Circulation: Genomic and Precision Medicine

REVIEW

Data Interoperability and Harmonization in Cardiovascular Genomic and Precision Medicine

C. Anwar A. Chahal, MBChB, MRCP, PhD; Fares Alahdab, MD, MSc; Babken Asatryan, MD, PhD; Daniel Addison, MD; Nay Aung, MBBS, MRCP, PhD; Mina K. Chung, MD; Spiros Denaxas, PhD; Jessilyn Dunn, PhD; Jennifer L. Hall, PhD; Nathalie Pamir, PhD; David J. Slotwiner, MD, PhD; Jose D. Vargas, MD, PhD; Antonis A. Armoundas, PhD

ABSTRACT: Despite advances in cardiovascular care and improved outcomes, fragmented healthcare systems, nonequitable access to health care, and nonuniform and unbiased collection and access to healthcare data have exacerbated disparities in healthcare provision and further delayed the technological-enabled implementation of precision medicine. Precision medicine relies on a foundation of accurate and valid omics and phenomics that can be harnessed at scale from electronic health records. Big data approaches in noncardiovascular healthcare domains have helped improve efficiency and expedite the development of novel therapeutics; therefore, applying such an approach to cardiovascular precision medicine is an opportunity to further advance the field. Several endeavors, including the American Heart Association Precision Medicine platform and public-private partnerships (such as BigData@Heart in Europe), as well as cloud-based platforms, such as Terra used for the National Institutes of Health All of Us, are attempting to temporally and ontologically harmonize data. This state-of-the-art review summarizes best practices used in cardiovascular genomic and precision medicine and provides recommendations for systems' requirements that could enhance and accelerate the integration of these platforms.

Key Words: big data ■ electronic health records ■ natural language processing ■ phenomics ■ translational research, biomedical

Cardiovascular genomic and precision medicine is going through a revolution with the availability of cutting-edge clinical and research technologies being introduced at a rapid pace. Furthermore, institutions and governments are funding and establishing biobanks and biorepositories to facilitate precision medicine. The cornerstone to making sense of genomic and precision technologies, in particular, omics, imaging, and monitoring data (telemetry, ambulatory ECG, smartphone, and wearable), is the electronic health records (EHRs).[1]

As access to health data increases exponentially, the adoption of regulatory frameworks and standards is critical for translating these health data to improve patient outcomes. In 2014, the US National Institutes of Health (NIH) introduced its Genomic Data Sharing Policy that was intended to encourage the broad and responsible sharing of genomic research data.[2]

While considerable and commendable efforts are underway to improve EHR interoperability, there is little progress tailored to cardiovascular genomic and precision medicine. Therefore, it is vital that cardiovascular-specific challenges are identified, and solutions are developed to harness the power of these technologies, in order to facilitate not only first-class clinical care, but also synergize research and outcomes' assessment.

The aim of this state-of-the-art review is to summarize the best practices used in cardiovascular genomic and precision medicine research, identify challenges and barriers and provide recommendations for systems requirements that could enhance and accelerate the integration

Correspondence to: C. Anwar A. Chahal, MBChB, MRCP, PhD, Center for Inherited Cardiovascular Diseases, WellSpan Health, 157 N Reading Rd, Ephrata, Lancaster, PA 17522, Email cchahal@wellspan.org; or Antonis A. Armoundas, PhD, Cardiovascular Research Center, Massachusetts General Hospitals, 149 13th St, Charlestown, MA 02129, Email armoundas.antonis@mgh.harvard.edu

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/CIRCGEN.124.004624.

For Sources of Funding and Disclosures, see pages 334 and 335.

Nonstandard Abbreviations and Acronyms

AI	artificial intelligence
CDM	common data model
DICOM	Digital Imaging and Communications
HER	electronic health record
FHIR	Fast Healthcare Interoperability Resources
ICD	*International Classification of Diseases*
NIH	National Institutes of Health
OMOP	Observational Medical Outcomes Partnership
SNOMED CT	Systematized Nomenclature of Medicine Clinical Terms

of data (Figure 1), and enhance their uses in translational research and applicability in real-world settings.

THE FOUNDATION

Accurate phenotypic and outcome data are vital to harness the power of precision medicine.[3] Connecting multiomics data is expected to become the cornerstone in precision diagnosis, whether this is about known diseases involving the identification of novel genetic variants, the illumination of new biological pathways, or in pursuing the understanding of unknown diseases (diagnostic odysseys). This section summarizes the role and challenges of the EHR in harnessing its potential.[4] Used correctly, complementing multiomics with the EHR can redefine current nosology and ontology.

