


REVIEW ARTICLE OPEN


# Genomics and multiomics in the age of precision medicine

 Srinivasan Mani <sup>1</sup>✉, Seema R. Lalani<sup>2</sup> and Mohan Pammi<sup>3</sup>

© The Author(s) 2025

Precision medicine is a transformative healthcare model that utilizes an understanding of a person's genome, environment, lifestyle, and interplay to deliver customized healthcare. Precision medicine has the potential to improve the health and productivity of the population, enhance patient trust and satisfaction in healthcare, and accrue health cost-benefits both at an individual and population level. Through faster and cost-effective genomics data, next-generation sequencing has provided us the impetus to understand the nuances of complex interactions between genes, diet, and lifestyle that are heterogeneous across the population. The emergence of multiomics technologies, including transcriptomics, proteomics, epigenomics, metabolomics, and microbiomics, has enhanced the knowledge necessary for maximizing the applicability of genomics data for better health outcomes. Integrative multiomics, the combination of multiple 'omics' data layered over each other, including the interconnections and interactions between them, helps us understand human health and disease better than any of them separately. Integration of these multiomics data is possible today with the phenomenal advancements in bioinformatics, data sciences, and artificial intelligence. Our review presents a broad perspective on the utility and feasibility of a genomics-first approach layered with other omics data, offering a practical model for adopting an integrated multiomics approach in pediatric health care and research.

*Pediatric Research* (2025) 97:1399–1410; <https://doi.org/10.1038/s41390-025-04021-0>

## IMPACT:

- Precision medicine provides a paradigm shift from a conventional, reactive disease control approach to proactive disease prevention and health preservation.
- Phenomenal advancements in bioinformatics, data sciences, and artificial intelligence have made integrative multiomics feasible and help us understand human health and disease better than any of them separately.
- The genotype-first approach or reverse phenotyping has the potential to overcome the limitations of the phenotype-first approach by identifying new genotype-phenotype associations, enhancing the subclassification of diseases by widening the phenotypic spectrum of genetic variants, and understanding functional mechanisms of genetic variations.

## INTRODUCTION

Precision medicine, also known as personalized or individualized medicine, is a novel healthcare approach that utilizes the understanding of a person's genome, environment, lifestyle, and interplay to deliver customized healthcare choices for prevention, diagnosis, and treatment.<sup>1</sup> Precision medicine provides a paradigm shift from a conventional, reactive disease control approach to proactive disease prevention and health preservation.<sup>2</sup> We can understand an individual's short—and long-term disease risks at a molecular level, make predictions regarding health trajectories, and implement preventive measures tailored to the individual, considering their environmental influences and genomic profile.<sup>3</sup> We can develop biomarkers for early detection of diseases, monitor their progression, and develop novel targeted therapies that could interrupt disease progression and restore health. We may be able to select therapeutics best suited to an individual's molecular profile, maximizing efficacy and limiting adverse effects. The healthcare transformation based on the precision medicine approach has

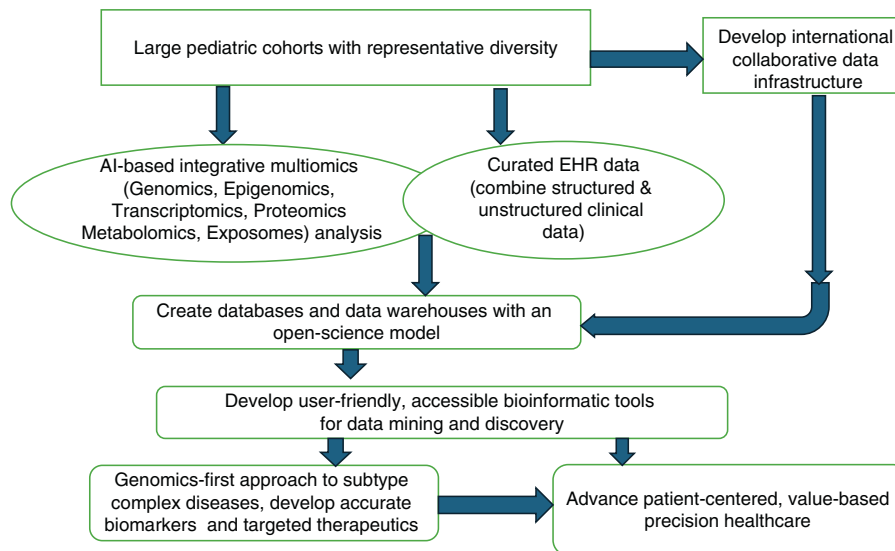
the potential to provide rich dividends for the upfront costs involved in establishing the infrastructure.<sup>4</sup> Precision medicine may improve health and productivity of the population, enhance patient trust and satisfaction in healthcare, and accrue health cost-benefits both at an individual and population level.<sup>5–7</sup>

The genomics revolution has laid the foundation for realizing the promise of precision medicine and P4 (predictive, preventive, personalized, and participatory) healthcare.<sup>8</sup> The Human Genome Project in 2003 helped scientists understand the framework of human biology better and gave them a deeper insight into the etiology of common non-communicable diseases. Through faster and cost-effective genomics data, next-generation sequencing provided the impetus to understand the nuances of complex interactions between genes, diet, and lifestyle that are heterogeneous across the population. The other omics technologies, including transcriptomics, proteomics, epigenomics, metabolomics, and microbiomics, have emerged, enhancing the knowledge necessary for maximizing the applicability of genomics data for better health outcomes. Integration of these multiomics data is

<sup>1</sup>Department of Pediatrics, University at Buffalo, Buffalo, NY, USA. <sup>2</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. <sup>3</sup>Division of Neonatology, Department of Pediatrics, Texas Children's Hospital, Houston, TX, USA. ✉email: [smani3@buffalo.edu](mailto:smani3@buffalo.edu)

Received: 24 April 2024 Revised: 6 March 2025 Accepted: 10 March 2025

Published online: 4 April 2025



**Fig. 1** A road map to AI-based integrative multiomics approach for precision medicine in pediatric healthcare and research.

possible today with the phenomenal advancements in bioinformatics, data sciences, and artificial intelligence (AI).<sup>9</sup>

Integrative genomics (with other omics) can help us understand the heterogeneous etiopathogenesis of complex pediatric diseases and create a framework for a precision medicine approach. Breaking down overlapping disease spectrums into definitive subtypes based on an integrative multiomics approach incorporated with clinical data from the EMR can be very valuable and may lead to targeted therapy. Integrated multiomics data of large cohorts with representative population diversity could provide valuable insights into the epidemiology of a disease.<sup>10–12</sup> Our review is organized into nine broad sections following the roadmap for a practical model for adopting an AI-based integrative multiomics approach that we propose in this review to deliver precision medicine in pediatric health care and research. (Fig. 1) First, we discuss the importance and current state of large-scale longitudinal cohorts, followed by a discussion on innovations and advances in genomics and other omics, and the role of bioinformatics and AI in integrating multiomics. Then, we describe the scope of integrating electronic health records with multiomics data, the current state of genomic databases, data mining technologies, and bioinformatics. Finally, we highlight the clinical and research applications of integrative multiomics, discuss the utility and feasibility of a genomics-first approach with a practical model for adoption, and discuss the challenges and future of precision medicine.

### LONGITUDINAL COHORTS

Large prospective cohorts are the backbone of clinical epidemiology. They help us understand the genetic determinants of health and disease, environmental exposures and risk factors, the natural history of a disease, modifiers of disease progression, response to treatment, and long-term prognosis at a population level. Understanding the importance of such longitudinal data in the context of genomics and multiomics, several national public-funded research agencies have invested in developing huge longitudinal cohorts. Table 1 depicts the various longitudinal cohorts available for researchers to access genomic data. Unfortunately, many of the available cohorts have not included children. Developing pediatric cohorts and expanding the recruitment of the pediatric population into the existing cohorts will be crucial in understanding the genetic epidemiology of diseases in children, including infants and newborns.

