HHS Public Access

Author manuscript

Pediatr Res. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

Pediatr Res. 2023 January ; 93(2): 308–315. doi:10.1038/s41390-022-02181-x.

Multiomics, Artificial Intelligence and Precision Medicine in Perinatology

Mohan Pammi[1,*], **Nima Aghaeepour**[2], **Josef Neu**[3]

[1]Section of Neonatology, Department of Pediatrics Baylor College of Medicine and Texas Children's Hospital Houston, Texas.

[2]Departments of Anesthesiology, Pediatrics, and Biomedical Data Sciences, Stanford University School of Medicine, Stanford, California.

[3]Section of Neonatology, Department of Pediatrics, University of Florida Gainesville, Florida.

Abstract

Technological advances in omics evaluation, bioinformatics and artificial intelligence have made us rethink ways to improve patient outcomes. Collective quantification and characterization of biological data including genomics, epigenomics, metabolomics and proteomics is now feasible at low cost with rapid turnover. Significant advances in the integration methods of these multi-omics datasets by machine learning promises us a holistic view of disease pathogenesis and yield biomarkers for disease diagnosis and prognosis. Using machine learning tools and algorithms, it is possible to integrate multiomics data with clinical information to develop predictive models that identify risk before the condition is clinically apparent, thus facilitating early interventions to improve the health trajectories of the patients. In this review, we intend to update the readers on the recent developments related to the use of artificial intelligence in integrating multiomic and clinical datasets in the field of Perinatology, focusing on neonatal intensive care and the opportunities for precision medicine. We intend to briefly discuss the potential negative societal and ethical consequences of using artificial intelligence in healthcare. We are poised for a new era in medicine where computational analysis of biological and clinical datasets will make precision medicine a reality.

Introduction

Technological advances have changed the way we live and work. Today, a device worn around our wrists carries more computing capability than devices that required large, cooled rooms in the early 1950's. The emergence of self-driving vehicles, facial and

[*]**Corresponding author**: Mohan Pammi MD, PhD, MRCPCH, Section of Neonatology, Department of Pediatrics, Baylor College of Medicine, Houston, USA, mohanv@bcm.edu, Tel: 832-824-3206 Fax: 832-825-2799.

Author contributions

MP and JN wrote the initial draft of the manuscript. NA made revisions and provided critical intellectual input for the review

Disclosure statement: We have no conflicts of interest to disclose.

Consent Statement: Patient consent was not required.

voice recognition and the integration of human cognition with the rapid processing of artificial intelligence (AI) are all examples of these advances. Newly developed AI strategies in radiology, retinal scanning, diagnosis of skin lesions, and interpretation of normal or abnormal cardiac rhythmicity have improved diagnostic capabilities that exceeds those of even the most skilled clinicians[1–3].

Concomitant to and aided by the development of AI is the emergence of the various “omic” disciplines. The sequencing of the first human genome during the 1990s took nearly a decade to accomplish and cost nearly a billion dollars. This task can now be done in less than a day for slightly over 100 US dollars. As described by Euan Angus Ashely, in his book “the Genome Odyssey,” the relative cost of having one’s genome sequenced compared to a couple of decades ago can be equated to being able to buy a new Ferrari automobile for a few coins[4]. Other omics and non-omics data can now be integrated with genomic data using various bioinformatic platforms melding nature (the genome) with nurture (the effects of the environment) as a means to explain the mechanisms of normal physiology and what goes awry during disease[5].

History has helped us discern three eras of medicine (Figure 1). We have gone through an era of “intuition-based medicine,” which has relied largely on what is known about signs and symptoms, pathophysiology of disease, the experience of the clinician and their diagnostic and therapeutic skills. Currently, we strive to make judgements using “evidence-based medicine” where diagnostics and therapeutics are based on evidence from clinical studies and gaussian statistics. More recently, there has been a drive to refocus our efforts toward prevention rather than treatment of illnesses and health disorders. Prevention requires early detection of features both clinical and laboratory that can be used to minimize adverse consequences associated with a developing illness. An example related to neonatal nutrition would be a paradigm shift from following growth curves and acting when growth failure is detected (delayed intervention) to identification of biomarkers that could predict growth failure even before it occurs. Biomarkers could inform individualized pre-emptive strategies that prevent even the beginnings of growth faltering seen on growth curves[6].

The value of AI and multiomics for studying various aspects of normal and pathologic pregnancy and the newborn period is being increasingly recognized [7–9]. In this review, we intend to discuss recent research integrating multiomics using machine learning tools in the field of Perinatology. Not only will we discuss the technical advances but will discuss ethical and societal considerations that should be addressed.

Description of Multiomics

The term ‘omics’ describes the collective characterization and quantification of large datasets including the genome, transcriptome, proteome, microbiome and epigenome that influence the structure, function, and dynamics of a biological process[5,10]. Recent biotechnological advances have enabled researchers to generate systems-level profiling of patients at multiple omics levels with increasing dimensionality (Figure 2). Table 1 shows some of the types of “omics” that are currently being explored in clinical and basic science research.

Many of these “omics” have been studied individually (single omics) and correlated to phenotypic data under different stressors, which have yielded some interesting associations, but do not provide enough mechanistic information to support causal relationships. An interesting corollary is found in the Indian parable of the “six blind men of Hindustan”. Each of these six men were asked to feel a portion of an elephant. One felt the side and was convinced this was a wall. Another felt the tusk and was convinced this was a spear. Another felt the tail and stated this was a rope and so on. They argued about what they felt, and their conclusions differed. So, in addition to being blind, they were also intellectually short sighted in that they did not open their minds to the others’ perspectives and thus were not able to identify the whole animal, the elephant. Similarly, if we attempt to associate individual omics only to a particular disease or phenotype, we will fail to identify the holistic mechanistic pathway that leads to differences in the phenotype. Integration of multiple omics using ‘state of the art’ bioinformatic techniques can yield networks that provide mechanistic clues and answers, which relate to causality (Figure 3).

