

OPINION ARTICLE

Open Access



Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis

Zeeshan Ahmed^{1,2}

Abstract

Precision medicine aims to empower clinicians to predict the most appropriate course of action for patients with complex diseases like cancer, diabetes, cardiomyopathy, and COVID-19. With a progressive interpretation of the clinical, molecular, and genomic factors at play in diseases, more effective and personalized medical treatments are anticipated for many disorders. Understanding patient's metabolomics and genetic make-up in conjunction with clinical data will significantly lead to determining predisposition, diagnostic, prognostic, and predictive biomarkers and paths ultimately providing optimal and personalized care for diverse, and targeted chronic and acute diseases. In clinical settings, we need to timely model clinical and multi-omics data to find statistical patterns across millions of features to identify underlying biologic pathways, modifiable risk factors, and actionable information that support early detection and prevention of complex disorders, and development of new therapies for better patient care. It is important to calculate quantitative phenotype measurements, evaluate variants in unique genes and interpret using ACMG guidelines, find frequency of pathogenic and likely pathogenic variants without disease indicators, and observe autosomal recessive carriers with a phenotype manifestation in metabolome. Next, ensuring security to reconcile noise, we need to build and train machine-learning prognostic models to meaningfully process multisource heterogeneous data to identify high-risk rare variants and make medically relevant predictions. The goal, today, is to facilitate implementation of mainstream precision medicine to improve the traditional symptom-driven practice of medicine, and allow earlier interventions using predictive diagnostics and tailoring better-personalized treatments. We strongly recommend automated implementation of cutting-edge technologies, utilizing machine learning (ML) and artificial intelligence (AI) approaches for the multimodal data aggregation, multifactor examination, development of knowledgebase of clinical predictors for decision support, and best strategies for dealing with relevant ethical issues.

Keywords: Precision medicine, Clinics, Genomics, Metabolomics, Integrative analysis, Artificial intelligence, Machine learning

Background

Since the beginning of scientific discoveries, it has been central to understand the cause of disease and senescence [1]. Pain is one of the key triggers for patients to

seek diagnosis and treatment. However, when dealing with some of the life-threatening diseases, patients may not feel pain. To identify and help patients with known diseases and symptoms, and those heading toward late stages of novel infectious (e.g., COVID-19), chronic (e.g., diabetes, heart disease), acute (e.g., flu, stroke, heart attack), and complex (e.g., cancer) diseases, it is essential to provide timely personalized treatment [2–10]. Our evolving understanding of the complex nature has led us

Correspondence: zahmed@ifh.rutgers.edu

¹Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick, NJ, USA

²Department of Medicine, Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences, 125 Paterson Street, New Brunswick, NJ, USA



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

to realize that to effectively diagnose and treat patients with these conditions, it is essential to provide personalized utilize a precision medicine approach [10]. Progress in the molecular technology developments have led to vast amounts of human health-related data that are expected to greatly expand our understanding of human biology and health, and to drive personalized medicine. We hypothesize that on-demand access and analysis of clinical, genetic, and metabolic data will align biomarker identification with treatment windows necessary for real-time personalized care and enhance prediction of potential disease risks [11]. Despite current advancements, there is still no platform available that can efficiently integrate clinical, multi-omics, and epidemiological data acquisition, and enable effective management of data analytics with a user-friendly physician-oriented clinical interface [12, 13]. Platforms like The Cancer Genome Atlas (TCGA) [14] provide a great resource for scientific data (i.e., genomics or epigenetics sequence data) but offer limited capacity for clinical information, because they are not directly integrated to clinical health systems like Epic, NextGen, and Cerner etc. The inability of disparate platforms to effectively integrate is largely due to the high volume and heterogenous nature of the different types of data they

contain, which is acquired from variable sources, each with unique data structures. It is essential to address a major gap in developing precision diagnostics and therapeutic agents in healthcare by establishing a digital solution for practicing precision medicine. Intelligent big data platforms are necessary to improve the quality of care-delivery process by increasing permeation of electronic health record (EHR) systems into clinical environments, focusing on predictive diagnosis, enabling real-time telemedicine, and precise treatment resulting in lower spending on life-threatening complex and chronic diseases [15].