Diversity in genomic and multiomics datasets is an essential factor necessary to achieve equity in genomic healthcare and to bring the benefits of precision medicine to the entire population.<sup>13</sup> It is estimated that participants of European descent constitute 86.3% of all the genomic studies ever conducted worldwide.<sup>14</sup> The participants of African, South Asian, and Hispanic descent together constitute less than 10% of the studies. In addition to the limited representation in the existing datasets, the applicability of the human reference genome to diverse populations is questionable because of the origin of nearly three-fourths of the reference sequences from a single donor in the US.<sup>15</sup> Underrepresented populations in genomic research belong to the minority and indigenous groups in the research-intensive nations of the developed world and populations from low- and middle-income countries.<sup>14</sup> Common barriers include lack of a diverse, skilled workforce, funding, trust among the underrepresented population, institutional capacity, and political will. A community-based participatory research framework should include identifying the context of the genomic research relevant to the stakeholders, establishing a diverse cross-sector stakeholder team, creating genomic infrastructure to answer community-centered research questions, collecting data that are culture-sensitive and adaptable to stakeholder feedback, and utilizing the research results to positively impact the health of the community and local health policy.<sup>16</sup> The existing research funding mechanisms in the Organization of Economic Cooperation and Development (OECD) countries should be capitalized for equitable collaborations and capacity building in resource-limited settings.<sup>17</sup> The investments in establishing diverse international consortia for genomic research by OECD countries will return mutual benefits.<sup>18</sup> While aiming to create diverse datasets for genomic research, investigators should understand that race and ethnicity are sociopolitical constructs, and they do not equate to the genetics and ancestry of an individual.<sup>19</sup>

Accurate genomic sequence interpretation is as important as having a representative reference genome for comparison.<sup>20</sup> There are only 5200 known disease-associated genes, but more than 90,000 variants exist. Only a quarter of these variants have their pathological significance classified, while the rest of them are classified as variants of unknown significance. The classification of a huge number of known sequence variants in the population and storing them in easily accessible databases is important for the accurate interpretation of genomic data. The Genome Aggregation Database (gnomAD) is one of the largest and most widely

**Table 1.** Longitudinal Cohorts.

No.	Longitudinal Cohort	Aim	Sample size	Population	Data type	Country
1.	Million Veterans Program	To study the effects of genes, lifestyle, exposures, and military experience on veterans	Greater than 1 million (as of November 2023)	Veterans Age > 18 yrs Race—25% non-white Female—10%	Microarray, WES, WGS, methylation, proteomics, and metabolomics	United States
2.	Gabriella Miller Kids First Pediatric Research Program	To understand the biology of childhood cancers and structural birth defects	Greater than 31000	Age—0–99 yrs Race—11% non-white Female—46%	WES, WGS, RNA-Seq, linked-read WGS, long-read sequencing	United States
3.	'All of Us' research program	To collect data from EHR, clinical measurements, digital health technology data, and biospecimens genomic analysis.	Greater than 413,000	Age > 18 yrs Race—49% non-white Sexual or gender minority—13%	WGS, Genetic microarray	United States
4.	UK Biobank	To understand the genetic and non-genetic determinants of health	Greater than 500,000	Age—40–69 yrs Race—14% non-white Female – 38%	WGS, WES, and Whole-genome microarray	United Kingdom
5.	Canadian Partnership for Tomorrow's Health'	To determine the effect of genes, environment, behavior, and family health history on chronic diseases	307,017	Age—30–74 yrs Race—16% non-white Female—62%	Genotyping, WGS, WES, RNA sequencing, single-cell sequencing	Canada
6.	National Cancer Institute Cohort Consortium	To create an extensive international collaboration to combine a massive quantity of data and biospecimens and enable the conduct of cancer research at a scale and pace not possible before	58 cohorts	Age—mostly adults > 25 yrs Race- predominantly white	Genetic Array WGS, WES, and epigenomic marks	United States, Canada, Europe, Southeast Asia, Mexico, Australia, Iran
7.	International Hundred Thousand Plus Cohort Consortium	To form a consortium of large cohorts and their data linked across them	50 million participants (103 cohorts)		WGS, WES, Genetic array, transcriptomics, proteomics data	43 countries

WES whole exome sequencing, WGS whole genome sequencing.

used public resources for sequencing-based population variation data.<sup>21</sup> It is particularly valuable as a source of putatively benign variants found in the human population. Among its key features are constraint metrics, which are extensively used to support the interpretation of genes and variants linked to rare diseases. Databases such as ClinVar, the Human Gene Mutation Database (HGMD), OMIM, GeneReviews, the Clinical Genome Resource (ClinGen), DECIPHER, and Mouse Genome Informatics (MGI) are important resources for the clinical interpretation of genetic variants.<sup>22</sup> To further assist in variant interpretation, various tools have been developed. These include software that complies with ACMG/AMP guidelines and machine learning (ML) models designed to differentiate between pathogenic and benign variants. Prediction tools such as MutationTaster2, PolyPhen, MutationAssessor, CADD, DeepVariant, GATK, REVEL, and phyloP are often employed for analyzing genetic variants. ML-based variant classification and interpretation tools are advantageous compared to statistics-based predictors because they are data-driven and yield probabilistic pathogenicity scores for prioritizing variants of unknown significance.<sup>23</sup>

### GENOMICS AND OTHER OMICS

The Human Genome Project (HGP) was an ambitious collaborative team effort funded by the US National Institutes of Health involving 20 centers across six countries. It was launched in 1990 and completed in 2003.<sup>24</sup> This project used the Sanger sequencing method to sequence the open chromatic portion of the human genome and reported a sequence of 2.85 billion nucleotides (approximately 85–90% of the human genome) with minimal interruptions. HGP reported that the human genome contains only 20,000–25,000 protein-coding genes and provided a reference sequence for the future. HGP reshaped the research undertaken to understand human biology, disease states, and their treatment.<sup>25</sup> It created the need to decipher complex interactions within the human body system at microscopic and macroscopic levels, highlighting the importance of the systems biology approach in biomedical research. HGP spurred the rapid development of technologies necessary for this approach in computation, mathematics, genomic data science, and AI. Another significant contribution of HGP to biomedical research is showcasing the success and potential of team science and open-source data sharing.<sup>26</sup>

The chain termination method developed by Frederick Sanger in 1977 is considered the gold standard for DNA sequencing because of its excellent accuracy in base calling. However, this method yields low throughput because the DNA can be sequenced only in small fragments at a given time. This disadvantage was overcome by newer high throughput technologies that use massively parallel DNA sequencing platforms collectively called next-generation sequencing (NGS).<sup>27,28</sup> NGS includes various methods like sequencing by synthesis, pyrosequencing, sequencing by ligation, and ion semiconductor sequencing. Sequencing by synthesis using polymerase chain reaction (PCR) is the most widely used method for genome and exome sequencing. Technological refinements aimed at improved accuracy and reduced cost have led to significant advancements in the NGS platforms. For example, the popular HiSeq technology by Illumina had an output capacity in the range of 1.6–1.8 terabytes (Tb), 5.3–6 billion reads per run with a maximum read length of  $2 \times 150$  base pairs (bp) in the year 2014 have advanced to NovaSeq technology that can give an output in the range of 6–16 Tb, 20–52 billion reads per run with maximum read lengths of up to  $2 \times 250$  bp in 2022. There was a significant cost reduction during the same period of technological advancement. The cost of genome sequencing was reduced from 1392 USD in November 2018 to 525 USD in May 2022, per the recent US National Human Genome Research Institute data.<sup>29</sup> Recently, a newer method of

NGS developed by Ultima Genomics utilizing single-end sequencing technology using silicon wafers instead of flow cells has decreased costs to as low as \$100 per sample.<sup>30</sup>

Alternative sequencing technology that uses novel methods such as single-molecule real-time sequencing (SMRT) and nanopore sequencing (ONT) without DNA amplification by PCR has been developed in recent years.<sup>31</sup> The advantages of SMRT and ONT include the ability to sequence long fragment DNA molecules with a read length of 30–50 KB, single-nucleotide resolution, detection of base modifications like methylation patterns used in epigenetic studies, direct RNA sequencing, and proven utility in metagenomics and transcriptomics.<sup>32</sup> Another cutting-edge sequencing technology on the horizon is single-cell in-situ nucleic acid sequencing. This technology overcomes the loss of spatial information of the transcriptome when cells are separated from the tissue for RNA sequencing.<sup>33,34</sup>

*Transcriptomics*, the study of RNA transcripts including their patterns of splicing, polyadenylation, fusion, isoform quantification and non-translated functional forms, has advanced from microarray to RNA-sequencing (RNA-Seq) technology.<sup>35,36</sup> RNA-Seq uses massively parallel sequencing to identify the RNA expression at a genome level. RNA-Seq can be classified into bulk RNA-Seq, the study of transcriptional profile of whole tissue, single cell RNA-Seq (scRNA-Seq), study of transcriptional profile at an individual cellular level and spatial RNA-Seq (spRNA-Seq), study of transcriptional profile with a spatial context.<sup>37</sup> The above-mentioned technological advancements in transcriptomics have impacted our understanding and clinical care of several pediatric diseases. Recognition of gene fusions are vital for risk-based therapies to improve outcomes in children with acute lymphoblastic leukemia.<sup>38</sup> An Australian study incorporated the RNA-Seq into the routine clinical diagnostic pipeline and found enhanced risk-based classification in a cohort of pediatric ALL patients.<sup>39</sup> As a part of an ongoing multicenter clinical trial by the Dana Farber Cancer Institute ALL consortium, RNA-Seq was used to study the utility of transcriptomics to augment conventional cancer diagnostics.<sup>40</sup> The study reported that RNA-Seq identified 56 gene fusions and 141 missense mutations of clinical significance missed by routine tests. A recent study on autism spectrum disorder (ASD) utilized single nucleus RNA sequencing analysis and identified broad transcriptomic dysregulation in the brain specifically the excitatory neurons and glial cells.<sup>41</sup>