Description of Machine Learning

Machine learning (ML) has been described as the ability of machines to “think” and includes any computer program that improves its performance at some task through experience[11]. The different terminologies related to machine learning are tabulated in Table 2. Machine learning models can be categorized as supervised and unsupervised. **Supervised learning** methods use label information from samples or clinical outcome data for model training. **Unsupervised learning** is used when there are no labelled data or outcomes and identifies patterns (or clusters) in the data by principal co-ordinate analysis. Supervised and less commonly unsupervised learning have been most commonly applied to multi-omics data integration in perinatology.

Clinical research studies often evaluate many ML models and choose the one based on predictive accuracy. Most commonly evaluated ML models are logistic regression, k-nearest neighbors, random forest, gradient boosting machine, (kernel) support vector machine and artificial neural networks (ANNs or NNs)[12,13]. ANNs are ML models that conceptually resemble the organization of the brain and neurons. Neural models are able to represent complex, non-linear functions with reasonable computational costs. More recently, kernelized version of NNs have been developed that restrict the massive expressive power of NNs while still capturing nonlinear relationships, and also control the smoothness of the resulting predictive models[14].

An example of selecting the best possible model is by hyper-parameter optimization[13]. Hyper-parameters are additional, model-dependent parameters that are specified before the learning phase, e.g., the value of k in k-Nearest Neighbor or the number of hidden units in a Neural Network. Ensuring proper generalization involves finding sub-optimal values of the hyper-parameters. Each model is subjected to the hyper-parameter optimization procedure and the model with the highest area under the receiver operated characteristic curve (AUROC) is chosen. The model is validated with a separate data set other than the one it is trained on (internal or external dataset validation).

AUROC is the most commonly used metric of choice to compare models for predictive accuracy. A receiver operated characteristic curve (ROC) is a plot of the true positive rate (sensitivity) as a function of the false positive rate (1 - specificity) for different values at each threshold parameter. If a classifier (or test) is perfect, the ROC curve passes through the upper left corner (sensitivity and specificity of 1) and therefore the AUROC measures the overall accuracy of the classifier/test. An AUROC value of 0.5 indicates that the model is equivalent to a coin toss, while a value of 1.0 indicates perfect predictions. The **goodness of fit** of the considered models is assessed by calculating the Brier loss. The lower the Brier loss, better the goodness-of-fit, with values of 0.25 equivalent to a coin toss and 0 that of a perfect forecast[15].

Machine learning approaches like neural networks are a new frontier in clinical decision-making, having recently made dramatic advances in medical image analysis and other fields[16–18]. Recent evidence suggests that machine learning approaches may improve clinical outcomes in a variety diseases including congenital cataracts, metastatic breast cancer, post-prandial glucose prediction, and diabetic retinopathy[19–25].

Integration of Multiomics with Machine Learning Strategies

As with the six blind men parable, it would have helped had they collaborated and discussed their findings and then developed a conclusion. Multiomic integration approaches that evaluate changes over time are needed, especially in Perinatology, which presents critical windows for nutrition, growth and epigenetic modelling[6].

Multi-omics data integration approaches may be categorized as early, intermediate and late (Figure 4)[10,26]. **Early integration** or early concatenation although not complicated, may have problems with very few data points but numerous features, often known as the "**curse of dimensionality**" [27]. For example, multi-omics datasets may contain more than thousands of features when the genome, transcriptome, and proteome are combined but the number of patient samples may be relatively small (hundreds or less). Another problem that needs to be addressed is heterogeneity; omics datasets can have variable data distribution or data types (e.g., numerical, categorical, continuous, discrete) and differ significantly in the number of features. **Dimensionality reduction** is the process of reducing the number of variables or features in order to decrease the dimensionality and noise of a dataset. Early and intermediate integration processes often require prior dimensionality reduction to be more effective and functional.

The intermediate integration strategy transforms each omics dataset independently into a simpler representation, thus overcoming some issues with the early integration strategy. Transformation facilitates analysis by converting the data set to a less dimensional and less noisy one and also decreases heterogeneity between omics datasets such as the datas' type or size differences. The transformed and combined representation can then be analyzed by classical ML models, which include three transformation methods namely kernel-based methods, graph-based methods and deep learning[28]. Machine learning methods used for intermediate integration are either general-purpose methods that couple dimensionality

reduction with different downstream algorithms for a variety of tasks or applications or end-to-end models designed for one specific task [10].

Late integration combines the results from each omics layer or each omics dataset by machine learning tools (or manually) and the predictions performed terminally[10,28]. Since each omics dataset is analyzed by omic-specific machine learning tools, the problems of noise and heterogeneity found in other strategies are not present. However, the downside of the late integration strategy is that it cannot capture inter-omics interactions as the different machine learning models (for the different omics datasets) do not share complementary information between omics[28]. Combining predictions as in the late integration strategy may not fully bring out the details and complexity of multi-omics data analysis and interpretation to understand biological mechanisms of diseases.

Various software for multi-omics integrations already exist and innovative methods and procedures are being developed to integrate multi-omics data with the clinical phenotype. Using these techniques one can delineate and understand the strength of correlations between the individual's omics data including the genome, microbiome, metabolome and the epigenome. *Multiset sparse partial least squares path modeling (msPLS) for high dimensional omics data analysis* (https://doi.org/10.1186/s12859-019-3286-3)[29] is a novel statistical method that models and help us understand biological pathways and genetic architectures of complex phenotypes. The effect of multiple molecular markers, from multiple omics domains, on the variation of multiple phenotypic variables are simultaneously modelled. The sparsity in the model provides interpretable results from analyses of hundreds of thousands of biomolecular variables. msPLS has been shown to outperform Multi-Omics Factor Analysis (MOFA)[30] in terms of variation explained in a chronic lymphocytic leukemia dataset. Another method called MOTA (https://doi.org/10.3390/metabo10040144) is a network-based method that uses multi-omics data to rank candidate *disease biomarkers*[31]. Thus, researchers can use MOTA to investigate the biological significance of the highly ranked biomarkers[31].

Clinical Applications

Challenges of Pregnancy

Pregnancy associated pathology may lead to complications such as preterm birth, preeclampsia, and fetal growth restriction. Prematurity is the leading cause of neonatal morbidity and mortality. Many babies born preterm require prolonged neonatal hospital stays and have complications such as necrotizing enterocolitis, bronchopulmonary dysplasia, neurodevelopmental disorders, and growth delays. The application of omics data from high throughput analysis such as with epigenomics data that show modifications of DNA or histone configurations may help us understand cellular processes of different cell types in the developing fetus and placenta and whether they are functioning properly for a successful pregnancy. Longitudinal multiomic data analysis during the various stages of pregnancy is likely to lead to early diagnosis of risk and novel methods for therapy and prevention. Some examples for the use of multi-omics and machine learning in pregnancy are discussed below.

