Crimson Publishers
Wings to the Research

Review Article

Integration of Bioinformatics Approaches and Experimental Multi-Omics Studies to Support Personalized Medicine

Eugenio Del Prete[1,2] and Angelo Facchiano[1]*

[1]Institute of Food Science, Italy

[2]Institute for Applied Mathematics "Mauro Picone", Italy

ISSN: 2689-2707

***Corresponding author:** Angelo Facchiano, Institute of Food Science, Italy

Submission: March 4, 2020

Published: November 09, 2020

Volume 2 - Issue 4

How to cite this article: Eugenio Del Prete and Angelo Facchiano. Integration of Bioinformatics Approaches and Experimental Multi-Omics Studies to Support Personalized Medicine. Trends Telemed E-Health 2(4). TTEH. 000542. 2020. DOI: 10.31031/TTEH.2020.02.000542

Abstract

The integration of computational approaches, omics and multi-omics studies, and bioinformatics resources offers great opportunities for clinical research. Biomarkers discovery for novel diagnostics devices, drug development and personalized medicine are the novel target of research in medicine. The effectiveness of their integration is the key to success of the future research in medicine.

Keywords: Clinical bioinformatics; Translational bioinformatics; Medical informatics; Multi-omics data; Data integration, Clinical metadata

Introduction

Emerging techniques in molecular biology open grand challenges in the development of new diagnostic procedures. Omics techniques include the capability of analysing a large number of molecules with a single analysis, and continuous development of technology offers the chance to perform the analysis at low cost, so that it becomes possible the screening of large number of patients with a given pathology, looking for novel biomarkers [1-4]. Integration of experimental studies, computational approaches, and bioinformatics tools represents an effective processing of the omics information, especially for multi-omics studies, to investigate complex pathologies [5]. Further integration with clinical data opens to the precision medicine perspective. In this mini review, we describe how the integration of computational and bioinformatics methods can support the investigation on human diseases and drive to the most effective clinical research.

Clinical bioinformatics take advantage of bioinformatics methods and technologies for the elaboration of clinical data. Physicians, clinicians, and other researchers with expertise in the management of patients collect heterogeneous data, with the aim of achieving underlying information about a disease, usually not trivial or easy to obtain with standard analyses. Such structured information can help in studying not only a single pathology, but also other diseases (comorbidities or not) connected with the starting one [6]. The evolution of clinical research involves also the application of bioinformatics and computational biochemistry for drug discovery and development [7]. Virtual screening is a computational approach aimed to screen a database of molecules searching for potential biological activity [8] and it is object of interest in bioinformatics for the development of appropriate tools [9], as well as it represents an active field for the search of molecules of interest for drug development [10-13]. Machine learning approaches have been developed for novel drug discovery [14]. Molecular simulations are strongly based on computational approaches and offer many examples of success in supporting the drug development steps [15,16]. Molecular simulations are also useful in investigating molecular mechanisms underlying pathologies, also in the case of rare diseases, that is of particular interest in our laboratory [17-24]. The low number of patients make difficult to find resources to support studies, so that in many cases rare diseases are also indicated as orphan diseases. In these cases, the bioinformatics and bio-computational approach offers the opportunity to investigate disease as well clinical cases [20,21,23] with

an approach that is an example of personalized medicine. Clinical bioinformatics represents an interface between healthcare data and the disciplines suitable for their analysis (statistics, mathematics, informatics, molecular biology, biochemistry, and so on), essential to cope with the goal of personalized medicine. From a biological point of view, omics data and technologies, together with research strategies (such as cancer research and system biology), are included in the field of clinical bioinformatics [25]. The increasing importance of clinical bioinformatics in medical laboratories is related to the diagnosis of complex disease: the integration of omics data with patients' Electronic Health Record (EHR) helps physicians in improving diagnoses and designing suitable therapies [26].

Many factors contribute in extracting knowledge from stored data, useful in finding new evidences on a given pathology. High-throughput experimental techniques (e.g. Next Generation Sequencing, NGS) provide a huge amount of data [27,28] to be integrated with the clinical information. Discovering new molecular biomarkers complete the knowledge and strengthen the strategy of approaching a particular disease [3-5]. Scientific and structured bibliography reassesses the importance of semantic algorithms for the automated extraction of information from dedicated documents [29,30]. The decrease in costs of technologies, making protocols and machines available for many clinical and research centres. An example of a complete clinical bioinformatics pipeline [31] starts from the collection of the patients by clinical trials, the extraction of biological samples and the conversion in omics data (by means of microarray technologies) provide the material for the in silico post-processing, by the study of the differential expressed genes (DEG), necessary for the discovery of new biomarkers related to a disease.