The variations in the translated proteins coded by a single gene called the proteoforms, play an essential role in the substantial diversity in the phenotype provided by the limited genes.<sup>42</sup> Advances in proteomic analysis, such as mass spectrometry and nuclear magnetic resonance mass spectroscopy, allow in-depth analysis of an individual's proteomic profile, aiming to provide phenotype-genotype correlation that helps us better understand disease pathophysiology and provide new drug targets.<sup>43</sup> A small pilot study ( $n = 6$ ) evaluated the association of focal segmental glomerulosclerosis (FSGS) with peritoneal membrane fibrosis in children undergoing peritoneal dialysis. The study used machine learning to perform comparative proteomic analysis of mesothelial exosomes in the peritoneal fluid effluent and found 40 distinct proteins to identify FSGS patients with 100% accuracy.<sup>44</sup> A retrospective observational study aimed to describe the changes in the genome and proteome in pediatric patients with acute lymphoblastic leukemia (ALL) during disease progression.<sup>45</sup> The study reported the stability of the tumor clones' genome and proteome during disease progression, demonstrating the benefit of combining proteomics with genomics in the diagnosis of pediatric ALL.

The study of metabolome, molecules smaller than proteins from biological samples, provides real-time insights into phenotypic variations in the disease processes. Small molecule metabolites are increasingly recognized for their role in physiological and pathological processes through cellular signaling mechanism,

immune function, and mediating the effects of environmental stimuli.<sup>46</sup> Metabolomics is an emerging “omics” field that can take us closer to correlating the genotype with the phenotype, including the interplay of enviotype in its analysis.<sup>47</sup> The technological advancements in mass spectrometry combined with liquid or gas chromatography have taken this field from targeted small-scale analysis of selected metabolites that are tissue-specific to a non-targeted approach that has great potential to subclassify a disease based on variations in metabolic dysfunction, identify novel biomarkers and drug targets.<sup>48,49</sup> A large prospective metabolomic study ( $n = 708$ ) including children between 18 and 48 months of age with ASD aimed to classify participants based on metabolic alterations, discovered and validated 34 metabolotypes, ratios of 39 plasma metabolite pairs related to amino acid and mitochondrial energy metabolisms, that could reliably differentiate children with ASD and typical development.<sup>50</sup> A cross-sectional study ( $n = 3290$  families) comparing the diagnostic rates of untargeted metabolomic profiling with traditional newborn metabolic testing (combination of plasma amino acids, plasma acylcarnitine profile, and urine organic acids) to identify inborn errors of metabolism found that the case finding by untargeted metabolomic approach was approximately 6-fold higher (7.1% vs 1.3%).<sup>51</sup> A sub-analysis of the Canadian Healthy Infant Longitudinal Development Study (CHILD) cohort ( $n = 647$ ) compared the fecal metabolites analyzed by nuclear magnetic resonance spectroscopy and secretory immunoglobulin A measured at 3–4 months of age, with BMI z scores at 1 and 3 years of age after controlling for environmental factors such as type of delivery, intrapartum antibiotic prophylaxis, birthweight, breastfeeding and solid food introduction.<sup>52</sup> The study found that fecal formate and butyrate levels in non-exclusively breastfed infants at 3–4 months correlated with the BMI z score at preschool age.

*Epigenomics* systematically studies the reversible, mostly heritable DNA modifications or the proteins attached to the DNA, such as DNA methylation, histone modification, chromatin accessibility, and non-coding RNA that regulate the gene expression without altering the DNA sequence.<sup>53</sup> Understanding the regulatory component that makes up 98% of our genome is extremely important to make an accurate cell-specific interpretation of the information contained in the genome. Recent technological leaps, including the NGS, have enabled the development of reference epigenomic maps for various cell and tissue types.<sup>54,55</sup> DNA methylation studies utilize platforms such as whole genome bisulfite sequencing, methylated DNA immunoprecipitation sequencing, and methylation-sensitive restriction enzyme sequencing. Studies looking at histone modifications use chromatin immunoprecipitation assay with sequencing. Investigating chromatin accessibility utilizes DNase cleavage with sequencing, high throughput chromosome conformation capture, transposase-accessible chromatin with sequencing, and chromatin loop reorganization using clustered regularly interspaced short palindromic repeats technology.<sup>56</sup> Non-coding RNAs of interest in epigenomic studies are long non-coding RNA and micro RNA. They are studied using NGS-based RNA sequencing platforms.<sup>57</sup> A US birth cohort ( $n = 954$ ) study found that in-utero exposure to maternal diabetes was associated with distinct DNA methylation patterns measured in the cord blood using genome-wide CpG analysis.<sup>58</sup> The investigators were able to map the methylated CpG sites to specific genes and showed their predictive capacity for preterm birth. An epigenome-wide meta-analysis conducted by the Pregnancy and Childhood consortium, including 17 cohorts with 1299 participants (668 newborns and 631 children), found an association between DNA methylation patterns (9 CpG sites and 35 regions) in the newborn period and risk of school-age asthma.<sup>59</sup> In children, 179 differentially methylated CpG sites and 36 regions predicted the risk and severity of asthma.

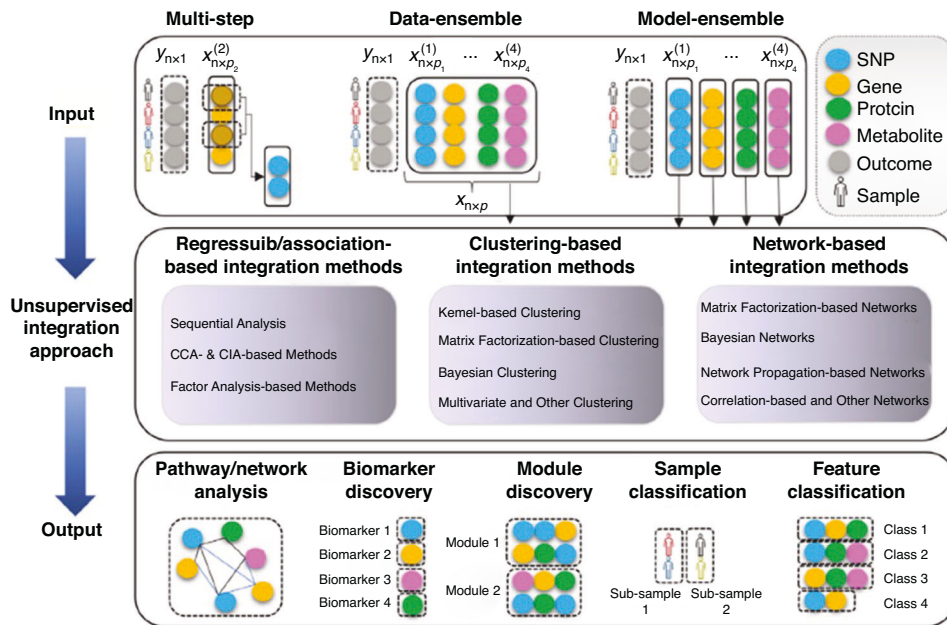
Exposomes are an aggregate of environmental, dietary, behavioral, and microbial exposures in a person's life combined with the biological responses to such exposures.<sup>60</sup> The systematic study of exposomes can be considered an extension of metabolomics, including chemicals in the environment and diet and those produced in the body as a response to sleep, regular exercise, or a lack thereof. The emerging field of exposomes holds excellent promise but currently lags behind other omics technologies.<sup>61</sup> Hu et al. recently proposed a novel workflow to analyze and quantify human exposomes using an untargeted approach using gas chromatography high-resolution mass spectrometry.<sup>62</sup>

The adverse in-utero and early life environment, exposures, and nutrition can impact the epigenome of the fetus and newborn.<sup>63</sup> For example, intraamniotic infection results in systemic fetal inflammatory response characterized by cytokine storm that affects the DNA methylation pattern in the fetus.<sup>64,65</sup> Maternal exposure to illicit drugs like opioids, cocaine, and cannabis can cause epigenetic changes in the fetus.<sup>66,67</sup> Infants born prematurely are exposed to life saving measures with yet unknown long term effects such as excessive oxygen and corticosteroids that may have an effect on their epigenome.<sup>68</sup> The epigenetic changes secondary to periconceptual, in-utero and early-life environmental stressors of the fetal and newborn period should be factored in the interpretation of the integrated multiomics data and their clinical relevance while taking an genome-first approach to precision medicine.