Multiomics modeling and delineating metabolome adaptations during human pregnancy may provide the framework for future studies[32]. Ghaemi et al performed a multiomics analysis of the immunome, transcriptome, microbiome, proteome and metabolome from 17 pregnant women (51 samples), delivering at term[32]. The investigators used multivariate predictive modeling using the Elastic Net algorithm to measure the ability of each dataset to predict gestational age. Multi-omics datasets were combined using stacked generalization, which increased the predictive power and revealed novel interactions among multi-omic datasets.

A healthy pregnancy is a cumulation of complex interrelated biological adaptations involving placentation, maternal immune responses, and hormonal homeostasis. Understanding these biological adaptations by multi-omics integration using state-of-the-art machine-learning methods is imperative to predict short- and long-term health trajectories for a mother and the offspring. These principles are delineated by data-driven modeling of pregnancy-related complications[33]. Maric et al investigated **modeling of preeclampsia** using longitudinal multiomics data and developed a multi-omics model using machine learning [34]. The multiomics model had high accuracy (AUROC of 0.94 [95% confidence intervals (CI) 0.90 to 0.99]). Ten urine metabolites were found to be major players in this model, which was validated using an independent cohort of 16 women (AUROC of 0.87, [95% CI 0.76 to 0.99]). The prediction accuracy of the urine metabolome model could be improved with the integration of clinical variables (AUC = 0.90, [95% CI 0.80 to 0.99]). This multi-omics integration study identified several biological pathways associated with preeclampsia.

Integration of multi-omics trajectories of the maternal metabolome, proteome, and immunome may be useful to **predict labor onset**[35]. Progression of pregnancy towards and birth is associated with major transitions occur in feto-maternal immune, metabolic, and endocrine systems. In a longitudinal study conducted in 63 women who went into spontaneous labor, serial blood samples were collected during the last 100 days of pregnancy. More than 7000 plasma analytes and peripheral immune cell responses were analyzed using untargeted mass spectrometry, aptamer-based proteomic technology, and single-cell mass cytometry. An integrated multiomic model predicted the time to spontaneous labor [($R = 0.85$, 95% CI [0.79 to 0.89], $P = 1.2 \times 10^{-40}$, $N = 53$, training set); ($R = 0.81$, 95% CI [0.61 to 0.91], $P = 3.9 \times 10^{-7}$, $N = 10$, independent test set)]. Coordinated alterations in maternal metabolome, proteome, and immunome marked a molecular shift from pregnancy maintenance to pre-labor biology, 2 to 4 weeks before delivery. Regulation of inflammatory responses preceded labor that was associated with a surge in steroid hormone metabolites and interleukin-1 receptor type 4.

Multi-omics integration and modelling of transcriptomics and proteomics profiling of plasma and urine metabolomics may identify **preterm birth** (PTB, delivery before 37 weeks gestation)[36]. Plasma and urine samples were analyzed from 81 pregnant women in 5 biorepository cohorts in low- and middle-income countries (LMICs). Of the 81 pregnant women, 39 had PTBs (48.1%) and 42 had term pregnancies (51.9%) (mean [SD] age of 24.8 [5.3] years). A cohort-adjusted machine learning algorithm was applied to each biological data set, and results combined into a final integrative model using machine

learning. The integrated model was more accurate, with an AUROC of 0.83 [95% CI, 0.72–0.91] compared with the models derived for each independent omic data set (transcriptomics AUROC, 0.73 [95% CI, 0.61–0.83]; metabolomics AUROC, 0.59 [95% CI, 0.47–0.72]; and proteomics AUROC, 0.75 [95% CI, 0.64–0.85]). Primary features associated with PTB included a serum inflammatory module and a urine metabolomic module associated with glutamine and glutamate metabolism and valine, leucine, and isoleucine biosynthesis pathways. This is an excellent example of integration of multi-omic data sets with machine learning tools that will lead to predictive tests and intervention candidates for preventing PTB.

The Neonatal Intensive Care Unit (NICU)

There are numerous areas in neonatal intensive care where AI combined with multiomics show promise and will likely become an important adjunct for disease prediction, diagnostics, and therapeutics. Longitudinal continuous monitoring of heart rate tracings, pulse oximetry and ventilation patterns generate data that can be interrogated by ML. Several poorly defined “diseases” in neonatal intensive care such as sepsis, necrotizing enterocolitis (NEC), bronchopulmonary dysplasia (BPD), and various forms of encephalopathy for which use of AI and multiomic based methods have the potential to markedly alter our current definitions and improve management.

Retinopathy of Prematurity

An excellent example of the use of AI is retinopathy of prematurity (ROP), which can now be detected much earlier and more accurately using these methods[37–39]. As an illustration (Figure 5), features seen in the retina during development in preterm infants can be utilized using user defined definitions or machine learned features to provide accurate diagnosis of this disease[1].

Necrotizing Enterocolitis

Defining NEC remains an enigma and 8 different definitions have been proposed[40], but none are considered satisfactory. Definitions attempt to define distinct entities and what we term “NEC” today is not a distinct entity[41], thus is not readily amenable to a definition. The need to revise the outdated modified Bell staging criteria is crucial in improving NEC management[41,42]. Recent data suggests that genetic susceptibility and stool microbiota signatures may help identify preterm infants at increased risk of the disease.[43,44] Ongoing studies using single or multi-omic approaches may help to characterize biomarkers that could predict different intestinal injuries currently being termed as “NEC”. Intestinal ultrasound may improve diagnostic accuracy for NEC but has been slow in adoption. Patient family perspectives are key in accelerating our efforts to integrate newer diagnostic methods into practice[42]. At the very least, we are beginning to recognize that the Bell staging criteria are not optimal and better criteria are needed[40].