Translational bioinformatics is strictly connected to clinical bioinformatics, with a particular reference on storage, analysis and interpretation of biomedical data from an informatics point of view, in order to ease all the health management [32]. The cooperation of both fields is targeted to the personalized medicine. Many challenges of the personalized medicine are still under fixing: treating large-scale genomic data, interpreting the effect of variations and the differences in biological functions, creating robust models for complex systems, converting evidences in medical practice, and so on [33]. The 'fourth paradigm', data-intensive science, is oriented towards personalized medicine, since nowadays in silico analyses are feasible all over the world, with affordable computational power and suitable infrastructures, together with the possibility of collaboration among the scientists, and the availability of data and results from public online repositories [34].

From these perspectives, personalized medicine can be encouraged by integrating different kinds of omics data, helping in the prediction of phenotypic outcomes. The main omics areas are genomics, transcriptomics, proteomics and metabolomics. Genomics studies genome biological function, genes distribution on the genome, modifications in their expression, and relationships with biological pathways, towards the increase of the therapeutic efficiency. Clusterization of cells and tissues by expression profiles is based on transcriptomics, such as in single-cell experiments, in order to classify diseases by their similarity (e.g. by means of microarray technology and Polymerase Chain Reaction protocol). Proteomics studies the proteins with their relationships in biological pathways, the modifications between structure and functionality, and the interactions among them, with goals such as drug discovery or discrimination of patients by mass spectrometry data. Metabolomics study the set of metabolites, key regulators in system homeostasis, in specific conditions, with emphasis on changes caused by genetic or environmental variations, analysing the profiles by technologies such as Gas Chromatography Mass Spectrometry (GC-MS) or Nuclear Magnetic Resonance (NMR) spectroscopy [35,36].

Cleaning, integrating and analysing multi-omics datasets are important tasks in improving the personalized medicine and need more and more updated tools and algorithms, to find intra-layer and inter-layer connections among different omics, with references to biological systems and, consequentially, to clinical evidences. Currently, in silico integration methods are mainly divided into unsupervised and supervised methods, with techniques focused on dimensionality reduction, classification, clustering, variable selection and network representation [37-39]. The integration of omics data is helpful in cancer analysis [40,41] and in tissue analysis, even considering imaging data [42]. From a statistical, mathematical and informatics point of view, the work in [43] well explains how the concept of integration can be enlarged in different directions, in terms of omics data (P-integration) or in terms of patients (N-integration), highlighted how the same problem can be afforded by considering different features or the availability of the data. From this perspective, the patient can be an active part of the integration process, with all the information stored in the form of metadata, that is available for a deeper post-processing analysis of the obtained results. It is obvious how the metadata should be standardized, creating an ontology where possible [44,45], not only to ease the availability for the physicians and clinicians, but also for the researchers that need these features, which are important in studies as clustering, classification and outliers discovery. This process has important implications also in the research reproducibility [46].

A clinical bioinformatics pipeline is useful in the analysis of a complex disease. The definition of a complex disease is not trivial, especially because many chronic diseases have not Mendelian behaviour. A representative list for describing a complex disease can be caused by a combination of different factors (genetic, environmental and lifestyle); not simple patterns of inheritance; difficulty in predicting insurgence and transmission; complicated treatment. For example, two patients can have different mutations in their genomes, thus the challenge is to extract the phenotypes and unravel specific casual mutations with association studies [47]. An example of clinical bioinformatics pipeline applied to a complex disease is reported in [48], with the analysis of celiac disease together with some specific comorbidities. In this study, microarray data available online are selected in order to extract

DEGs from transcriptomics data, and Gene Set Enrichment Analysis (GSEA) is performed to connect the most important DEGs to Gene Ontology (GO) terms, extracting the biological process (BP) mainly related to the pathology. Moreover, the GO and Disease Ontology (DO) trees are compared by semantic similarity, to show which datasets (celiac disease or other autoimmune diseases) are more correlated. Finally, the pathways analysis is used to correlated DEGs to Kyoto Encyclopaedia of Genes and Genomes (KEGG) database, to search for biological pathways with strong relationships between celiac disease and its comorbidities.