An ongoing prospective population-based birth cohort study in six European countries recruited 31,472 mother-child dyads and selected a sub-cohort of 1033 mother-child dyads for the initial exposome-wide association study. Eighty-five prenatal and 125 postnatal exposures and the child's lung function at 6–12 years of age were evaluated.<sup>69,70</sup> The study found that prenatal exposure to perfluorononanoate and perfluorooctanoate was associated with abnormal lung function. Postnatal exposure to copper, ethylparaben, phthalate metabolites, house crowding, and facility density around schools were associated with abnormal lung function. Another large prospective cohort study is underway in Europe to characterize the human exposomes in working life.<sup>71</sup> The Exposome Project for Health and Occupational Research aims to develop detection and analytical platforms with high-throughput technologies to enable future research in this domain.

## INTEGRATED MULTIOMICS

Integrative multiomics, the combination of multiple omics data layered over each other, including the interconnections and interactions between them, helps us understand human health and disease better than any of them separately.<sup>72,73</sup> Several approaches and computational platforms are available to achieve this integration. The integration process can be classified in several ways. One type of categorization of the integrative approach is early (concatenation-based), mixed (transformation based), intermediate, late (model-based), and hierarchical.<sup>74</sup> Another broad way of classifying the integration platforms is supervised and unsupervised based on whether machine learning models are trained using data labels.<sup>75</sup> Multiomics integration methodologies can be classified into the following approaches: 1. Regression/association-based methods, 2. Clustering-based methods, and 3. Network-based methods. Each approach uses one of the three methodologies, namely a. multistep and sequential analysis, b. data-ensemble, and c. model-ensemble. (Fig. 2) In the data ensemble method, the different multiomics data layers are linked and combined to form a single input for analysis. In contrast, in the model ensemble method, each omics data layer is analyzed separately and fused for the integrated analysis. The bioinformatics tools available to achieve integration are ever-expanding, and the choice depends on the analysis's objective.



**Fig. 2 Approaches to unsupervised multiomics data integration.** 1. Multi-step approach utilizes one of the three regression/association-based statistical methods for integration—sequence analysis, canonical correlation analysis/ co-inertia analysis, or factor analysis. 2. The data ensemble approach creates unified multiomics input data using one of the four clustering-based statistical integration methods (kernel-based, matrix factorization-based, Bayesian, or multivariate clustering). 3. The model ensemble approach utilizes one of the four network-based statistical methods (matrix factorization, Bayesian, Network propagation, or correlation-based) to analyze each multiomics data separately and combine the results to discover novel biomarkers, functionally similar modules, and classify samples. Reproduced from open access source Vahabi, N., & Michailidis, G. (2022). Unsupervised Multiomics Data Integration Methods: A Comprehensive Review.

These complex tools are extensively used in pediatric oncology and other pediatric diseases. A multicenter prospective cohort study ( $n = 221$ ) of respiratory syncytial virus bronchiolitis used affinity matrix, similarity network fusion, and spectral clustering method to integrate multiomics data (clinical, virus, nasopharyngeal microbiome, transcriptome, and metabolome data) and found four distinct clinically relevant disease endotypes.<sup>76</sup> A Canadian study based on a South Asian birth cohort ( $n = 50$ ) found novel biomarkers of early-onset childhood obesity by integrating gut microbiome (55 bacterial amplicon sequencing variants) and serum metabolome (73 serum metabolites) profiles using a supervised machine learning tool called Data Integration Analysis for Biomarker discovery using Latent cOMponent (DIABLO).<sup>77</sup>

### EHR DATA INTEGRATION WITH MULTIOMICS

Integrated multiomics data coupled with clinical context is essential to accomplish the aim of precision medicine. Electronic Health Records (EHR) contain structured data such as administrative information, laboratory parameters, vital signs, anthropometric measurements, medications, and diagnosis codes.<sup>78</sup> The clinical notes from various healthcare providers (primary and consulting physicians, registered nurses, respiratory therapists, pharmacists, dieticians, and social workers) and medical imaging constitute the unstructured data. The combined data from the EHR can help provide the clinical context for interpreting multiomics, and hence, integrating EHR with multiomics is crucial for creating clinically relevant and actionable data. The first step in this process, combining structured and unstructured data, could be accomplished using technologies such as Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR).<sup>79</sup> The next step is integrating curated EHR data with multiomics utilizing AI-powered Big Data analytics. To accomplish this, we need high-quality data from EHR that incorporates patient heterogeneity and diversity to allow machine learning (ML) tools to reproduce

reliable output. One way to achieve this is by national or international multi-center studies, but patient privacy and data security issues are potential barriers. Newer ML platforms, such as Federated Learning allow institutions to train their models locally with patient data. Then, the institutions share trained models without patient data.<sup>80</sup> The models shared by various institutions are aggregated to create a robust consensus model.<sup>81</sup> Application programming interface (API) and related Substitutable Medical Apps, Reusable Technologies (SMART®) platform are showing promise in this process of integration.<sup>82</sup> A bioinformatics study demonstrated that an integrated SMART clinic-genomics app can be created by developing separate SMART on FHIR clinical and genomic data adapters to directly payload data from the EHR and sequencer.<sup>83</sup> Examples of clinic-genomic apps include Precision Medicine Genes +AI and SMART Precision Cancer Medicine (PCM).<sup>84</sup> PCM is an EHR-agnostic open-source software that helps clinicians visualize patient-specific genomic data in the context of a population-level spectrum within their clinical workflow.

### DATABASES

The first step in harnessing the potential of integrative multiomics is creating a scalable, collaborative, secure digital environment that can be used to identify, collect, share, and retrieve genomic and multiomics data. Such digital infrastructures are essential for making the best use of large datasets to help deliver customized healthcare at a population level. Table 2 lists a few examples of such genomic data infrastructure development initiatives.

### DATA MINING AND BIOINFORMATICS

Bioinformatics, the science of using computational methods to analyze and understand biological data, utilizes data mining and knowledge discovery in databases, two interlinked processes for unearthing trends and patterns in large genomic and multiomics databases to generate novel clinically valuable insights and

**Table 2.** Genomic Data Infrastructure Development Initiatives.

No.	Name	Description	Collaborators
1.	The Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative	A cloud-based data-sharing undertaking by NIH to help biomedical researchers	Google Cloud, Amazon Web Services, and Microsoft Azure
2.	NIH Cloud Platform Interoperability (NCPI) program	A federated genomic data ecosystem in which the researchers can discover and merge data from different platforms, such as Analysis Visualization and Informatics Lab-space (AnVIL), BioData Catalyst, Cancer Research Data Commons (CRDC), Gabriella Miller Kids First Pediatric Research Program, Database of Genotypes and Phenotypes (dbGaP), Sequence Read Archive (SRA)	NHGRI, NHLBI, NCI, NIH Common Fund, NCBI and NIH
3.	European Genomic Data Infrastructure (GDI) program	Transnational access to genomic, related phenotypic, and clinical data to support the 1+ Million Genomes initiative	20 member states of the European Union

NHGRI National Human Genome Research Institute, NHLBI National Heart, Lung, and Blood Institute, NCI National Cancer Institute, NCBI National Center for Biotechnology Information, NIH National Institutes of Health.

applications.<sup>85</sup> Data mining relies on machine learning and advanced statistical methods to recognize patterns in clinical big data.<sup>86</sup>

The investments made in developing large bioinformatics datasets worldwide have provided a tremendous opportunity for data mining. Some public genomic databases that could be used for data mining were mentioned earlier. Others include Gene Expression Omnibus (GEO), Cancer Genome Atlas, International Cancer Genome Consortium, Comparative Toxicogenomic Database, and Biologic Specimen and Data Repositories Information Coordinating Center. The data mining process involves three main steps—identifying an appropriate model (predictive or descriptive), choosing a task to implement the model (classification, estimation, prediction, association, clustering, and descriptive visualization), and using analytical methods to complete the process.