Several attempts have been made to differentiate what we term as classical NEC from some of the “imposters.” For example, NEC can be confused with spontaneous intestinal perforation (SIP) even with the availability of clinical and laboratory information[45,46]. Supervised machine learning was used to evaluate NEC versus spontaneous intestinal

perforation (SIP), which can be differentiated at laparotomy, but the increasing use of peritoneal drains without direct visualization of the bowel limits the accuracy of the diagnosis. In a recent study that evaluated several clinical and radiographic features in infants who had undergone surgery where necrosis was clearly found versus isolated ileal perforation, machine learning was able to readily delineate several features that are helpful in differentiating these two outcomes prior to surgery[47].

Other studies have evaluated the stool microbiome prior to NEC using machine learning techniques that suggest the possibility of their use as biomarkers[48]. Another study employed clinical and imaging preterm infant data from six neonatal centers using AI approaches to determine their relationship to the development of NEC and found that such an approach may be feasible[49]. Now, it is reasonable to infer that a completely new approach to intestinal injuries in the neonatal intestine is in order. Use of multiomic approaches along with AI[47] can be used to differentiate the different forms of intestinal injury or dysfunction that are labeled as NEC[50,51]. AI strategies may make it possible to redefine the different forms of intestinal injury commonly seen in preterm infants into different clusters that derive from different pathophysiology. By doing this, each one of these clusters can be better evaluated for discrete mechanisms, biomarkers and preventative strategies.

Precision Nutrition as an example of integration of omics and AI

One of the greatest challenges in Neonatology remains how to optimize the timing and composition of nutrition in preterm infants to achieve appropriate growth, while minimizing devastating outcomes such as intestinal injury, LOS, BPD, and ROP. Progress has been made since the era of delayed feeding of high risk preterm infants. Many guideline based approaches have been developed that have good evidence based foundations (as described in the second era of medicine in the Introductory section of this review). Such approaches may be adequate for many of these infants, but a significant percentage require a more personalized (precision) approach based on the individual's genetic makeup and omics data rather than traditional dogma, anecdotal clinical experience, or even good clinical studies that evaluate populations rather than individuals. In accordance with concerns raised by the National Institute of Health (NIH) Precision Nutrition Initiative, the "one size fits all" approach marginalizes many of these infants and may even lead to harm. For example, not until recently has sex been evaluated as an important variable that relates to nutritional needs. Additionally, preterm infants may exhibit weight gains commensurate with standardized growth curves, but still develop adverse outcomes such as NEC, ROP, BPD and LOS. Some of these outcomes may depend on previous in utero developmental events that are not considered when simply using a postnatal growth curve approach.

The development of precision-based nutritional strategies that leverage newly developed integration of "multiomics," systems biology, and machine learning will significantly improve and optimize growth and reduce adverse outcomes in preterm infants. Using multi-omics (microbiome, metabolome, and inflammasome) based systems biology network analysis along with AI will then identify biomarkers to guide personalized interventional strategies for high-risk infants in order to mitigate adverse short- and long-term outcomes experienced by these infants. ML Integration of blood parameters, dietary habits,

anthropometrics, physical activity, and gut microbiota has shown to predict personalized postprandial glycemic response to real-life meals in adults[21,52]. Similar studies promoting precision nutrition in neonates have not been performed.

Precision nutrition explores and incorporates the effects of the complex interplay among genetics, microbiome, antibiotic and probiotic use, metabolism, food environment, and physical activity, as well as economic, social, and other behavioral characteristics[53]. The NIH is leading efforts to advance the field of precision nutrition with a plan 2020–2030 with strategic goals to address precision nutrition research[54].

Societal and Ethical Issues related to the use of AI in healthcare

The intent of AI/ML use in healthcare is to improve diagnosis or predict clinical phenotypes that require intervention, which will then translate to better patient outcomes. However, we should also be mindful of the potential adverse consequences of the use of AI and ML in healthcare as they are increasingly explored for patient care. Societal issues due to relate to **fairness, explainability, privacy, ethics** and **legislation,** which should be sufficiently addressed in new projects[55]. **Fairness** is a central perspective in healthcare and the models developed by AI/ML should be fair to human characteristics including gender, race and ethnicity. For example, a predictive model that is developed with one gender, race or ethnic group only is not likely be applicable to the population at large and is bound to fail in the real world. **Privacy** is another critical issue in healthcare and the anonymity and privacy of the patients should be respected, if the ML modelling is performed using private information of the patients e.g., identifiers such as medical record number, social security numbers or even the zip code where they live. **Explainability and interpretability** are important aspects of clinical medicine e.g., how a certain exposure leads to disease or a drug leads to side effects. However, ML algorithms, especially deep learning models are considered black boxes where the connection between input and output is obscure, and the function is not explainable. This might be all the more important when the model does not work very well, and the issues are not explainable. **Ethical research** is the cornerstone of medicine and healthcare, and several questions need to be addressed. Will it be possible for AI and ML models to follow or address the basic biomedical ethical principles of respect for autonomy, non-maleficence, beneficence and justice? Could the use of AI, if the model is poorly developed, cause more harm than good? Could the use of AI in some way by their predictions increase psychological stress in parents, make the patients poor insurance clients if we predict long term poor outcomes early? Are there ethical concerns, conflicts of interest, individual on the part of the researcher or clinician, hospital or institution – financial or otherwise?[56]. How are medicolegal risks that arise from use of AI managed? [57] We as proponents for the use of AI/ML should make sure that we address the societal and ethical issues adequately to maximize risk-benefit ratios.

Summary and the future

There are tremendous opportunities for enhancing patient outcomes that technological advances have brought forth, integrating multiomics and clinical datasets with modern computing platforms using machine learning. The advances and capabilities have the

potential to bring about paradigm shifts in how we practice medicine in the fields of Perinatal and Neonatal medicine. Predictive analytics hold promise for the prevention of disease and improved diagnostics and therapeutics. Potential ethical and societal issues due to the use of machine learning techniques in healthcare need to be addressed with foresight and wise judgement.

Statement of financial support.

This study was funded by the following extramural sources:

NIH R03HD098482, R21HD091718, CA125123; NCI P30CA125123; NIH/NCI R01CA220297, R01CA216426; NIEHS P30 ES030285, P42 ES027725, NIMHD P50MD015496; CPRIT RP170005, R35GM138353, Burroughs Wellcome Fund (1019816), and the March of Dimes.

