Lever" [1]

Recent Publications:

1. Wang J D (2018) A Novel Approach to Mine for Genetic Markers via Comparing Class Frequency Distributions of Maximal Repeats Extracted from Tagged Whole Genomic Sequences. Bioinformatics in the Era of Post Genomics and Big Data. IntechOpen. Pages:71-91. Doi:10.5772/intechopen.75113.
2. Wang J D (2016) Extracting significant pattern histories from timestamped texts using MapReduce. The Journal of Supercomputing. 72(8):3236-3260. Doi:10.1007/s1127-016-1713-z.
3. Wang C T (2017) Method for extracting maximal repeat patterns and computing frequency distribution tables. Patent No: US 2017/0255634 A1.
4. Wang J D and Hwang M C (2017) A novel approach to extract significant patterns of travel time intervals of vehicles from freeway gantry timestamp sequences. Applied Sciences. 7(9):878. Doi:10.3390/app7090878.

Biography

Jing Doo Wang obtained his BS Degree in Computer Science and Information Engineering from the Tatung Institute of Technology, Taiwan (1989); MS and PhD Degree in Computer Science and Information Engineering from the National Chung Cheng University in 1993 and 2002 respectively. He has been associated with Asia University since 2003, where he is currently an Associate Professor in the Department of Computer Science and Information Engineering, and also holds a joint appointment with the Department of Bioinformatics and Medical Engineering. His research interests lies in the areas of bioinformatics, text mining for trend analysis, class ambiguity analysis and cloud computing. He has developed a scalable approach to extract maximal repeat patterns from a huge amount of sequential data via MapReduce programming model. He has to his credit USA patent published (US-2017-0255634-A1) on the topic entitled “Method for extracting maximal repeat patterns and computing frequency distribution tables”.

jdwang@asia.edu.tw