Let us outline an example to help understand the opportunities and challenges in bioinformatics data mining. GEO is a public repository containing functional genomics datasets by the global research community, maintained by the US National Center for Biotechnology Information. The original data supplied by the researchers are curated into study-level records called GEO datasets and gene-level records called GEO profiles that can be queried and analyzed by other researchers using internet-based data mining tools.<sup>87</sup> A GEO dataset-based study used gene set enrichment analysis to find the relevance of MCYN-associated genes in predicting the prognosis of pediatric neuroblastoma.<sup>88</sup> Although genomic data mining opens up many research avenues, the data in the public repositories are underutilized, especially in pediatrics. The challenge researchers face in multiomics data reuse from these huge repositories is the technical complexity of accessing and analyzing data. Several user-friendly software applications have been developed to overcome this problem. Examples of such applications include DataMed – an open-source biomedical data discovery system; GEO RNA-seq Experiments Interactive Navigator (GREIN) – an interactive web platform for re-analyzing GEO RNA-seq data; OMICtools- directory of more than 4400 web-accessible tools related to genomics, transcriptomics, proteomics and metabolomics, Datasets2Tools – search engine for bioinformatics datasets, tools and canned analyses.<sup>89–92</sup>

## CLINICAL AND RESEARCH APPLICATIONS

### Clinical applications

We will review some of the advances in integrative multiomics and genomics that have been clinically studied that promise to deliver precision medicine to the bedside in specific areas of

pediatric oncology, pediatric critical care, and pediatric diabetes. Tissue-specific transcriptome analysis or RNA-sequencing has been increasingly utilized in research studies for resolving undiagnosed diseases with diagnostic yield ranging from 7 to 35%.<sup>93</sup> More recently, it is being implemented in clinical laboratory reporting for the resolution of non-coding and splice variants detected by genome sequencing. The use of RNA-seq with genome sequencing promises to enhance the interpretation of such genetic variants for improved clinical care.

A study utilized a multiscale RNA clustering approach to define pediatric cancers at a molecular level and applied the method to a retrospective dataset of 13313 transcriptomes to classify pediatric cancers. The new diagnostic classification was used to design a deep learning model and validated in a prospective cohort with 85% accuracy.<sup>94</sup> A population-based pediatric neuro-oncology study aimed to identify the molecular neuropathology of pediatric CNS tumors.<sup>95</sup> The authors developed a neuro-oncology-specific NGS gene panel and a DNA methylation-based classification system to achieve this. Both the omics data were integrated with the histopathology reports. Multiomics integration improved diagnostic accuracy, aided detection of mutations relevant to diagnosis and treatment, identified precancerous syndromes, and predicted prognosis better. A study aimed to evaluate the combination of exome and transcriptome sequencing and SNP array to provide precision cancer therapy for children and young adults ( $n = 59$ ) with relapsed and refractory non-CNS solid tumors.<sup>96</sup> The investigators found that 51% of the study participants had clinically actionable mutations, including germline and somatic events (including a single nucleotide variant, an amplification, a deletion, an indel, or a fusion gene), that could be targeted in the ongoing clinical trials.

An Australian cohort study included 290 critically ill infants and children referred from neonatal intensive care units (47%), pediatric intensive care units (39%), and other hospital units nationwide to the Acute Care Genomics program comprising a national panel of experts.<sup>97</sup> The trio analysis (patient and both parents) was done in 94% of the cohort. Ultra-rapid genome sequencing was the initial testing offered, followed by transcriptomics and proteomics analysis if they remained undiagnosed after genome sequencing alone. The integrative multiomics approach increased the diagnostic yield from 47% to 54%, with an impactful alteration in the management plan in approximately 1/3rd of the cohort. A systematic review including 21 prospective studies ( $n = 1654$ ) aimed to synthesize the clinical utility of exome and genome sequencing in critically ill infants younger than 1 year. The study categorized the utility of genomics results as treatment change, redirection of care, prognostic information,

reproductive information, and screening or subspecialty referral. The review found that a significant percentage of critically ill infants (mean—37% [13–61%]) undergoing genomic evaluation experienced clinical utility.

Armenteros et al. reported the molecular signatures in newly diagnosed type 1 diabetes patients ( $n = 97$ ) from the pan-European consortium that predicted the rate of beta cell loss after the diagnosis using multiomics (integrated transcriptomic, genomic, targeted proteomic, lipidomic, metabolomic, and immunomic data) factor analysis.<sup>98</sup> The two specific molecular signatures predictive of rapidly declining beta cell functional loss were identified. One of them was associated with immune pathways, and the other correlated with viral infection. The study showed the promise of using multiomics to identify newly diagnosed type 1 diabetic patients at risk of rapid disease progression and design multiomics-informed clinical trials, enabling the development of precision therapeutics.

The selected studies discussed above showed the usefulness of integrated multiomics in the diagnosis and management of pediatric cancers, critical illness, and diabetes. However, the relevance of the availability of these technologies, their utility, and applicability in clinical care spans various spheres of pediatric healthcare.

### Research applications

Traditionally, genomics research hypotheses have been postulated with the phenotype-first approach. The reasons for the phenotype-first approach in genomics research include readily available clinical data from EHR, easily traceable family records, information regarding exposomes that can be inferred from routinely collected information such as zip codes, and infrastructures for data sharing that already exists or at least easier to establish.<sup>99</sup> However, this approach creates challenges and limits the yield of clinically helpful knowledge from the research endeavors. The EHR records are incomplete and not uniform across institutions, zip code/ insurance-based estimation of exposomes, including psychosocial stress, provides cross-sectional data and does not consider the lifetime experience of the participants.<sup>100,101</sup> There are difficulties in scaling the datasets

on a global stage because of the differences in the healthcare delivery models in LMIC countries and ascertainment bias limiting our ability to delineate penetrance, expressivity, and variant pathogenicity.<sup>100,101</sup> Thus, the phenotype-first approach has drawbacks and is not ideal for research or for advancing precision medicine in children.

The genotype-first approach in genomic and multiomics research can overcome the limitations of the phenotype ascertainment and clinical informatics approaches and enhance the subclassification of diseases by widening the phenotypic spectrum and reverse phenotyping.<sup>101,102</sup> Reverse phenotyping, where the genotypes of the research participants are known at recruitment, and a hypothesis is formulated to correlate with new phenotypic data, forms the basis for genomic ascertainment research. Genomic ascertainment research can help us uncover unrecognized or undiagnosed phenotypes, find an association between a new genotype and disease, and conduct an ex vivo phenotype analysis. The National Human Genome Research Institute (NHGRI) Reverse Phenotyping Core has piloted this research approach, mostly in adults utilizing the ClinSeq cohort, highlighting the feasibility of this approach.<sup>103,104</sup> The tenets of the reverse phenotyping research strategy include an informed consent process that allows recontact and data sharing, sustaining the participation of the research subjects, institutions, and stakeholders, clarity and transparency in sharing results with participants, and networking to recruit large cohorts. A framework and challenges of conducting reverse phenotyping research and their possible solutions are presented in Table 3.<sup>101</sup>

A study aimed to find the relationship of gene-gene interactions (epistasis) in the pathophysiology of ASD used the genotype-first approach and found that single nucleotide polymorphisms in the Ras/MAPK pathway contribute to idiopathic ASD.<sup>105</sup> The study identified that dysregulation of the GPR141 gene may have a role in idiopathic ASD. A study of steroid-resistant nephrotic syndrome utilized exome sequencing and a reversed phenotyping strategy to identify novel clinical signs and prognostic factors associated with genetic nephropathies that could be differentiated from podocytopathies.<sup>106</sup> Thus, a genome-first approach overcomes the drawbacks of a phenotype-first approach and can identify

**Table 3.** A. A Framework for a Genotype-First Approach B. Challenges in Genotype-First or Reverse Phenotyping Approach in Pediatrics.

A		
A. Create a strategic plan at the outset and secure written agreement for pledged commitments from stakeholders		
B. Include broad data sharing and the ability to recontact participants in the genomic research informed consent process		
C. Establish long-term, trusting relationships with study participants: success depends on the participants' willingness to return for studies		
D. Generate and maintain institutional engagement and support for follow-up studies: establishing and maintaining a reverse phenotyping resource requires institutional commitment and material resources of staff, money, and time.		
E. Define what results will be returned, to whom, and by whom: lack of clarity around which study participants will receive results and who is responsible for delivery of these results can diminish participants' trust and risk their participation in future studies.		
F. Invest in adding new cohorts through networking with other investigators: there is power in numbers, especially for rare variants, so it is crucial to have a large pool of participants from which to draw.		
B		
No.	Challenges	Potential Solution
1.	Need for long-term follow-up of the pediatric patients and their families	Include a follow-up provision in all informed consent forms for research with a clear definition of the purpose
2.	Need for storing genomic data in open-science databases	Strengthen the data infrastructure to prevent misuse and help parents make an informed choice about data sharing in the research community
3.	Cost burden of long-term follow-up and data storage	Communicate the impact and return for investment to stakeholders in tangible terms
4.	Ethical conundrum in sharing results of genomic research with patients and their families	Investigators should develop a strong plan for returning research findings during the design phase and should be mandated by the institutional review boards before study approval
5.	Recruiting large prospective cohorts of representative diversity	Foster global networks and community partnerships

Adapted from "Genotype first: Clinical genomics research through a reverse phenotyping approach" by Wilczewski et al. Cell Press Open Access.

multiple phenotypes for the same gene defect, quantify the gene-dose effect, and help in prognostication.