Data Availability Statement:

This is a review and all data discussed are included in the review.

LITERATURE CITED

1. Scruggs BA, Chan RVP, Kalpathy-Cramer J, Chiang MF & Campbell JP Artificial Intelligence in Retinopathy of Prematurity Diagnosis. Transl Vis Sci Technol 9, 5 (2020).
2. Titano JJ et al. Automated Deep-Neural-Network Surveillance of Cranial Images for Acute Neurologic Events. Nat Med 24, 1337–1341 (2018). [PubMed: 30104767]
3. Terranova N, Venkatakrishnan K & Benincosa LJ Application of Machine Learning in Translational Medicine: Current Status and Future Opportunities. Aaps j 23, 74 (2021). [PubMed: 34008139]
4. Ashley EA The Genome Odyssey (Celadon Books, 2021).
5. López de Maturana E et al. Challenges in the Integration of Omics and Non-Omics Data. Genes (Basel) 10 (2019).
6. Lugo-Martinez J et al. Integrating Longitudinal Clinical and Microbiome Data to Predict Growth Faltering in Preterm Infants. J Biomed Inform, 104031 (2022). [PubMed: 35183765]
7. Clapp MA & McCoy TH The Potential of Big Data for Obstetrics Discovery. Curr Opin Endocrinol Diabetes Obes 28, 553–557 (2021). [PubMed: 34709211]
8. Ramakrishnan R, Rao S & He JR Perinatal Health Predictors Using Artificial Intelligence: A Review. Womens Health (Lond) 17, 17455065211046132 (2021). [PubMed: 34519596]
9. Garcia-Canadilla P, Sanchez-Martinez S, Crispi F & Bijnens B Machine Learning in Fetal Cardiology: What to Expect. Fetal Diagn Ther 47, 363–372 (2020). [PubMed: 31910421]
10. Cai Z, Poulos RC, Liu J & Zhong Q Machine Learning for Multi-Omics Data Integration in Cancer. iScience 25, 103798 (2022). [PubMed: 35169688]
11. Mitchell T Machine Learning 1st edition edn(McGraw-Hill, 1997).
12. Podda M et al. A Machine Learning Approach to Estimating Preterm Infants Survival: Development of the Preterm Infants Survival Assessment (Pisa) Predictor. Sci Rep 8, 13743 (2018). [PubMed: 30213963]
13. Hastie T, Tibshirani R & Friedman JH The Elements of Statistical Learning: Data Mining, Inference, and Prediction 2nd edn(Springer, 2009).
14. Sahs J et al. Shallow Univariate Relu Networks as Splines: Initialization, Loss Surface, Hessian, & Gradient Flow Dynamics. https://arxivorg/abs/200801772 (2020).
15. Brier GW Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review 78, 1–3 (1950).
16. Anwar SM et al. Medical Image Analysis Using Convolutional Neural Networks: A Review. Journal of Medical Systems 42, 226 (2018). [PubMed: 30298337]

17. Lundervold AS & Lundervold A An Overview of Deep Learning in Medical Imaging Focusing on Mri. Zeitschrift für Medizinische Physik 29, 102–127 (2019). [PubMed: 30553609]
18. Munir K, Elahi H, Ayub A, Frezza F & Rizzi A Cancer Diagnosis Using Deep Learning: A Bibliographic Review. Cancers (Basel) 11 (2019).
19. Long E et al. An Artificial Intelligence Platform for the Multihospital Collaborative Management of Congenital Cataracts. Nature Biomedical Engineering 1, 0024 (2017).
20. Sayres R et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. Ophthalmology 126, 552–564 (2019). [PubMed: 30553900]
21. Mendes-Soares H et al. Model of Personalized Postprandial Glycemic Response to Food Developed for an Israeli Cohort Predicts Responses in Midwestern American Individuals. Am J Clin Nutr 110, 63–75 (2019). [PubMed: 31095300]
22. Wang D, K. A, Gargeya R, Irshad H, Beck AH. Deep Learning for Identifying Metastatic Breast Cancer. Http://Arxiv.Org/Abs/1606.05718, 2016).
23. Wang DD & Hu FB Precision Nutrition for Prevention and Management of Type 2 Diabetes. Lancet Diabetes Endocrinol 6, 416–426 (2018). [PubMed: 29433995]
24. Wang P et al. Real-Time Automatic Detection System Increases Colonoscopic Polyp and Adenoma Detection Rates: A Prospective Randomised Controlled Study. Gut 68, 1813–1819 (2019). [PubMed: 30814121]
25. Wu L et al. Randomised Controlled Trial of Wisense, a Real-Time Quality Improving System for Monitoring Blind Spots During Esophagogastroduodenoscopy. Gut 68, 2161–2169 (2019). [PubMed: 30858305]
26. Zitnik M et al. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. Inf Fusion 50, 71–91 (2019). [PubMed: 30467459]
27. Bellman R Dynamic Programming. Science 153, 34–37 (1966). [PubMed: 17730601]
28. Picard M, Scott-Boyer MP, Bodein A, Périn O & Droit A Integration Strategies of Multi-Omics Data for Machine Learning Analysis. Comput Struct Biotechnol J 19, 3735–3746 (2021). [PubMed: 34285775]
29. Csala A, Zwinderman AH & Hof MH Multiset Sparse Partial Least Squares Path Modeling for High Dimensional Omics Data Analysis. BMC Bioinformatics 21, 9 (2020). [PubMed: 31918677]
30. Argelaguet R et al. Multi-Omics Factor Analysis-a Framework for Unsupervised Integration of Multi-Omics Data Sets. Mol Syst Biol 14, e8124 (2018). [PubMed: 29925568]
31. Fan Z, Zhou Y & Ressom HW Mota: Network-Based Multi-Omic Data Integration for Biomarker Discovery. Metabolites 10 (2020).
32. Ghaemi MS et al. Multiomics Modeling of the Immunome, Transcriptome, Microbiome, Proteome and Metabolome Adaptations During Human Pregnancy. Bioinformatics 35, 95–103 (2019). [PubMed: 30561547]
33. Espinosa C et al. Data-Driven Modeling of Pregnancy-Related Complications. Trends Mol Med 27, 762–776 (2021). [PubMed: 33573911]
34. Marj I et al. Multiomics Longitudinal Modeling of Preeclamptic Pregnancies, 2021.
35. Stelzer IA et al. Integrated Trajectories of the Maternal Metabolome, Proteome, and Immunome Predict Labor Onset. Sci Transl Med 13 (2021).
36. Jehan F et al. Multiomics Characterization of Preterm Birth in Low- and Middle-Income Countries. JAMA Netw Open 3, e2029655 (2020). [PubMed: 33337494]
37. Coyner AS et al. Single-Examination Risk Prediction of Severe Retinopathy of Prematurity. Pediatrics 148 (2021).
38. Tsai AS et al. Assessment and Management of Retinopathy of Prematurity in the Era of Anti-Vascular Endothelial Growth Factor (Vegf). Prog Retin Eye Res, 101018 (2021). [PubMed: 34763060]
39. Li J et al. Evaluation of Artificial Intelligence-Based Quantitative Analysis to Identify Clinically Significant Severe Retinopathy of Prematurity. Retina 42, 195–203 (2022). [PubMed: 34387234]
40. Lueschow SR, Boly TJ, Jasper E, Patel RM & McElroy SJ A Critical Evaluation of Current Definitions of Necrotizing Enterocolitis. Pediatr Res (2021).