Practicing precision medicine and AI

Precision medicine has the potential to improve the traditional symptom-driven practice of medicine, and allow earlier interventions using predictive diagnostics and tailoring better-personalized treatments [2–9, 16, 17]. However, practicing precision medicine is not straightforward, as significant efforts are required from the experts in multidisciplinary sciences. This necessitates development of progressive healthcare environment that will enable clinicians and researchers to gain a complete picture of the patient to deepen their understanding, using additional details from healthcare and

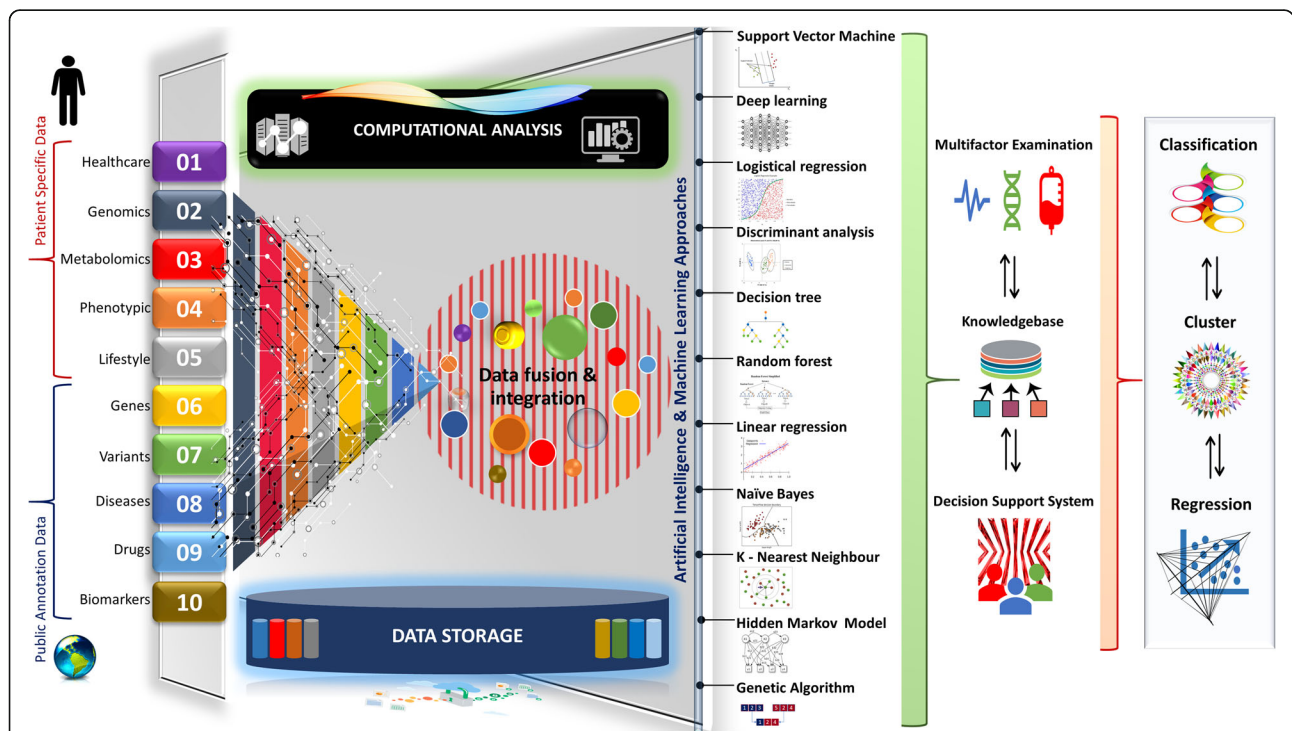


Fig. 1 Design modeling of heterogenous patient-specific healthcare, genomics, metabolomics, phenotypic, and lifestyle data, and publicly available annotation data including genes, variants, diseases, drugs, and biomarkers. Analysis using AI and ML approaches (Support Vector Machine, Deep Learning, Logistic Regression, Discrimination Analysis, Decision tree, Random Forest, Linear Regression, Naïve Bayes, K-Nearest Neighbor, Hidden Markov Model, and Genetic Algorithm), multifactor examination, knowledgebase and decision support system for data classification, cluster, and regression analysis. Furthermore, resource allocation for data storage and computational analysis

multi-omics data. We hypothesize that clinical information will enrich genomics and metabolomics data such that combined predictors will perform better than the individual classifiers only based on either genomics, metabolomics, or clinical data. Practically supporting the hypothesis, we need to design research methodology, which includes modeling of patient-specific (healthcare, genomics, metabolomics, proteomics, and lifestyle) and publicly available annotation (genes, variants, diseases, drugs, biomarkers) data storage, management, integration, knowledgebase creation, and analysis using different artificial intelligence (AI) and machine learning (ML) approaches (Fig. 1) [18, 19].