### ROAD MAP TO AI-BASED INTEGRATIVE MULTIOMICS APPROACH FOR PRECISION MEDICINE IN PEDIATRIC HEALTHCARE AND RESEARCH

We present a road map for implementing integrative multiomics in pediatric clinical care. (Fig. 1) Large prospective pediatric cohorts with representative diversity should be developed with global public and private partnerships with secure and scalable data infrastructure. Multiomics data from such a cohort of patients should be integrated using artificial intelligence/machine learning tools described earlier in the review. Such integrated multiomics data should be combined with curated EHR data to create databases and data warehouses with an open-science model. Clinicians and researchers should be equipped with user-friendly, accessible bioinformatics tools to access the data warehouses with a genomics-first approach. This method helps clinicians precisely subtype patients with complex diseases with overlapping phenotypes, and researchers discover accurate biomarkers and targeted therapeutics. Such a road map would advance patient-centered, value-based precision healthcare in pediatrics.

### PRECISION MEDICINE—CURRENT CHALLENGES AND THE FUTURE

The challenges facing implementing precision medicine in routine clinical practice can relate to 1. Data, 2. Cost 3. Bioethics, and 4. Legal issues. We need huge scalable, and interoperable data for the AI/ML models to prove they can make accurate predictions for widespread implementation in clinical care.<sup>4</sup> We need to establish accessible digital infrastructures to collect, store, and share data of large magnitudes in compliance with FAIR guidelines.<sup>107</sup> We need data that would inform us about the human genome's regulatory/non-coding component, which will complete the puzzle and enhance our understanding.<sup>108</sup> We need a diverse population in our datasets so that the ML models are not biased in their prediction.<sup>109</sup> We must develop capabilities for out-of-sample cross-validation of the AI/ML models to produce generalizable data. AI models must recognize heterogeneity in the disease mechanisms from population data having the same phenotypic labels. Data intended for precision medicine must capture patient-centered outcome measures, and the predictions should be tuned to factor in those outcomes while predicting treatment responses. Implementing multiomics data collection and integration with the EHR at a population level would incur excessive costs. The analytic tools needed to handle integrated multiomics-EHR big data cannot be made accessible without the will and investment of policymakers. Healthcare systems dependent on medical insurance reimbursement need the payers' acceptance and support for the practice of precision medicine. Existing reimbursement models will be insufficient to sustain the widespread implementation of precision medicine.<sup>110</sup> The biggest ethical challenge facing the implementation of precision medicine is the balance that needs to be struck between inclusivity and diversity in the big datasets and the protection from potential harms from breaches in data safety and privacy, implications in future insurance premiums, discrimination in society, and employment. On the legal and regulatory front in the US, the genomics and multiomics tests are categorized as laboratory-developed tests (LDT).<sup>111,112</sup> LDTs are currently not regulated by the Food and Drug Administration. So, LDTs can be marketed without proof of clinical validity and utility.

The first step in realizing the hope of precision medicine is delineating the existing infrastructure and barriers to advancing precision medicine at a local, regional, and global level.<sup>113</sup> The second step is to create a collaborative network of various stakeholders - physicians, other healthcare providers, researchers,

public and private research funding agencies, healthcare systems, academic institutions, healthcare payers, industry partners from genomics and multiomics diagnostic, pharmaceuticals and biological, bioinformatics, governmental regulatory agencies, and public health policymakers—with a shared vision of precision medicine. Such a network must include patients, their families, and community support groups. The third step is to invest in the education and training of the workforce, including clinical geneticists, computational genomics, and data scientists, to implement precision medicine sustainably.<sup>114</sup> The current generation of healthcare providers has to be provided with the necessary knowledge and skills to adopt and incorporate the practice of precision medicine into traditional practice. A diverse precision medicine workforce should be developed in healthcare, medical research, data science, bioinformatics, computational, and systems biology for the future demands of the field. Patients have to be educated about the benefits of precision medicine and encouraged to participate and contribute to the diversity of the data. Fourth, an informed consent process should be streamlined to enable patients from diverse socioeconomic strata to participate in genomics research. Fifth, multiomics and stored EHR data should be regulated by law to protect patient privacy, prevent discrimination, and make data access and reuse easier. Health Insurance Portability and Accountability Act (HIPAA) laws ensure genomic data in EHRs are protected from unauthorized access, requiring EHR systems to follow strict regulations for the secure handling, storage, and transmission of genetic data to maintain patient privacy. The Genetic Information Nondiscrimination Act of 2008 (GINA) prohibits discrimination based on genetic information in health insurance and employment. However, its protections are limited and do not extend to other types of insurance, such as long-term disability coverage or life insurance. These patient safeguards have to be strengthened further. Sixth, a combination of government-initiated “top-down” and healthcare systems-initiated “bottom-up” approaches should be pursued in precision medicine research, innovation, and implementation.<sup>115</sup> Seventh, for the successful implementation of precision medicine, the insurance reimbursement model must change from fee-for-service to value-based healthcare.<sup>116</sup> The health systems assessment framework should become patient-centered and reformed by identifying and obtaining consensus from various stakeholders.<sup>117</sup>

Innovations in genomics, multiomics, and AI have begun to create a paradigm shift in our understanding of human biology. The patient-centered precision approach to diagnosing and treating certain diseases like cancer has shown us a glimpse of the future landscape of healthcare. Research aimed at translating the successes of precision oncology into other common public health problems like neuropsychiatric and cardiometabolic disorders, has informed us of the potential to implement precision medicine at a population level. Stakeholders' confidence in the promise of precision medicine has enabled national and global collaborations on a scale unseen in the past to establish the infrastructure needed for the sustained innovation and implementation of precision medicine. Adopting the principles of precision medicine in the research of disorders affecting newborn infants, children, and adolescents has shown us how to bring about patient-centered, value-based transformation in the practice of pediatrics and its subspecialties. We are far from implementing precision medicine in our day-to-day practice. However, we know the pathway to be taken to get there. We envision datasets similar to the scale of All of Us that would be created to include children with built-in ethical protections. Common pediatric public health burdens such as neurodevelopmental disorders, asthma, obesity, and prematurity would be subtyped based on distinct etiopathogenic mechanisms, diagnosed early by novel, accurate biomarkers, treated effectively with targeted therapeutics, prognosticated better by leveraging integrated multiomics linked with EHR and AI.