41. Neu J, Modi N & Caplan M Necrotizing Enterocolitis Comes in Different Forms: Historical Perspectives and Defining the Disease. Seminars in fetal & neonatal medicine 23, 370–373 (2018). [PubMed: 30100524]
42. Kim JH, Sampath V & Canvasser J Challenges in Diagnosing Necrotizing Enterocolitis. Pediatr Res 88, 16–20 (2020). [PubMed: 32855507]
43. Pammi M et al. Intestinal Dysbiosis in Preterm Infants Preceding Necrotizing Enterocolitis: A Systematic Review and Meta-Analysis. Microbiome 5, 31 (2017). [PubMed: 28274256]
44. Talavera MM et al. Single Nucleotide Polymorphisms in the Dual Specificity Phosphatase Genes and Risk of Necrotizing Enterocolitis in Premature Infant. Journal of neonatal-perinatal medicine 13, 373–380 (2020). [PubMed: 31985475]
45. Berrington J & Embleton ND Discriminating Necrotising Enterocolitis and Focal Intestinal Perforation. Archives of disease in childhood Fetal and neonatal edition (2021).
46. Berrington JE & Embleton ND Time of Onset of Necrotizing Enterocolitis and Focal Perforation in Preterm Infants: Impact on Clinical, Surgical, and Histological Features. Frontiers in pediatrics 9, 724280 (2021). [PubMed: 34540772]
47. Lure AC et al. Using Machine Learning Analysis to Assist in Differentiating between Necrotizing Enterocolitis and Spontaneous Intestinal Perforation: A Novel Predictive Analytic Tool. J Pediatr Surg (2020).
48. Hooven T LY, Salleb-Aouissi A & Multiple Instance Learning for Predicting Necrotizing Enterocolitis in Premature Infants Using Microbiome Data. CHIL '20: Proceedings of the ACM Conference on Health, Inference, and LearningApril 2020 Pages 99–109https://doiorg/101145/33685553384466 (2020).
49. Ji J et al. A Data-Driven Algorithm Integrating Clinical and Laboratory Features for the Diagnosis and Prognosis of Necrotizing Enterocolitis. PLoS One 9, e89860 (2014). [PubMed: 24587080]
50. Neu J Necrotizing Enterocolitis: A Multi-Omic Approach and the Role of the Microbiome. Dig Dis Sci 65, 789–796 (2020). [PubMed: 32008132]
51. Beck LC, Granger CL, Masi AC & Stewart CJ Use of Omic Technologies in Early Life Gastrointestinal Health and Disease: From Bench to Bedside. Expert Rev Proteomics 18, 247–259 (2021). [PubMed: 33896313]
52. Zeevi D et al. Personalized Nutrition by Prediction of Glycemic Responses. Cell 163, 1079–1094 (2015). [PubMed: 26590418]
53. Rodgers GP & Collins FS Precision Nutrition-the Answer to "What to Eat to Stay Healthy". Jama 324, 735–736 (2020). [PubMed: 32766768]
54. NIH. Us Department of Health & Human Services. 2020–2030 Strategic Plan for Nih Nutrition Research. Https://Www.Niddk.Nih.Gov/About-Niddk/Strategic-Plans-Reports/Strategic-Plan-Nih-Nutrition-Research, 2020).
55. Vellido A Societal Issues Concerning the Application of Artificial Intelligence in Medicine. Kidney Dis (Basel) 5, 11–17 (2019). [PubMed: 30815459]
56. Morley J et al. The Ethics of Ai in Health Care: A Mapping Review. Soc Sci Med 260, 113172 (2020). [PubMed: 32702587]
57. Oliva A et al. Management of Medico-Legal Risks in Digital Health Era: A Scoping Review. Front Med (Lausanne) 8, 821756 (2021). [PubMed: 35087854]
58. Hasin Y, Seldin M & Lusis A Multi-Omics Approaches to Disease. Genome Biol 18, 83 (2017). [PubMed: 28476144]

Impact

- Biotechnological advances have made multi-omic evaluations feasible and integration of multi-omics data may provide a holistic view of disease pathophysiology.
- Artificial Intelligence and machine learning tools are being increasingly used in healthcare for diagnosis, prognostication and outcome predictions.
- Leveraging artificial intelligence and machine learning tools for integration of multi-omics and clinical data will pave the way for precision medicine in perinatology

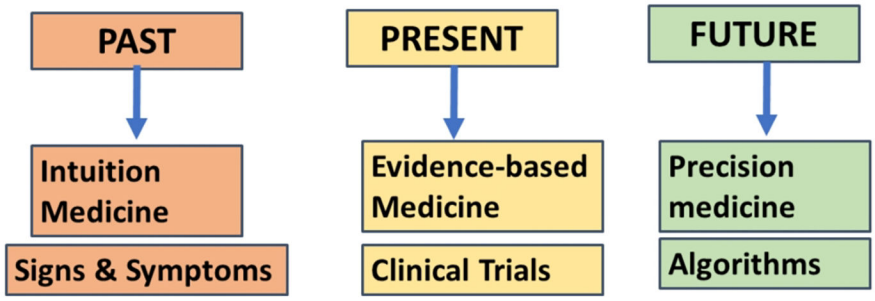

Figure 1: Three Eras of Medicine

Evolution of medicine in the past, present and future. In the past, medicine was practiced by evaluation of signs and symptoms and based solely on the knowledge of the individual physician and was intuition medicine. Currently, medicine is based on scientific research including clinical trials and is evidence-based medicine. In the future, medicine will be practiced by algorithms based on the patient's phenotype, genome, epigenome or other omics data that will individualize treatment and that which constitutes precision medicine.