STRUCTURED VERSUS FREE TEXT: USE OF NATURAL LANGUAGE PROCESSING IN CODING AND ONTOLOGY

The EHRs' complexity, attributed to the heterogeneity of clinical patient data generated and collected by disparate systems, presents challenges for data entry, management, and use (Figure 2).[5] EHRs exist in both structured and unstructured formats. Structured EHR data are recorded using clinical terminologies, ontologies, and coding systems (eg, *International Classification of Diseases* [*ICD*] and current procedural terminology) and enable quick analysis and systematic comparison across healthcare organizations, including meta-analyses. Unstructured data include free text (eg, clinical notes) and provide detailed narratives that would not otherwise be possible to capture using ontological terms. While structured data are crucial for computational tasks and align with specific standards, free text often contains insights that may be lost in structured data. Natural language processing bridges this gap by encoding free text, for example, translating human language into machine-readable formats, for example, by assigning accurate ontological concepts.[6,7] This type of language embedding, combined with ontology structures, allows for better alignment and translation of medical concepts across different languages and standards.[8]

CODING SYSTEMS AND MEDICAL ONTOLOGIES

Since the passage of the Health Information Technology for Clinical and Economic Health Act in 2009 aiming to adopt the use of EHRs,[9] various coding systems have evolved and now serve as the backbone of health informatics, each with their unique complexity and purpose. The World Health Organization adopted and began developing the *ICD* in 1948 for recording, reporting, and analyses of mortality and morbidity data.[10,11] *ICD*, *Ninth Revision*[12] was developed in 1983 and remained in use in the United States until 2015 (as *ICD*, *Ninth Revision*, Clinical Modification, a specific version that was developed for use in the US healthcare system *ICD*, *Tenth Revision*, encompasses around 68 000 codes; this refined revision enabled a more precise coding of diseases; Table S1).[13,14] Nevertheless, certain limitations remain with the use of *ICD* coding such as errors in coding, variability in coder expertise, and the occasional imprecision in the coding system itself. Among the strategies used to mitigate these challenges have been the frequentist approach in using multiple coding (diagnosis defined by being coded in multiple settings) and the temporal consistency (confirming diagnosis coded similarly over time).

In the United Kingdom, the analogous procedural coding system is the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures.[15] Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) provides an international and comprehensive medical ontology that allows the precise recording of information by healthcare professionals and the exchange of health information between healthcare providers.[16,17] SNOMED CT offers a multiaxial (compositional) approach incorporating: (1) concepts; (2) descriptions; and (3) relationships, which allows for > 500 000 possible concepts and >1 million clinically meaningful relationships between those concepts, overcoming many long-standing coding limitations. SNOMED CT's advantages over other coding standards may allow for better specificity of harmonization and is a standard for interoperability required by the United States. An adjacent standard that is well-structured and has similarly high interoperability is Logical Observation Identifiers, Names and Codes, which is the international standard for laboratory and clinical test results.