Precision medicine is moving forward but with many challenges that require addition of useful analytic tools, technologies, databases, and approaches to efficiently manage massive heterogeneous data, augment networking and interoperability of clinical, laboratory, and public health systems. A major barrier to implementation of precision medicine is the data analysis requirement. Most of the precision medicine efforts today are manual or semi-automated, time-consuming, and unable to facilitate on-demand analysis of diverse human datasets to impact critical treatment windows and predict potential disease risks [19–27]. The traditional way of computational analysis is based on running a series of command-

line applications, which require good programming skills and ability to work in the UNIX environment. It hinders linking information generated at different stages of treatments and experiments conducted at levels of sampling, sequencing, and analysis. While precision medicine analyses require complex coordinated efforts between disparate groups with non-aligned data formats and massive amounts of computing time that is essential in many cases to positively impact treatment outcomes. Furthermore, it is difficult but mandatory to address ethical and social issues related to healthcare data collection, privacy, and protection with effective balance [18, 28]. Further, current potential pitfalls are given in attached Table 1.

The efficient use of information technology, data science, and AI has the potential to enhance public health surveillance and tracking, with systematic collection, management, analysis, and interpretation of data within accelerated timelines [6, 19, 29–31]. We need detailed bioinformatics and AI platforms for supporting real-time processes involved in multisource heterogeneous raw data generation, mathematical modeling, computational analysis, data fusion, integration, management, and visualization (Fig. 1). Platforms need to be user-friendly, multi-functional, and multi-roles-based to address complex and big data-oriented problems in clinical settings.

Table 1 Current potential pitfalls

Number	Potential pitfalls
1	Uneven distribution of informatics resources.
2	Integration of biomedical data located among heterogeneous sources.
3	Hazards in dehumanization of healthcare data.
4	Handling of extensively available irrelevant, error prone, and missing data.
5	Intelligent and user-friendly interface development.
6	Applying regulations and policies for data collection, usage and sharing.
7	Harmonizing big data with the definitions of clinical phenotypes and diagnosis.
8	Inflexible EHR database schemas not geared for precision medicine.
9	Lack of data availability on social determinants of health.
10	Unstandardized genomics tools and modifications in their versions and outcome format.
11	Overloaded Data generated during unnecessary follow-up diagnoses and treatments.
12	Augmented computational complexity with increasing number of attributes.
13	Slow SQL based high volume data processing speed.
14	Determining optimal parameters and understanding structures of AI and ML algorithms.
15	Handling continuous explanatory variables with more than two levels and understanding odds and probabilities in AI and ML algorithms.
16	Possibility of too many overfitting attributes in AI and ML algorithms.
17	Handling redundant attributes, distribution of statistically independent attributes, and management of class frequencies affecting accuracy.
18	Reduced evidence and reproducibility.
19	Correct predictor variables selection, and evidence-based observational data analysis and screening.
20	Gaining confidence of clinicians at AI produced results.
21	Ethical and social issues related to healthcare data collection, privacy and protection.

It will support categorizing interaction patterns among variables, learning from experiences, and strategizing and predicting better orientations. Multiple AI and ML algorithms (e.g., Support Vector Machine, Deep Learning, Logistic Regression, Discrimination Analysis, Decision tree, Random Forest, Linear Regression, Naïve Bayes, K-Nearest Neighbor, Hidden Markov Model, and Genetic Algorithm, etc.) are available for multifactor examination, scientific knowledge extraction, and decision support system (Fig. 1) [32–36]. However, determining which AI and ML approaches to use for which task is a challenge in itself [37]. We suggest classifying tasks based on the available predictor variables, as a key to correctly address this problem. Best fitting use of ML and AI algorithms have the potential to predict the existence of life-threatening diseases risk susceptibility, starting from the most common to rare among the population data [19].