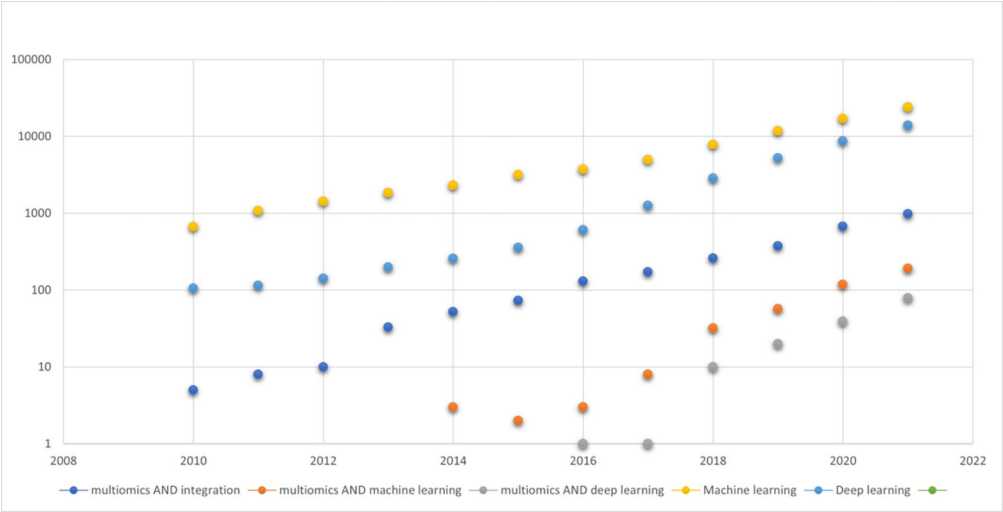

Figure 2: Number of publications per year on search keywords from 2010 to 2021.
Search words used were: ‘multiomics,’ ‘multiomics AND integration,’ ‘multiomics AND machine learning’ and ‘multiomics AND deep learning’.

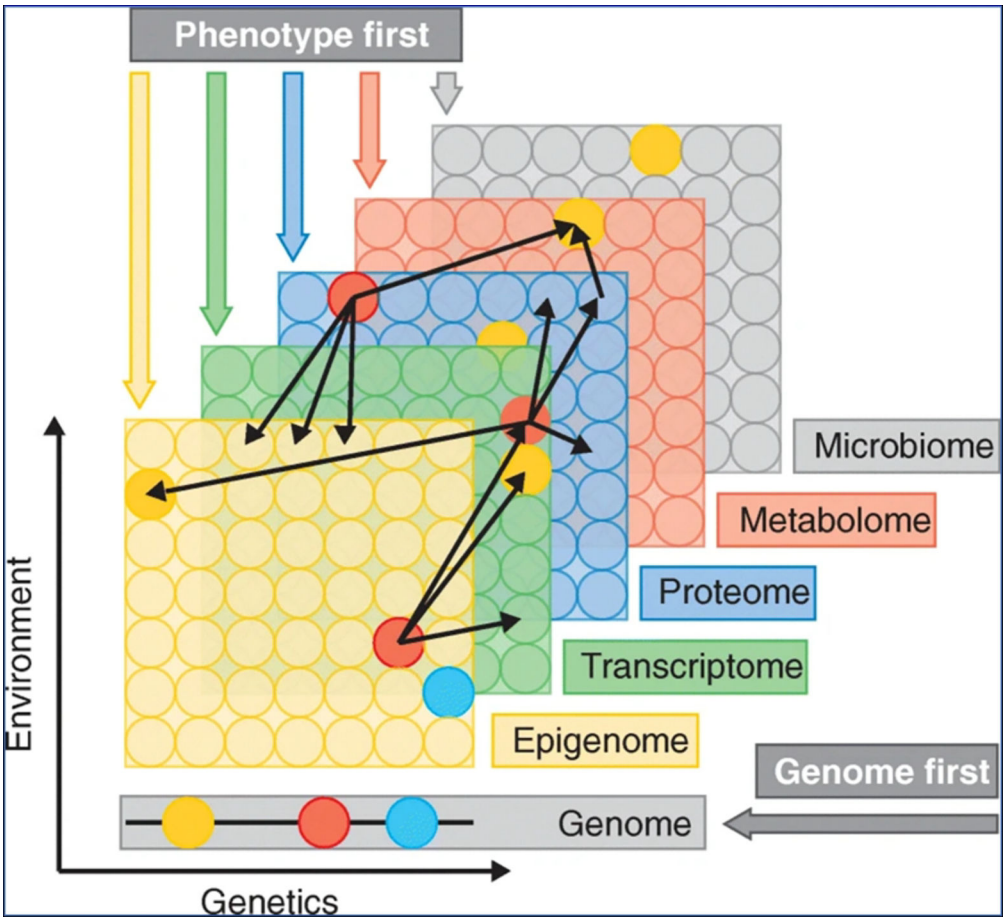

Figure 3: Schema for integration of omics data and approaches to disease (from Hasan et al reproduced, open access data)[58]

Multiple omics data types are depicted in the different layers. Circles in each colored panel represent the entire pool of molecules from which the omics data are collected. Genetic regulation and environment impact the molecules in each layer except the genome layer. The **thin black arrows** represent potential interactions or correlations detected between molecules in different layers, for example, the red transcript can be correlated to multiple proteins. **Thick colored arrows** near the top of the panel point to different potential starting points for consolidating multiple omics data to understand biological systems and pathogenesis of disease. The genome first approach (thick gray arrow) implies that one starts from associated genetic locus, while the phenotype first approach may start from any other omics layer. The environment first approach (not shown) examines environmental perturbations.