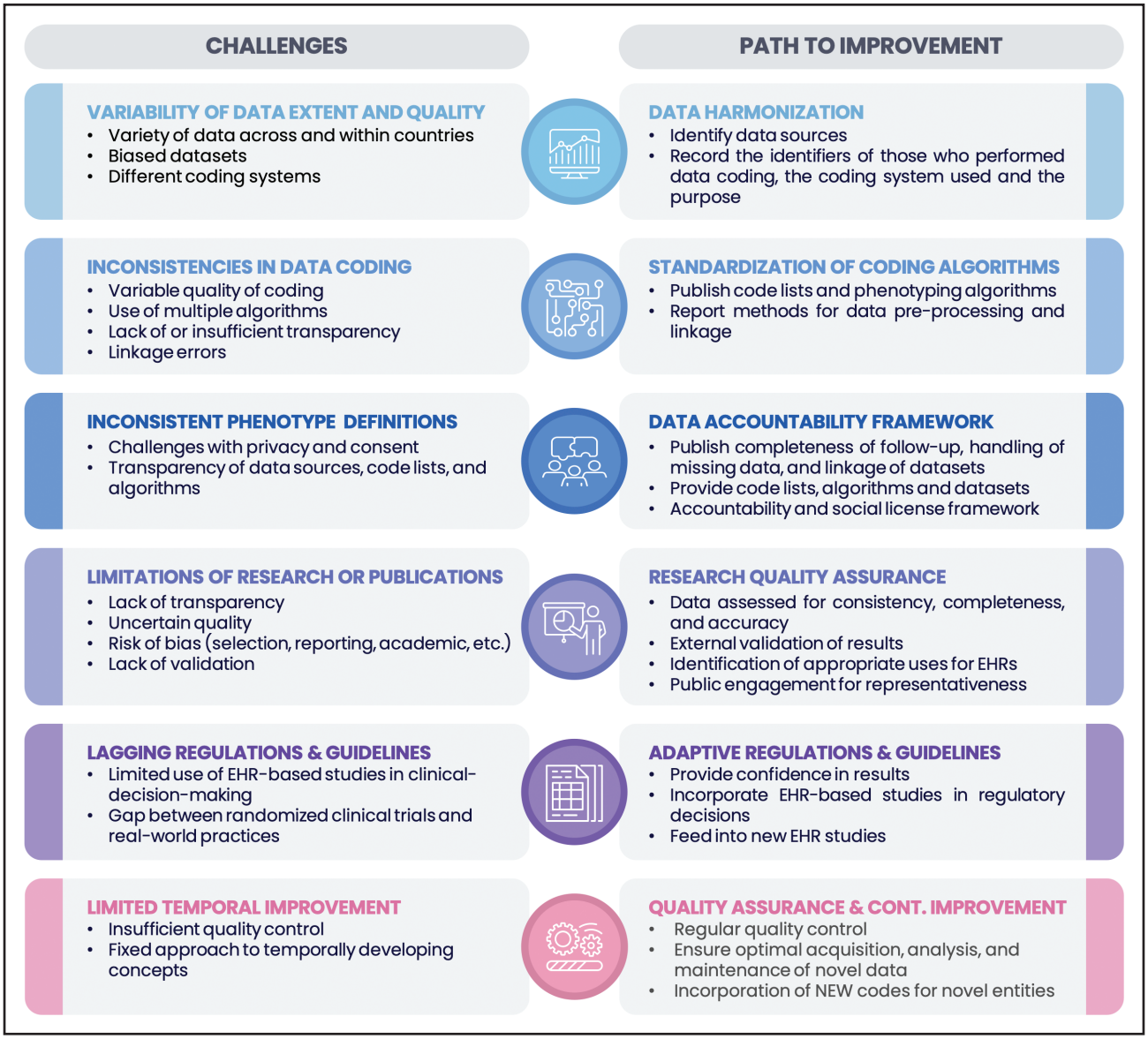

Figure 1. Summary of present challenges and possible solutions to improve data harmonization and interoperability for genomic and precision medicine.
EHR indicates electronic health record.

CHALLENGES

The availability of different coding standards, as well as the multitude of vendor-created proprietary EHR information systems, inevitably leads to wide gaps in how data are collated, stored, transferred, curated, and used. Furthermore, adopting subsequent updated coding systems, although helping to overcome prior coding limitations, is inherently difficult and costly. Not all health systems are equally able to adopt newer updates and better EHR systems, as digital maturity is a major component that influences transitions to complex information systems.[18] These lead to several data harmonizing challenges that jeopardize the integrity and reliability of data use. The purpose of coding, be it for billing or registry, often affects the quality and level of detail, and several studies have shown concerning inconsistencies that include a lack of or incorrectly coded data.[19] Furthermore, a lack of consensus on coding for diagnosis and procedures due to different coding practices, underlying incentives, or clinical pathways may lead to systematic biases in the data. This causes confusion and a lack of interoperability at a national and international level.[20,21] Finally, coding systems and medical ontologies may lack the nuance to capture complex medical realities, leading to imprecise representations and maintaining the need for the inclusion of natural language processing[22,23] (Figure 2).