Novel efforts are applied to omics (and integration) in reducing computational time, providing and storing big data, through modern workflows and pipelines, which cover all the study from the in vivo/in vitro experiment to the in-silico prediction and analysis [49,50]. The standardization of different approaches for the in-silico elaboration is important in each subsequent step (getting and cleaning data, aggregation of data, statistical analysis and validation, presentation of result) [51]. Nevertheless, a cooperation between the 'bio' area (biology, medicine, chemistry) and the 'info' area (mathematics, physics, statistics, informatics, engineering) is compulsory, in terms of knowledge and collaboration among the researchers. Fortunately, many programming languages are clinical/biological data-oriented, providing ad hoc libraries for specific problems: a valuable example is R (and the related Bioconductor repository) [52-54], with lots of libraries conceived for the sake of the reproducible research. For example, the class of data 'Summarized Experiment' [55] was conceived to carry both expression values and patients' metadata, and library such as 'TCGAbiolinks' [56] provides the possibility to download, arrange, analyze, and integrate with clinical information, cancer data from the Genomic Data Commons (GDC) Data Portal online repository [57].

Conclusion

In conclusion, clinical bioinformatics connect bioinformatics approaches to clinical data from patients (and healthy control) in order to extract principal features (e.g. biomarkers) that represent a sort of fingerprint for the subject. Such features are important for the prediction, diagnosis and treatment of the disease, with a view to personalized medicine, towards the last goal of treating every single patient on the base of the abovementioned specific evidences. Integration of this area with telemedicine and e-health services can represent an effective step towards personalized medicine approaches, especially taking into account the possibility to reduce the time in curing the patient, and to help physicians and clinicians in taking decisions using information, which has an automated extraction and a fast availability on electronical devices.

References

1. Subramanian I, Verma S, Kumar S, Jere A, Anamika K, et al. (2020) Multi-omics data integration, interpretation, and its application. Bioinform Biol Insights 14: 1177932219899051.
2. Solovev I, Shaposhnikov M, Moskalev A (2020) Multi-omics approaches to human biological age estimation. Mech Ageing Dev 185: 111192.
3. Simpkins AN, Janowski M, Oz HS, Roberts J, Bix G, et al. (2019) Biomarker application for precision medicine in stroke. Transl Stroke Res 11(4): 615-627.
4. Gan WZ, Ramachandran V, Lim CSY, Koh RY (2019) Omics-based biomarkers in the diagnosis of diabetes. J Basic Clin Physiol Pharmacol 31(2).
5. Arcangelo D, Facchiano F, Nassa G, Stancato A, Antonini A, et al. (2016) PDGFR-alpha inhibits melanoma growth via CXCL10/IP-10: A multi-omics approach. Oncotarget 7(47): 77257-77275.
6. Schwarz E, Leweke FM, Bahn S, Liò P (2009) Clinical bioinformatics for complex disorders: A schizophrenia case study. BMC Bioinformatics 10(Suppl 12): 1-9.
7. Gill SK, Christopher AF, Gupta V, Bansal P (2016) Emerging role of bioinformatics tools and software in evolution of clinical research. Perspect Clin Res 7(3): 115-122.
8. Good A (2007) Virtual screening. Reference in Comprehensive Medicinal Chemistry II 4: 459-494.
9. Glaab E (2016) Building a virtual ligand screening pipeline using free software: A survey. Brief Bioinform 17(2): 352-366.
10. Sousa ACC, Combrinck JM, Maepa K, Egan TJ (2020) Virtual screening as a tool to discover new β-haematin inhibitors with activity against malaria parasites. Sci Rep 10(1): 3374.
11. Dubey A, Facchiano A, Ramteke PW, Marabotti A (2016) In silico approach to find chymase inhibitors among biogenic compounds. Future Med Chem 8(8): 841-851.
12. Dubey A, Marabotti A, Ramteke PW, Facchiano A (2016) Interaction of human chymase with ginkgolides, terpene trilactones of ginkgo biloba investigated by molecular docking simulations. Biochem Biophys Res Commun 473(2): 449-454.
13. Dubey A, Dotolo S, Ramteke PW, Facchiano A, Marabotti A, et al. (2019) Searching for chymase inhibitors among chamomile compounds using a computational-based approach. Biomolecules 9(1): 5.
14. Lima AN, Philot EA, Trossini GH, Scott LP, Maltarollo VG, et al. (2016) Use of machine learning approaches for novel drug discovery. Expert Opin Drug Discov 11(3): 225-239.
15. Wade RC, Ahen OMH (2019) Molecular modeling in drug design. Molecules 24(2): 321.
16. Liu X, Shi D, Zhou S, Liu H, Liu H, et al. (2018) Molecular dynamics simulations and novel drug discovery. Expert Opin Drug Discov 13(1): 23-37.
17. Marabotti A, Facchiano AM (2005) Homology modelling studies on human galactose-1-phosphate uridylyltransferase and on its galactosemia-related mutant Q188R provide an explanation of molecular effects of the mutation on homo- and heterodimers. J Med Chem 48(3): 773-779.
18. d'Acierno A, Facchiano A, Marabotti A (2009) GALT Protein Database, a bioinformatics resource for the management and analysis of structural features of a galactosemia-related protein and its mutants. Genomics Proteomics Bioinformatics 7(1-2): 71-76.
19. Facchiano A, Marabotti A (2010) Analysis of galactosemia-linked mutations of GALT enzyme using a computational biology approach. Proteins Eng Des Sel 23(2): 103-113.
20. Boutron A, Marabotti A, Facchiano A, Cheillan D, Zater M, et al. (2012) Mutation spectrum in the French cohort of galactosemic patients and structural simulation of 27 novel missense variations. Mol Genet Metab 107(3): 438-447.