AI has the ability to improve identification of relevant variables for patient data stratification with timely detection of statistical patterns across millions of features to identify conditions that are likely to manifest later and discover modifiable risk factors that support the best utilization of known therapies [38]. Impactful and automated implementation of AI and ML can elevate investigating correlation and overlapping of reported diagnoses of a patient in clinical data, and assess genotype and phenotype associations among various diseases to find potential indistinct results for patient care from highly expressed genes and disease-causing variants [9, 39]. Understanding how genetic variations contribute to health is one important aspect of precision medicine, where additional approaches involve measuring levels of proteins and metabolic products. By harnessing the power of metabolomics, we need to profile a patient's metabolome and correlate it with their body mass index (BMI). Further, AI can assist in finding metabolite penetrance using listed features and abnormalities, and analyzing biochemical pathways in metabolites [40–42] with patterns of multimodal distributions for candidate genes [10, 43].

Conclusions

The scientific approach would be to perform analysis of individual genomes giving rise to a new form of preventive and personalized medicine in healthcare. Availability of gene-based designer drugs, precise targeting of molecular fingerprints for tumor, appropriate drug therapy, predicting individual susceptibility to disease, diagnosis, and treatment of mental illness are all a few of the many transformations expected in the decade to come. Precision medicine will timely enable clinicians to integrate healthcare data with targeted assays and tests to identify and assess disease biomarkers and risks, determine

actionable genetic variants in patients, obtain the entire picture of the metabolome, and map metabolites to disease pathways.

Abbreviations

AI: Artificial intelligence; BMI: Body mass index; COVID-19: Coronavirus disease 2019; EHR: Electronic health record; HIPAA: Health Insurance Portability and Accountability Act; ML: Machine learning; TCGA: The Cancer Genome Atlas

Acknowledgements

We appreciate great support by the Institute for Health, Health Care Policy and Aging Research (IFH), and Rutgers Robert Wood Johnson Medical School, Rutgers Biomedical and Health Sciences at the Rutgers, The State University of New Jersey.

Author's contributions

ZA conceived of the presented idea and wrote the manuscript. The author read and approved the final manuscript.

Authors' information

ZA is Assistant Professor of Medicine and Core Faculty Member at the Institute for Health, Health Care Policy and Aging Research (IFH), and the Department of Medicine, Rutgers Robert Wood Johnson Medical School (RWJMS), Rutgers Biomedical and Health Sciences (RBHS), at Rutgers, The State University of New Jersey.

Funding

This work was supported by the Institute for Health, Health Care Policy and Aging Research, and Robert Wood Johnson Medical School, at Rutgers, The State University of New Jersey.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The author declares that they have no competing interests.

Received: 2 April 2020 Accepted: 15 September 2020

Published online: 02 October 2020

References

1. Zeeshan S, Xiong R, Liang BT, Ahmed Z. 100 years of evolving gene–disease complexities and scientific debutants. *Brief Bioinform.* 2019;21(3):885–905.
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015;372:793–5.
3. Venter JC, Smith HO, Adams MD. The sequence of the human genome. *Clin Chem.* 2015;61:1207–8.
4. Long T, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet.* 2017;49:568–78.
5. Guo L, et al. Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *Proc Natl Acad Sci USA.* 2015;112(35):E4901–10.
6. Khoury MJ, Iademarco MF, Riley WT. Precision public health for the era of precision medicine. *American journal of preventive medicine.* 2016;50(3):398–401.
7. Lazaridis KN, et al. Implementing individualized medicine into the medical practice. *Am J Med Genet C Semin Med Genet.* 2014;166C:15–23.
8. Bakris G, Sorrentino M. Redefining hypertension — assessing the new blood pressure guidelines. *N Engl J Med.* 2018;378:497–9.
9. Hou YC, et al. Precision medicine integrating whole-genome sequencing, comprehensive metabolomics, and advanced imaging. *Proceed Natl Acad Sci U S A.* 2020;117(6):3053–62.
10. Ahmed Z, Zeeshan S, Foran DJ, Kleinman LC, Wondisford FE, Dong X. Integrative Clinical, Genomics and Metabolomics Data Analysis for