Common data models (CDMs) are utilized when there is a need to share and exchange data for particular uses. CDMs can be used to specify structure, format,

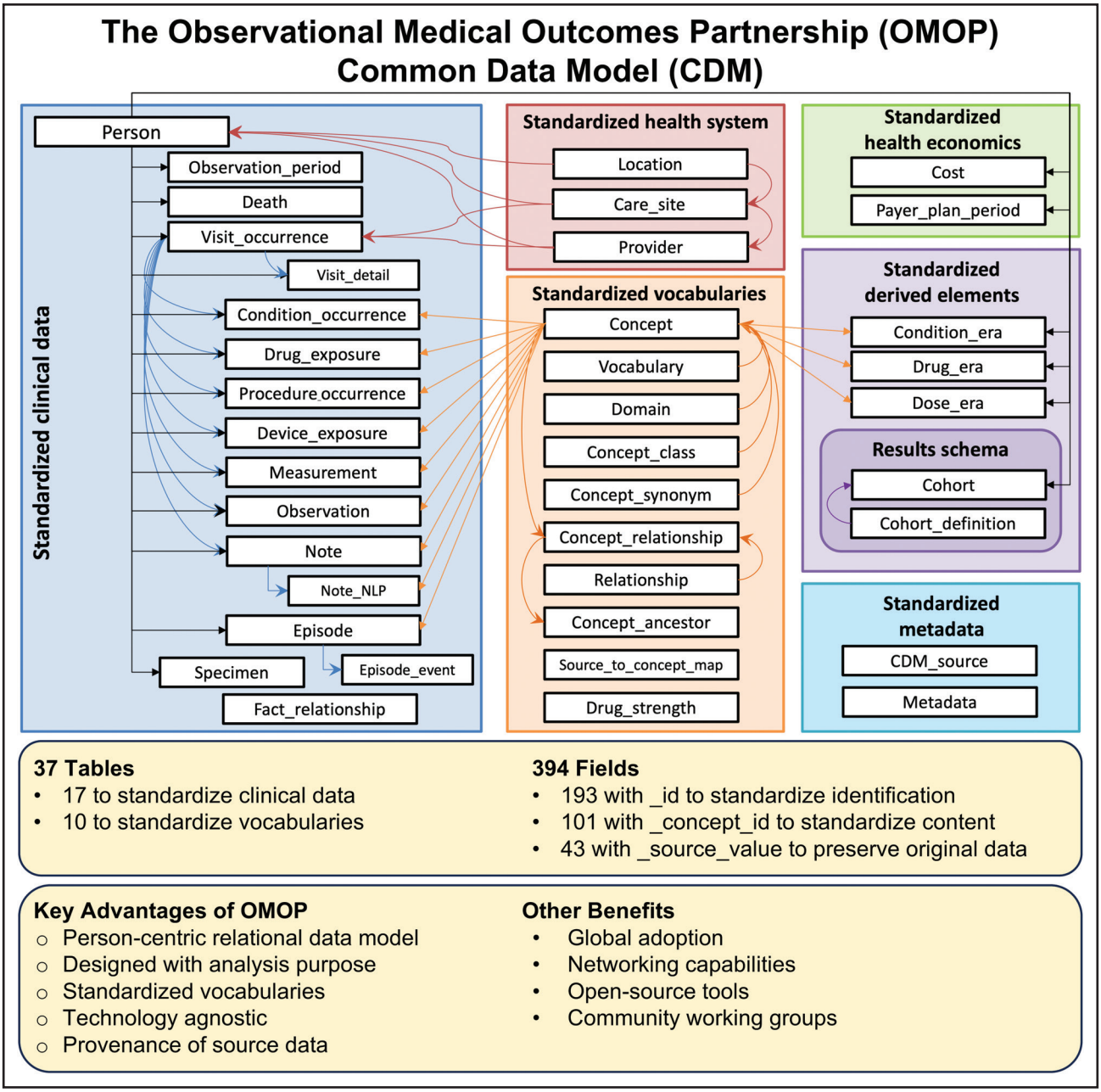

Figure 2. Current status of standardization for observational data.
The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is an open community data standard, designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence. A central component of the OMOP CDM is the Observational Health Data Sciences and Informatics (OHDSI) standardized vocabularies. The OHDSI vocabularies allow organization and standardization of medical terms to be used across the various clinical domains of the OMOP common data model and enable standardized analytics that leverage the knowledge base when constructing exposure and outcome phenotypes and other features within characterization, population-level effect estimation, and patient-level prediction studies.

and content and have formed the cornerstone of multicenter registries and data reporting in clinical research. Examples of CDMs include the Informatics for Integrating Biology and the Bedside, the Patient-Centered Clinical Research Network, the Observational Medical Outcomes Partnership (OMOP), the Clinical Data Interchange Standards Consortium Study Data Tabulation Model, and Sentinel. These aim to format, clean, harmonize, and standardize data mostly for research and discovery purposes (Figure 2),[24] across countries.[25] Head-to-head comparisons of CDM employing longitudinal EHR data consisted of 11 criteria, covering 6 categories, content coverage (completeness), integrity, flexibility, ease of querying (simplicity), standards compatibility (integration), and ease and implementation, have highlighted these challenges.[26] The OMOP CDM also includes a natural language processing element focused on medical notes and serves as a great

example of an effective standardized data model within health informatics.