21. Tang M, Facchiano A, Rachamadugu R, Calderon F, Mao R, et al. (2012) Correlation assessment among clinical phenotypes, expression analysis and molecular modeling of 14 novel variations in the human galactose-1-phosphate uridylyltransferase gene. Hum Mutat 33(7): 1107-1115.

22. d'Acierno A, Facchiano A, Marabotti A (2014) GALT protein database: Querying structural and functional features of GALT enzyme. Hum Mutat 35(9): 1060-1067.

23. Viggiano E, Marabotti A, Burlina AP, Cazzorla C, D'Apice MR, et al. (2015) Clinical and molecular spectra in galactosaemic patients from neonatal screening in northeastern Italy: Structural and functional characterization of new variations in the galactose-1-phosphate uridyltransferase (GALT) gene. Gene 559(2): 112-118.

24. d'Acierno A, Scafuri B, Facchiano A, Marabotti A (2018) The evolution of a web resource: The galactosemia proteins database 2.0. Hum Mutat 39(1): 52-60.

25. Mayer G, Heinze G, Mischak H, Hellemons HE, Heerspink HJ, et al. (2011) Omics-bioinformatics in the context of clinical data. Methods Mol Biol 719: 479-497.

26. Belmont JW, Shaw CA (2016) Clinical bioinformatics: Emergence of a new laboratory discipline. Expert Rev Mol Diagn 16(11): 1139-1141.

27. Kulkarni P, Frommolt P (2017) Challenges in the setup of large-scale next-generation sequencing analysis workflows. Computational and Structural Biotechnology Journal 15: 471-477.

28. Pereira R, Oliveira J, Sousa M (2020) Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. J Clin Med 9(1): 132.

29. Jovanović J, Bagheri E (2017) Semantic annotation in biomedicine: The current landscape. J Biomed Semantics 8(1): 44.

30. Naderi N, Witte R (2012) Automated extraction and semantic analysis of mutation impacts from the biomedical literature. BMC Genomics 13(Suppl 4): S10.

31. Chen H, Wang X (2011) Significance of bioinformatics in research of chronic obstructive pulmonary disease. J Clin Bioinforma 1(1): 35.

32. Bellazzi R, Masseroli M, Murphy S, Shabo A, Romano P, et al. (2012) Clinical bioinformatics: Challenges and opportunities. BMC Bioinformatics 13(Suppl 14): S1.

33. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB, et al. (2011) Bioinformatics challenges for personalized medicine. Bioinformatics 27(13): 1741-1748.