## REFERENCES

1. Delpierre, C. & Lefèvre, T. Precision and personalized medicine: what their current definition says and silences about the model of health they promote. implication for the development of personalized health. *Front. Socio.* **8**, 1112159 (2023).
2. Abrahams, E., Ginsburg, G. S. & Silver, M. The personalized medicine coalition: goals and strategies. *Am. J. Pharmacogenom.* **5**, 345–355 (2005).
3. Dzau, V. J. & Ginsburg, G. S. Realizing the full potential of precision medicine in health and health care. *Jama* **316**, 1659–1660 (2016).
4. Denny, J. C. & Collins, F. S. Precision medicine in 2030—seven ways to transform healthcare. *Cell* **184**, 1415–1419 (2021).
5. Henderson, R. H. et al. Delivering the precision oncology paradigm: reduced R&D costs and greater return on investment through a companion diagnostic informed precision oncology medicines approach. *J. Pharm. Policy Pract.* **16**, 84 (2023).
6. Kingsmore, S. F., Nofsinger, R. & Ellsworth, K. Rapid genomic sequencing for genetic disease diagnosis and therapy in intensive care units: a review. *NPJ Genom. Med.* **9**, 17 (2024).
7. Sanford Kobayashi, E. et al. Cost efficacy of rapid whole genome sequencing in the pediatric intensive care unit. *Front Pediatr.* **9**, 809536 (2021).
8. Hofker, M. H., Fu, J. & Wijmenga, C. The genome revolution and its role in understanding complex diseases. *Biochim. Biophys. Acta Mol. Basis Dis.* **1842**, 1889–1895 (2014).
9. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat. Rev. Mol. Cell Biol.* **24**, 695–713 (2023).
10. Rockowitz, S. et al. Children's rare disease cohorts: an integrative research and clinical genomics initiative. *NPJ Genom. Med.* **5**, 29 (2020).
11. Ma, X., Wang, P., Xu, G., Yu, F. & Ma, Y. Integrative genomics analysis of various omics data and networks identify risk genes and variants vulnerable to childhood-onset asthma. *BMC Med. Genomics* **13**, 123 (2020).
12. Choi, K. W. et al. Integrative analysis of genomic and exposomic influences on youth mental health. *J. Child Psychol. Psychiatry* **63**, 1196–1205 (2022).
13. Jooma, S., Hahn, M. J., Hindorff, L. A. & Bonham, V. L. Defining and achieving health equity in genomic medicine. *Ethn. Dis.* **29**, 173–178 (2019).
14. Fatumo, S. et al. A roadmap to increase diversity in genomic studies. *Nat. Med* **28**, 243–250 (2022).
15. Wong, K. H. Y. et al. Towards a reference genome that captures global genetic diversity. *Nat. Commun.* **11**, 5482 (2020).
16. Rebbeck, T. R. et al. A framework for promoting diversity, equity, and inclusion in genetics and genomics research. *JAMA Health Forum* **3**, e220603 (2022).
17. Martin, A. R. et al. Increasing diversity in genomics requires investment in equitable partnerships and capacity building. *Nat. Genet.* **54**, 740–745 (2022).
18. Skanharajah, N. et al. Equity, diversity, and inclusion at the Global Alliance for Genomics and Health. *Cell Genom.* **3**, 100386 (2023).
19. Duello, T. M., Rivedal, S., Wickland, C. & Weller, A. Race and genetics versus 'race' in genetics: a systematic review of the use of African Ancestry in genetic studies. *Evol. Med Public Health* **9**, 232–245 (2021).
20. Bean, L. H. & Hegde, M. Gene variant databases and sharing: creating a global genomic variant database for personalized medicine. *Hum. Mutat.* **37**, 559–563 (2016).
21. Gudmundsson, S. et al. Variant interpretation using population databases: lessons from Gnomad. *Hum. Mutat.* **43**, 1012–1030 (2021).
22. Landrum, M. et al. Clinvar: improvements to accessing data. *Nucleic Acids Res.* **48**, D835–D844 (2020).
23. Nicora, G., Zucca, S., Limongelli, I., Bellazzi, R. & Magni, P. A machine learning approach based on Acmg/Amp guidelines for genomic variant classification and prioritization. *Sci. Rep.* **12**, 2517 (2022).
24. International Human Genome Sequencing, C. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
25. Hood, L. & Rowen, L. The Human Genome Project: big science transforms biology and medicine. *Genome Med.* **5**, 79 (2013).
26. Maxson Jones, K., Ankeny, R. A. & Cook-Deegan, R. The Bermuda Triangle: the pragmatics, policies, and principles for data sharing in the history of the Human Genome Project. *J. Hist. Biol.* **51**, 693–805 (2018).
27. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).
28. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* **122**, e59 (2018).
29. KA., W. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (Gsp).
30. Pennisi, E. Upstart DNA sequencers could be a 'game changer. *Science* **376**, 1257–1258 (2022).
31. Xiao, T. & Zhou, W. The third generation sequencing: the advanced approach to genetic diseases. *Transl. Pediatr.* **9**, 163–173 (2020).
32. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 (2020).
33. Ke, R., Mignardi, M., Hauling, T. & Nilsson, M. Fourth generation of next-generation sequencing technologies: promise and consequences. *Hum. Mutat.* **37**, 1363–1367 (2016).
34. Jovic, D. et al. Single-cell RNA sequencing technologies and applications: a brief overview. *Clin. Transl. Med.* **12**, e694 (2022).
35. Wirka, R., Pjanic, M. & Quertermous, T. Advances in transcriptomics. *Circ. Res.* **122**, 1200–1220 (2018).
36. Tsimberidou, A. M., Fountzilas, E., Bleris, L. & Kurzrock, R. Transcriptomics and solid tumors: the next frontier in precision cancer medicine. *Semin Cancer Biol.* **84**, 50–59 (2022).
37. Li, X. & Wang, C. -Y. From bulk, single-cell to spatial RNA sequencing. *Int. J. Oral. Sci.* **13**, 36 (2021).
38. Lilljebjörn, H., Orsmark-Pietras, C., Mitelman, F., Hagström-Andersson, A. & Fioretos, T. Transcriptomics paving the way for improved diagnostics and precision medicine of acute leukemia. *Semin. Cancer Biol.* **84**, 40–49 (2022).
39. Brown, L. et al. The application of RNA sequencing for the diagnosis and genomic classification of pediatric acute lymphoblastic leukemia. *Blood Adv.* **4**, 930–942 (2020).
40. Tran, T. et al. Whole-transcriptome analysis in acute lymphoblastic leukemia: a report from the DFCI All Consortium Protocol 16-001. *Blood Adv.* **6**, 1329–1341 (2022).
41. Gandal, M. et al. Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
42. Smith, L. M. & Kelleher, N. L. Proteoform: a single term describing protein complexity. *Nat. Methods* **10**, 186–187 (2013).
43. Adhikari, S. et al. A high-stringency blueprint of the human proteome. *Nat. Commun.* **11**, 5301 (2020).
44. Bruschi, M. et al. Proteomic profile of mesothelial exosomes isolated from peritoneal dialysis effluent of children with focal segmental glomerulosclerosis. *Sci. Rep.* **11**, 20807 (2021).
45. Lorentzian, A. C. et al. Targetable lesions and proteomes predict therapy sensitivity through disease evolution in pediatric acute lymphoblastic leukemia. *Nat. Commun.* **14**, 7161 (2023).
46. Wishart, D. Metabolomics for investigating physiological and pathophysiological processes. *Physiol. Rev.* **99**, 1819–1875 (2019).
47. DeBerardinis, R. J. & Keshari, K. R. Metabolic analysis as a driver for discovery, diagnosis, and therapy. *Cell* **185**, 2678–2689 (2022).
48. Qiu, S. et al. Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduct. Target. Ther.* **8**, 132 (2023).
49. Chen, Z. Z. et al. Nontargeted and targeted metabolomic profiling reveals novel metabolite biomarkers of incident diabetes in African Americans. *Diabetes* **71**, 2426–2437 (2022).
50. Smith, A. M. et al. A metabolomics approach to screening for autism risk in the children's autism metabolome project. *Autism Res* **13**, 1270–1285 (2020).
51. Liu, N. et al. Comparison of untargeted metabolomic profiling vs traditional metabolic screening to identify inborn errors of metabolism. *JAMA Netw. Open* **4**, e2114155–e2114155 (2021).
52. Bridgman, S. et al. Childhood body mass index and associations with infant gut metabolites and secretory IGA: findings from a prospective cohort study. *Int. J. Obes.* **46**, 1712–1719 (2022).
53. Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).
54. Bernstein, B. E. et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
55. Satterlee, J. S. et al. The NIH Common Fund/Roadmap Epigenomics Program: successes of a comprehensive consortium. *Sci. Adv.* **5**, (2019). eaaw6507.
56. Wang, K. C. & Chang, H. Y. Epigenomics: technologies and applications. *Circ. Res* **122**, 1191–1199 (2018).
57. Sun, Y. -M. & Chen, Y. -Q. Principles and innovative technologies for decrypting noncoding RNAs: from discovery and functional prediction to clinical application. *J. Hematol. Oncol.* **13**, 109 (2020).
58. Wang, G. et al. Impact of intrauterine exposure to maternal diabetes on preterm birth: fetal DNA methylation alteration is an important mediator. *Clin. Epigenet.* **15**, 59 (2023).
59. Reese, S. E. et al. Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J. Allergy Clin. Immunol.* **143**, 2062–2074 (2019).
60. Miller, G. W. & Jones, D. P. The nature of nurture: refining the definition of the exposome. *Toxicol. Sci.* **137**, 1–2 (2014).
61. Jones, D. P. Sequencing the exposome: a call to action. *Toxicol. Rep.* **3**, 29–45 (2016).
62. Hu, X. et al. A scalable workflow to characterize the human exposome. *Nat. Commun.* **12**, 5575 (2021).
63. Hari Gopal, S., Alenghat, T. & Pammi, M. Early life epigenetics and childhood outcomes: a scoping review. *Pediatr. Res.* (2024).
64. Konwar, C. et al. DNA Methylation profiling of acute chorioamnionitis-associated placentas and fetal membranes: insights into epigenetic variation in spontaneous preterm births. *Epigenet. Chromatin* **11**, 63 (2018).