Closed-world assumptions in formal logic refer to an accurate statement known to be true; conversely, what is not known to be true is untrue. With an open-world assumption, what is not known to be true is simply unknown. Finally, in leveraging CDMs, a differential violation can be made with a closed-world assumption model.

However, appropriate funding underpins many of these initiatives, and because funding tends to favor historically larger and better-funded institutions, populations not served by these institutions are inadvertently excluded and remain poorly represented.

STEPS IN DATA PROCESSING

Data Collection

When individuals interact with the healthcare system, a wealth of data and metadata is generated,[27] including anything related to care, such as diagnoses, physical examination findings, drug administrations and prescriptions, laboratory measurements, procedures, and surgical interventions. Metadata include information about the method of admission, the discharge destination, spatiotemporal information (eg, date and time of admission and hospital identifiers), and other details around health care (eg, the specialty of the treating physician), which eventually become real-world evidence.[28]

Data Storage and Processing

Data are often stored in EHRs and health information management systems. EHR systems are specialized, and often proprietary, software applications that are used to store structured and unstructured information and use healthcare information exchange standards, such as Health Level Seven Fast Healthcare Interoperability Resources (FHIR),[29] to structure data and enable its portability. Images are stored often on Digital Imaging and Communications (DICOM) servers that are specialized servers designed to store, retrieve, and transmit medical images and related data in the DICOM format.[30] DICOM servers are used in healthcare settings, such as hospitals, clinics, and imaging centers, to manage medical imaging data produced by devices such as X-ray machines, magnetic resonance imaging scanners, and computerized tomography scanners.

Data Linkage

Disparate data sources are often linked using patient identifiers through a process known as data linkage and normalization, for example, linking data between primary care EHR and hospital admissions' data.[31] Data linkage is often performed using a unique healthcare identifier, such as the National Health Service Number in the United Kingdom, in a deterministic fashion. In the United States, such a common identifier in many cases does not exist,[32,33] and the medical record number is often used as an identifier unique to the patient and the health system; therefore, one unique patient may have multiple medical record numbers if they visit multiple health systems. When a unique healthcare identifier does not exist or does not coincide with another data set, which often occurs when linking data between health and nonhealth data (eg, educational records), the linkage is performed using nonunique information, such as name, postcode, and date of birth, using probabilistic approaches. Both deterministic and probabilistic approaches are prone to errors in the data (such as missing data and incorrect data) or nonuniqueness across patient information. Careful evaluation of the linked data is required to ensure that no biases or errors are introduced during the linkage process.[34]

Federated Studies

Federated studies in health care[35] refer to research initiatives that involve the collaborative analysis of data from multiple institutions or organizations while keeping the data localized and secure within each respective site. Unlike traditional centralized meta-studies where data are pooled into a single database, federated studies allow researchers to perform meta-analyses across disparate data sets without physically combining them. Instead, queries and analyses are distributed across different sites, ensuring data privacy and security. This approach is particularly valuable in healthcare research, enabling large-scale studies involving diverse patient populations and data sets without compromising individual privacy or violating data-sharing regulations.

Data Quality Control: Foundation of Good Scientific Data Management and Stewardship

Data Management

Data management plans are now required by the NIH and the National Science Foundation. A data management plan includes types of data, metadata, data and metadata standards, related tools, software, code, data preservation, access, timelines, distribution, or reuse considerations and oversight.