34. Hey T, Tansley S, Tolle K (2009) The fourth paradigm: Data-intensive scientific discovery. Microsoft Research, Redmond, Washington, USA.

35. Debnath M, Prasad GBKS, Bisen PS (2010) Molecular diagnostics: Promises and possibilities. Springer, Netherlands.

36. Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, et al. (2016) Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. Brief. Bioinform 19(2): 286-302.

37. Huang S, Chaudhary K, Garmire LX (2017) More is better: Recent progress in multi-omics data integration methods. Front Genet 8: 84.

38. Ang JC, Mirzal A, Haron H, Hamed HNA (2016) Supervised, unsupervised and semi-supervised feature selection: A review on gene selection. IEEE/ACM Trans Comput Biol Bioinforma 13(5): 971-989.

39. Buescher J M, Driggers EM (2016) Integration of omics: More than the sum of its parts. Cancer Metab 4(1): 1-8.

40. Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19(5): 299-310.

41. Serra A, Fratello M, Fortino V, Raiconi G, Tagliaferri R, et al. (2015) MVDA: A multi-view genomic data integration methodology. BMC Bioinformatics 16(1): 1-13.

42. Disselhorst JA, Krueger MA, Ud Dean SMM, Bezrukov I, Jarboui MA, et al. (2018) Linking imaging to omics utilizing image-guided tissue extraction. Proc Natl Acad Sci 115(13): E2980-E2987.

43. Rohart F, Gautier B, Singh A, Lê Cao KA (2017) Mixomics: An R package for omics feature selection and multiple data integration. PLoS Comput Biol 13(11): e1005752.

44. Kim H, Park Y, Lee K, Song Y, Kim J, et al. (2019) Clinical metadata ontology: A simple classification scheme for data elements of clinical data based on semantics. BMC Med Inform Decis Mak 19(1): 166.

45. Papatheodorou I, Crichton C, Morris L, Maccallum P, METABRIC Group, et al. (2009) A metadata approach for clinical data management in translational genomics studies in breast cancer. BMC Medical Genomics 2: 66.

46. Kolker E, Özdemir V, Martens L, Hancock W, Anderson G, et al. (2014) Toward more transparent and reproducible omics studies through a common metadata checklist and data publications. Omics 18(1): 10-14.

47. Stessman HA, Bernier R, Eichler EE (2014) A genotype-first approach to defining the subtypes of a complex disease. Cell 156(5): 872-877.

48. Del Prete E, Facchiano A, Liò P (2020) Bioinformatics methodologies for coeliac disease and its comorbidities. Briefings in Bioinformatics 21(1): 355-367.

49. Gomez CD, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M, et al. (2014) Data integration in the era of omics: Current and future challenges. BMC Syst Biol 8(Suppl 2): I1.

50. Gedela S (2011) Integration, warehousing, and analysis strategies of Omics data. Methods in Molecular 719: 399-414.

51. Kohl M, Megger DA, Trippler M, Meckel H, Ahrens M, et al. (2014) A practical data processing workflow for multi-OMICS projects. Biochim. Biophys. Acta-Proteins Proteomics 1844(1): 52-62.

52. Wanichthanarak K, Fahrmann JF, Grapov D (2015) Genomic, proteomic, and metabolomic data integration strategies. Biomark Insights 10(Suppl 4): 1-6.

53. Russo F, Righelli D, Angelini C (2016) Advantages and limits in the adoption of reproducible research and R-Tools for the analysis of omic data. In: Angelini C, Rancoita PMV, et al. (Eds.), Computational intelligence methods for bioinformatics and biostatistics. Springer International Publishing Group, Switzerland, pp. 245-258.

54. Peng RD (2011) Reproducible research in computational science. Science 334(6060): 1226-1227.

55. Morgan M, Obenchain V, Hester J, Pagès H (2019) Summarized experiment: Summarized experiment container.

56. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, et al. (2019) New functionalities in the TCGA biolinks package for the study and integration of cancer data from GDC and GTEx. PLoS Computational Biology 15(3): e1006701.

57. Jensen MA, Ferretti V, Grossman RL, Staudt LM (2017) The NCI genomic data commons as an engine for precision medicine. Blood 130(4): 453-459.