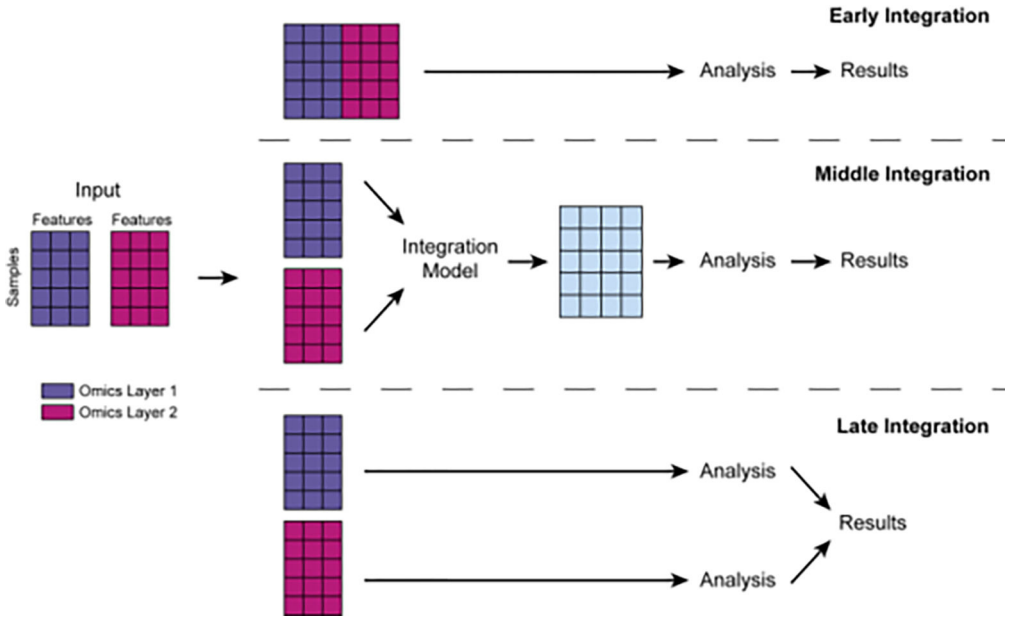

Figure 4: Illustration of the timing of integration for multi-omic data matrices (figure reproduced from Cai et al[10])

Muti-omics data integration may be categorized by the timing of integration related to analysis and interpretation as early, intermediate and late. The features from different data matrices are concatenated early in early integration but challenged by high dimensionality but few number of samples. Intermediate integration strategy involves each omic dataset being transformed into a simpler representation by dimensionality reduction and the combined data set is consolidated by machine learning without concatenating features or merging results. Late integration involves the results of each multi-omic data set combined terminally after each omics layer is analyzed independently.

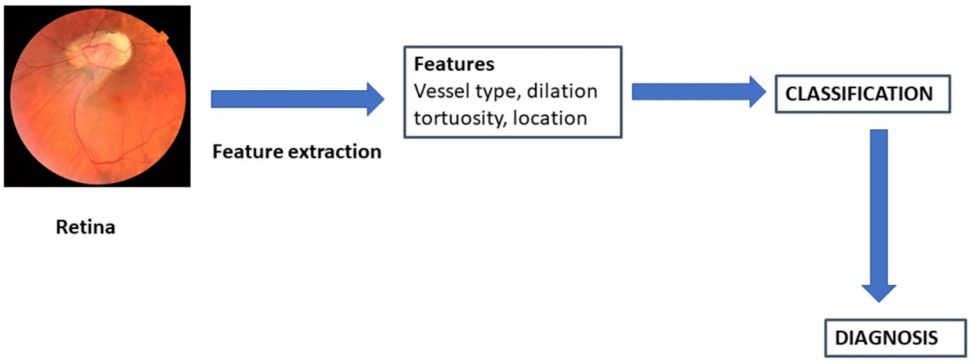

Figure 5: Diagnosis of Retinopathy of Prematurity by Machine learning

Early efforts to automate diagnosis of ROP used features such as vessel type, dilation, tortuosity and location and a score was developed without computer algorithm (e.g., ROPtool). Machine learning uses a classifier step that relates the features to the diagnosis. Deep convoluted neural networks (CNN) differ from traditional ML algorithms such as Support Vector Machine by letting the machine learn the features related to the input image that correlates to the diagnosis without human defined features.

Table 1:

Multiomics Technologies

“Omic” technology	**Description**
Genome	The basic template of DNA. Technologies can identify genetic (DNA) variants associated with diseases.
Microbiome	Allows for accurate quantitative determination of microbial taxa, their abundance and diversity that can be associated with healthy and diseased states.
Transcriptome	Examines RNA levels transcribed from DNA template. A small amount of RNA is transcribed for protein synthesis, a much larger amount is encoded for other purposes, which may be implicated in disease.
Proteome	Quantifies peptides which may be used as disease biomarkers.
Metabolome	Detects and quantifies small molecules which include carbohydrates, amino and fatty acids, and other products of cellular metabolism. Abnormally high or low levels may predict disease.
Epigenome	Characterizes modifications of DNA or DNA associated proteins.

Table 2:

Common computational terminologies

Machine Learning	An application of artificial intelligence that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning focuses on the development of computer programs** that can access data and use it to learn for themselves.
Deep Learning	Deep learning is a subset of machine learning (ML), where artificial neural networks, algorithms modeled to work like the human brain, learn from large amounts of data.
Neural Networks	Computer systems modeled on the human brain and nervous system. These are collections of interconnected nodes organized in multiple layers and are connected via weighted links. Learning is performed by adjusting the weights to perform a task at hand with maximum accuracy.
Supervised Learning	Algorithms that learn to predict a certain property or outcome associated with a given set of input features and outcome labels.
Unsupervised learning	Derives hidden structure from the data with no knowledge of an outcome or a label.
Semi-supervised learning	Combines a small amount of labeled data with a large amount of unlabeled data during training. Unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy.
High dimensional data	Data with many features that typically exceed the number of observations.
Dimensionality reduction	Transformation of data from a highdimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension.