

Multi-Omics Data Integration: A Modular Approach

MNV Prasad Gajula*

Institute of Biotechnology, Professor Jayashankar Telangana State Agricultural University, Rajendra Nagar, Hyderabad 500030, India

Abstract

The concept of data integration is critical to the advancement of agricultural and biomedical research in India, particularly due to an uniformity in the practices that depended upon on multi-disciplinary and multi-technological aspects. So far, no successful attempts were reported in this aspect especially at national level. This concept note briefly outlines one of the modules that were proposed to develop under multi-omics approach.

Keywords: Biomedical; Molecular; Genetics; Proteins; Nutritional

Introduction

Substantial extent of research has been carried out over past few decades to understand the molecular and genetic level information within the cell that help improving the methodologies in the fields of agricultural or medical research. However, often the research is focused at one particular area of expertise by an individual researcher or a work group and very less number of attempts were made to cross compare the experimental data with the data obtained from other platforms. Moreover, these technologies require specific expertise for data generation and analysis, and it is quite rare to find experiments that are performed using all of these technologies on the same sample. As it is already known that genes, transcripts, proteins, metabolites and all the processes performed in cells form a complex system that requires each of these and many other components to function collaboratively [1]. However, based on current technologies or methodologies direct multi-omics integration either at experimental level/ data analysis level is not easy [2].

In the proposed study, an attempt will be made to develop separate modules that include standards, methods and web tools pertaining to plants/animals/humans etc. in order to integrate the data e.g., on gene sequence, protein structural and metabolite information respectively. As a step forward in this process, a plant module for metabolomics data is being proposed. However, it is very challenging as there is no standard approach is yet followed by the plant metabolomics community. Even within the single metabolomics research group, it is identified that each specific user follows different approach right from the sample extraction till the metabolite identification. Therefore it is necessary to build the module as generic as possible that covers all the aspects of metabolomics information. During the development of the plant module it is identified that it is complicated to develop the plant data model together with nutritional or cell studies. Therefore, it is suggested to develop the modules separately on individual basis as generic as possible keeping an attention on common features that may exists directly or indirectly from each module. It is then essential to integrate these modules with the core system. Then the model is easily extended by selecting the specific module within the 'Bio-source type' of the Bio-source that acts as a gateway for the different modules (Figure 1).

All the necessary information for constructing the base models for each module can be collected by means of interviews with the experimentalists and literature search. Preexisting data sources and studies provide more information to build strong domain knowledge [3-12]. And any other such models that were being developed by other user communities were also taken into consideration. It is suggested to use Model-View-Controller based framework [5-12]. It simplifies the

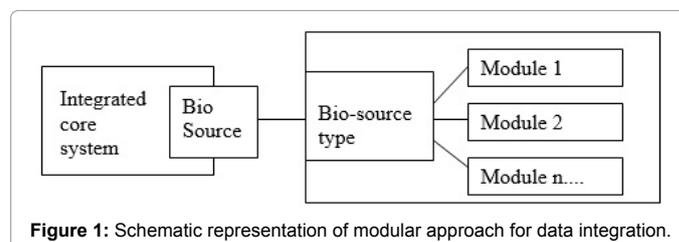


Figure 1: Schematic representation of modular approach for data integration.

development of web applications by reducing the need for configuration files and by generating a lot of the things needed in a database-backed web application.

Acknowledgments

Science and Engineering Research Board (SERB), DST, India is highly acknowledged for the support through Ramanujan fellowship project award no. SR/S2/RJN-22/2011. Sincere thanks to all the members of NMC for extensive discussions and for their efforts in metabolomics research. Sincere thanks to Dr. C. Cheralu, the director of IBT for the extended support.

References

1. Anwar N, Huntz E, Kolch W, Pitti A (2010) Semantic data integration for *Francisella tularensis* novicida. *Proteomic and Genomic Data*.
2. Gajula MNVP, Rai A (2012) It's time to integrate multi Omics data to understand real biology. *Int J Sys Algorithms & App* 2: 31-34.
3. Kumar A, Kumar S, Kumar U, Suravajhala P, Gajula MN (2016) Functional and structural insights into novel DREB1A transcription factors in common wheat (*Triticum aestivum* L): A molecular modeling approach. *Comput Biol Chem* 64: 217-226.
4. Gajula MNVP, Vogel KP, Rai AF, Dietrich S, Steinhoff HJ (2013) How far *in-silico* computing meets real experiments. A study on the structure and dynamics of spin labeled vinculin tail protein by molecular dynamics simulations and EPR spectroscopy. *BMC Genomics* 14(S2): 1-11.
5. Bhawna G, Chaduvula PK, Bhawna Bonthala VS, Manjusha V, Siddiqi EA, et al. (2015) CmMDb: A versatile database for *Cucumis melo* microsatellite markers and other horticulture crop research. *Plos One* 10(4): e0118630.
6. Nagar L, Kumar A, Vimala Y, Gajula MNVP (2016) Sequence to structure

*Corresponding authors: Dr. MNV Prasad Gajula, Institute of Biotechnology, Professor Jayashankar Telangana State Agricultural University, Rajendra Nagar, Hyderabad 500030, India, Tel: 04024015011; E-mail: gajula.ibt@gmail.com

Received October 25, 2016; Accepted November 17, 2016; Published November 21, 2016

Citation: MNV Prasad Gajula (2016) Multi-Omics Data Integration: A Modular Approach. *J Mol Genet Med* 10: 232 doi:10.4172/1747-0862.1000232

Copyright: © 2016 MNV Prasad Gajula. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

-
- analysis of DOPA protein from *Mucuna pruriens*: A computational biology approach. Int J Emerg Trends Technol Comp Sci 2(8): 3083-3089.
7. Ramprasad E, Rakesh G, Rani CVD, Vanisri S, Prasad Gajula MNV (2016) *In silico* analysis of chalcone synthase 1 protein sequences from different plant species. Int J Sci Environ Technol 5(4): 1968-1979.
 8. Prasad Gajula MNV (2016) Displacement of the tyrosyl radical in RNR enzyme: A sophisticated computational approach to analyse experimental data - Conference series paper.
 9. Prasad Gajula MNV (2014) Role of magnetic resonance methods in modern biology: A computational biology perspective. In: Biotechnology Volume 6: Bioinformatics and computational biology, Studium Press LLC, USA.
 10. Bhawna Bonthala VS, Prasad Gajula MNV (2016) PvTFDB: A *Phaseolus vulgaris* transcription factors database for expediting functional genomics in Legumes. Oxford Database Journal 2016: 114.
 11. Gajula MNVP, Soni G, Babu G, Rai A, Bharadwaj N (2013) Molecular interaction studies of shrimp antiviral protein, PmAV with WSSV RING finger domain *in-silico*. J Appl Bioinform Comput Biol.
 12. Gajula MNVP (2008) Computer simulation meets experiment: Molecular dynamics simulations of spin labeled proteins. Osnabrueck, University of Osnabrueck.