- Mainstream Precision Medicine to Investigate COVID-19. *BMJ Innov.* 2020. Published Online First: 04 September 2020. <https://doi.org/10.1136/bmjinnov-2020-000444>.
11. Ahmed Z, Kim M, Liang BT. MAV-clic: management, analysis, and visualization of clinical data. *JAMIA Open.* 2019;2:23–8.
 12. Sboner A, Elemento O. A primer on precision medicine informatics. *Brief Bioinform.* 2016;17:145–53. <https://doi.org/10.1093/bib/bbv032>.
 13. Karczewski KJ, Snyder MP. Integrative omics for health and disease. *Nat Rev Genet.* 2018;19(5):299–310. <https://doi.org/10.1038/nrg.2018.4>.
 14. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Poznan, Poland)*, 2015;19(1A):A68–A77. <https://doi.org/10.5114/wo.2014.47136>.
 15. Ahmed Z, et al. Human gene-disease associations for clinical-genomics and precision medicine research. *Clin Transl Med.* 2020;2020:1–22.
 16. Schüssler-Fiorenza RSM, Contrepois K, Moneghetti KJ, et al. A longitudinal big data approach for precision health. *Nat Med.* 2019;25:792–804. <https://doi.org/10.1038/s41591-019-0414-6>.
 17. Perkins BA, Caskey CT, Brar P, et al. Precision medicine screening using whole-genome sequencing and advanced imaging to identify disease risk in adults. *Proc Natl Acad Sci U S A.* 2018;115(14):3686–91. <https://doi.org/10.1073/pnas.1706096114>.
 18. Bradford W, Hurdle JF, LaSalle B, Facelli JC. Development of a HIPAA-compliant environment for translational research data and analytics. *Journal of the American Medical Informatics Association: JAMIA.* 2014;21(1):185–9.
 19. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database.* 2020;2020:baaa010.
 20. Lee SJ, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nature communications.* 2018;9(1):42.
 21. Van Panhuis WG, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health.* 2014;14:1144.
 22. Beltran H, et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA oncology.* 2015;1(4):466–74.
 23. Luo Y, Ahmad FS, Shah SJ. Tensor factorization for precision medicine in heart failure with preserved ejection fraction. *J Cardiovasc Transl Res.* 2017;10:305–12.
 24. Katsanis N. The continuum of causality in human genetic disorders. *Genome Biol.* 2016;17:233.
 25. Manrai AK, Ioannidis JP, Kohane IS. Clinical genomics: from pathogenicity claims to quantitative risk estimates. *JAMA.* 2016;315:1233–4.
 26. Boyle EA, Li Yi, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell.* 2017;169:1177–86.
 27. Shieh Y, et al. Breast cancer screening in the precision medicine era: risk-based screening in a population-based trial. *J Natl Cancer Inst.* 2017;109:djw290.
 28. Feero WG. Clinical application of whole-genome sequencing: proceed with care. *JAMA.* 2014;311:1017–9.
 29. Khoury MJ, Ioannidis JP. Medicine. Big data meets public health. *Science.* 2014;346(6213):1054–5.
 30. Bali J, Garg R, Bali RT. Artificial intelligence (AI) in healthcare and biomedical research: why a strong computational/AI bioethics framework is required? *Indian journal of ophthalmology.* 2019;67:3–6.
 31. Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *The British journal of general practice: the journal of the Royal College of General Practitioners.* 2018;68:143–4.
 32. Shah P, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med.* 2019;2:69.
 33. Rajkomar A, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med.* 2018;1:18.
 34. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 2015;33:831–8.
 35. Esteva A, et al. A guide to deep learning in healthcare. *Nat. Med.* 2019;25:24–9.
 36. Norgeot B, Glicksberg BS, Butte AJ. A call for deep-learning healthcare. *Nat. Med.* 2019;25:14–5.
 37. Kelly CJ, Karthikesalingam A, Suleyman M, Carrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17:195.
 38. Ahmed Z, Zeeshan S, Xiong R, Liang BT. Debutant iOS app and gene-disease complexities in clinical genomics and precision medicine. *Clin Transl Med.* 2019;8:26.
 39. Marouli E, et al. Rare and low-frequency coding variants alter human adult height. *Nature.* 2017;542(7640):186–90.
 40. Dandekar T, Fieselmann A, Majeed S, Ahmed Z. Software applications toward quantitative metabolic flux analysis and modeling. *Brief Bioinform.* 2014;15:91–107.
 41. Ahmed Z, et al. Software LS-MIDA for efficient mass isotopomer distribution analysis in metabolic modelling. *BMC Bioinformatics.* 2013;14:218.
 42. Ahmed Z, et al. "Isotopo" a database application for facile analysis and management of mass isotopomer data. *Database.* 2014;2014:bau077.
 43. Tonn MK, Thomas P, Barahona M, Oyarzún DA. Stochastic modelling reveals mechanisms of metabolic heterogeneity. *Communications biology.* 2019;2:108.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