65. Kleeman, E., Gubert, C. & Hannan, A. Transgenerational epigenetic impacts of prenatal infection on offspring health and disease susceptibility. *Trends Genet.* **38**, 662–675 (2022).
66. Knopik, V., Marceau, K., Bidwell, L. C. & Rolan, E. Prenatal substance exposure and offspring development: Does DNA methylation play a role?. *Neurotoxicol. Teratol.* **71**, 50–63 (2019).
67. Noble, A., Adams, A., Satsangi, J., Boden, J. & Osborne, A. Prenatal cannabis exposure is associated with alterations in offspring DNA methylation at genes involved in neurodevelopment, across the life course. *Mol. Psychiatr.* (2024).
68. Lorente-Pozo, S. et al. Oxygen in the neonatal period: oxidative stress, oxygen load and epigenetic changes. *Semin. Fetal Neonatal Med.* **25**, 101090 (2020).
69. Agier, L. et al. Early-life exposome and lung function in children in Europe: an analysis of data from the longitudinal, population-based Helix cohort. *Lancet Planet. Health* **3**, e81–e92 (2019).
70. Maitre, L. et al. Human early life exposome (Helix) study: a European population-based exposome cohort. *BMJ Open* **8**, e021311 (2018).
71. Pronk, A. et al. Applying the exposome concept to working life health: the Eu Euphor Project. *Environ. Epidemiol.* **6**, e185 (2022).
72. Sun, Y. V. & Hu, Y. J. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv. Genet.* **93**, 147–190 (2016).
73. Karczewski, K. J. & Snyder, M. P. Integrative omics for health and disease. *Nat. Rev. Genet.* **19**, 299–310 (2018).
74. Picard, M., Scott-Boyer, M. P., Bodein, A., Périn, O. & Droit, A. Integration strategies of multi-omics data for machine learning analysis. *Comput. Struct. Biotechnol. J.* **19**, 3735–3746 (2021).
75. Vahabi, N. & Michailidis, G. Unsupervised multi-omics data integration methods: a comprehensive review. *Front. Genet.* **13**, 854752 (2022).
76. Raita, Y. et al. Integrated omics endotyping of infants with respiratory syncytial virus bronchiolitis and risk of childhood asthma. *Nat. Commun.* **12**, 3601 (2021).
77. Rafiq, T. et al. Integrative multiomics analysis of infant gut microbiome and serum metabolome reveals key molecular biomarkers of early onset childhood obesity. *Heliyon* **9**, e16651 (2023).
78. Tong, L. et al. Integrating multi-omics data with EHR for precision medicine using advanced artificial intelligence. *IEEE Rev. Biomed. Eng.* **17**, 80–97 (2024).
79. Alterovitz, G. et al. FHIR genomics: enabling standardization for precision medicine use cases. *NPJ Genom. Med.* **5**, 13 (2020).
80. Sheller, M. J. et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* **10**, 12598 (2020).
81. Sarma, K. V. et al. Federated learning improves site performance in multicenter deep learning without data sharing. *J. Am. Med. Inf. Assoc.* **28**, 1259–1264 (2021).
82. Warner, J., Jain, S. & Levy, M. Integrating cancer genomic data into electronic health records. *Genome Med.* **8**, (2016).
83. Alterovitz, G. et al. Smart on FHIR genomics: facilitating standardized clinico-genomic apps. *J. Am. Med. Inf. Assoc.* **22**, 1173–1178 (2015).
84. Warner, J. L. et al. Smart precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. *J. Am. Med. Inf. Assoc.* **23**, 701–710 (2016).
85. Bensmail, H. & Haoudi, A. Data mining in genomics and proteomics. *J. Biomed. Biotechnol.* **2005**, 63–64 (2005).
86. Wu, W. T. et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil. Med. Res.* **8**, 44 (2021).
87. Clough, E. & Barrett, T. The gene expression omnibus database. *Methods Mol. Biol.* **1418**, 93–110 (2016).
88. Wang, H., Wang, X., Xu, L., Zhang, J. & Cao, H. Prognostic significance of mycn related genes in pediatric neuroblastoma: a study based on target and geo datasets. *BMC Pediatr.* **20**, 314 (2020).
89. Chen, X. et al. Datamed—an open source discovery index for finding biomedical datasets. *J. Am. Med. Inf. Assoc.* **25**, 300–308 (2018).
90. Mahi, N. A., Najafabadi, M. F., Pilarczyk, M., Kouril, M. & Medvedovic, M. Grein: An interactive web platform for re-analyzing geo RNA-seq data. *Sci. Rep.* **9**, 7580 (2019).
91. Henry, V. J., Bandrowski, A. E., Pepin, A. S., Gonzalez, B. J. & Desfeux, A. Omictools: an informative directory for multi-omic data analysis. *Database.* **2014** (2014).
92. Torre, D. et al. Datasets2tools, repository and search engine for bioinformatics datasets, tools and canned analyses. *Sci. Data* **5**, 180023 (2018).
93. Murdock, D. R. et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J. Clin. Investig.* **131**, e141500 (2021).
94. Comitani, F. et al. Diagnostic classification of childhood cancer using multiscale transcriptomics. *Nat. Med.* **29**, 656–666 (2023).
95. Sturm, D. et al. Multiomic neuropathology improves diagnostic accuracy in pediatric neuro-oncology. *Nat. Med.* **29**, 917–926 (2023).
96. Chang, W. et al. Multidimensional clinomics for precision therapy of children and adolescent young adults with relapsed and refractory cancer: a report from the Center for Cancer Research. *Clin. Cancer Res.* **22**, 3810–3820 (2016).
97. Lunke, S. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* **29**, 1681–1691 (2023).
98. Armenteros, J. et al. Multi-omics analysis reveals drivers of loss of B-cell function after newly diagnosed autoimmune type 1 diabetes: an Innodia multicenter study. *Diabetes/Metab. Res. Rev.* **40** (2024).
99. Kohane, I. S. Finding a new balance between a genetics-first or phenotype-first approach to the study of disease. *Neuron* **109**, 2216–2219 (2021).
100. Ranola, J. M. O., Tsai, G. J. & Shirts, B. H. Exploring the effect of ascertainment bias on genetic studies that use clinical pedigrees. *Eur. J. Hum. Genet.* **27**, 1800–1807 (2019).
101. Wilczewski, C. M. et al. Genotype first: clinical genomics research through a reverse phenotyping approach. *Am. J. Hum. Genet.* **110**, 3–12 (2023).
102. Seltz, S. et al. Reverse phenotyping facilitates disease allele calling in exome sequencing of patients with Cakut. *Genet. Med.* **24**, 307–318 (2022).
103. Lyons, J. J. et al. Elevated basal serum tryptase identifies a multisystem disorder associated with increased Tpsab1 copy number. *Nat. Genet.* **48**, 1564–1569 (2016).
104. Rees, M. et al. Correlation of rare coding variants in the gene encoding human glucokinase regulatory protein with phenotypic, cellular, and kinetic outcomes. *J. Clin. Investig.* **122**, 205–217 (2012).
105. Mitra, I. et al. Reverse pathway genetic approach identifies epistasis in autism spectrum disorders. *PLoS Genet.* **13**, e1006516 (2017).
106. Landini, S. et al. Reverse phenotyping after whole-exome sequencing in steroid-resistant nephrotic syndrome. *Clin. J. Am. Soc. Nephrol.* **15**, 89–100 (2020).
107. Wilkinson, M. D. et al. The fair guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
108. Liu, X., Luo, X., Jiang, C. & Zhao, H. Difficulties and challenges in the development of precision medicine. *Clin. Genet.* **95**, 569–574 (2019).
109. Petzschner, F. H. Practical challenges for precision medicine. *Science* **383**, 149–150 (2024).
110. Pixberg, C. et al. Reimbursement in the context of precision oncology approaches in metastatic breast cancer: challenges and experiences. *Breast Care* **19**, 10–17 (2024).
111. Miller, M. B., Watts, M. L. & Samuel, L. FDA's proposed rule for the regulation of laboratory-developed tests. *J. Clin. Microbiol.* **62**, e0148823 (2024).
112. Genzen, J. R. Regulation of laboratory-developed tests. *Am. J. Clin. Pathol.* **152**, 122–131 (2019).
113. Baird, A. M. et al. How can we deliver on the promise of precision medicine in oncology and beyond? A practical roadmap for action. *Health Sci. Rep.* **6**, e1349 (2023).
114. Precision Medicine: An Action Plan for California. *Executives for Health Innovation* (2019).
115. Stenzinger, A. et al. Implementation of precision medicine in healthcare—a European Perspective. *J. Intern Med* **294**, 437–454 (2023).
116. What Is Value-Based Healthcare? *Catalyst Carryover* **3** (2017).
117. Xu, Y., Lai, K. K. & Leung, W. K. J. A consensus-based decision model for assessing the health systems. *PLoS One* **15**, e0237892 (2020).

## AUTHOR CONTRIBUTIONS

Each author has met the *Pediatric Research* authorship requirements. S.M.—conceptualization, design, interpretation of data, drafting the article, and revision; S.L.—interpretation of data, critical review, revision, and final approval; M.P.—conceptualization, design, data acquisition, data interpretation, data analysis, revision, and final approval. All authors agree to be accountable for all aspects of the work.

## FUNDING

Mohan Pammi is funded by the following extramural source: NIH 1R01HD112886.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Srinivasan Mani.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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 Author(s) 2025