Adoption and employment of data standards, before data collection, are expected to facilitate interoperability and extend the reusability of data.[21] Table 1 provides resources for data and metadata standards. Table S2 provides examples of data platforms and repositories that provide data to end users with accompanying data dictionaries and documentation that are expected to improve data handling efficiency and usability. Quality

Table 1. Data Types, File Types, and Examples of Where to Find Data and Metadata Standards; Adopted and Revised From Repository (Meta) Data Standards Examples, Version 0.4, April 17, 2023, From NIDDK

Data type	File type	Data standards	Metadata standards
Clinical	Flat tabular files (eg, CSV and TSV)	Clinical Data Interchange Standards Consortium *ICD-10*-CM CTAE SNOMED CT Phenotype and Trait Ontology NIH Common Data Elements Repository	CDISC ADaM
Imaging	DICOM, NiFTi, mp4, PNG, and TIFF	DICOM NIH Common Data Elements Repository	Photo Metadata IPTC
Genomics	BAM, BED, FASTQ, and VCF	HUGO Gene Nomenclature Committee NIH Common Data Elements Repository	Minimum information about any sequence[36]
Transcriptomics	BAM and FASTQ	HUGO Gene Nomenclature Committee	Minimal information about a high-throughput sequencing experiment[37]
Metabolomics	imzML and mzTab	mzTab for Metabolomics[38] Nuclear Magnetic Resonance Markup Language	Core information for metabolomics reporting[38]
Proteomics	mzidentML, mzTab, RAW, and TSV	Mz Markup Language[39]	Minimal information about a proteomics experiment[40]

ADaM indicates Analysis Data Model; BAM, binary alignment/map; BED, browser extensible data; CDISC, Clinical Data Interchange Standards Consortium; CSV, comma-separated values; CTAE, Common Terminology Criteria for Adverse Events; DICOM, Digital Imaging and Communications in Medicine; FASTQ, a text-based format that stores both nucleotide sequences and quality scores, standard for raw sequencing reads; Format Mz, mass-to-change ratio; HUGO, Human Genome Organisation; *ICD-10*-CM, *International Classification of Diseases, Tenth Revision*, Clinical Modification; IPTC, Photo Metadata Standard; MP4, MPEG-4 Part 4; Mz, mass-to-charge ratio; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; NiFTi, Neuroimaging Informatics Technology Initiative; NIH, National Institutes of Health; PNG, portable network graphics; RAW, a generic term for unprocessed data from imaging devices; SNOMED CT, Systematized Nomenclature of Medicine Clinical Terms; TIFF, tagged image file format; TSV, tab-separated values; and VCF, variant call format.

control procedures and quality evaluation metrics for each data set are not normally provided in data documentation, yet such information is important for end users. Therefore, we propose that a reporting criterion should include quality evaluation metrics.

Data Collection, Preservation, Backup, and Security

Table S3 provides examples of open-source workstreams that are expected to improve interoperability and usability. As a case study, the American Heart Association Precision Medicine Platform adopted an open-source workstream for data access and data governance made available by the Broad Institute, Data Use Oversight System.[41] The Data Use Oversight System brings together researchers submitting and requesting data, as well as data access committees and institutional review boards.[41]

However, a major challenge appears to be how to manage, store, and back up large volumes of data. Data preservation ensures that data remain accessible and usable for future researchers. In that case, data backups and redundancy prevent data loss due to hardware failures or other unforeseen events. There are many different paths for data preservation and backup, and considerations for the different setups include data size, access speed, and budget constraints. On-premises solutions include Network Attached Storage, and off-premises options include cloud services (Figure 3). A combination of approaches may be chosen to optimize different types of data access against the cost of storage, and data lifecycle management is an approach that can be used to plan data lifecycles from data entry to either archival or destruction. Recently, the NIH released a policy (NOT-OD-21-013), which requires researchers generating NIH-funded data to create a strategy and budget for data management and sharing. Similarly, the National Human Genome Research Institute Genomic Data Science Analysis, Visualization, and Informatics Lab-Space is a secure, cloud-based environment where researchers are able to store, share, and analyze key unrestricted- and controlled-access genomic data sets and associated phenotypic data or metadata, particularly those generated with National Human Genome Research Institute funding or support (Analysis, Visualization, and Informatics Lab-Space, NOT-HG-19-024). The Analysis, Visualization, and Informatics Lab-Space is designed to minimize local analyses because of unsustainable storage needs, security concerns, and large volumes of data traditionally downloaded from a data warehouse by providing a cloud environment for the analysis of large genomic and other omic-related data sets.[42]

Data Storage and Organization

Hierarchical electronic file storage systems, often using a tree-based directory structure, underpin electronic data organization. The file tree stores data files in a logical structure that mirrors the workflow, which can either be the workflow, which generated the data, or the workflow by which the generated data will be used for downstream analyses. Such structured data storage is facilitated by strong file naming conventions, which often also provide metadata. Aside from the traditional tree-based directory systems, other database structures

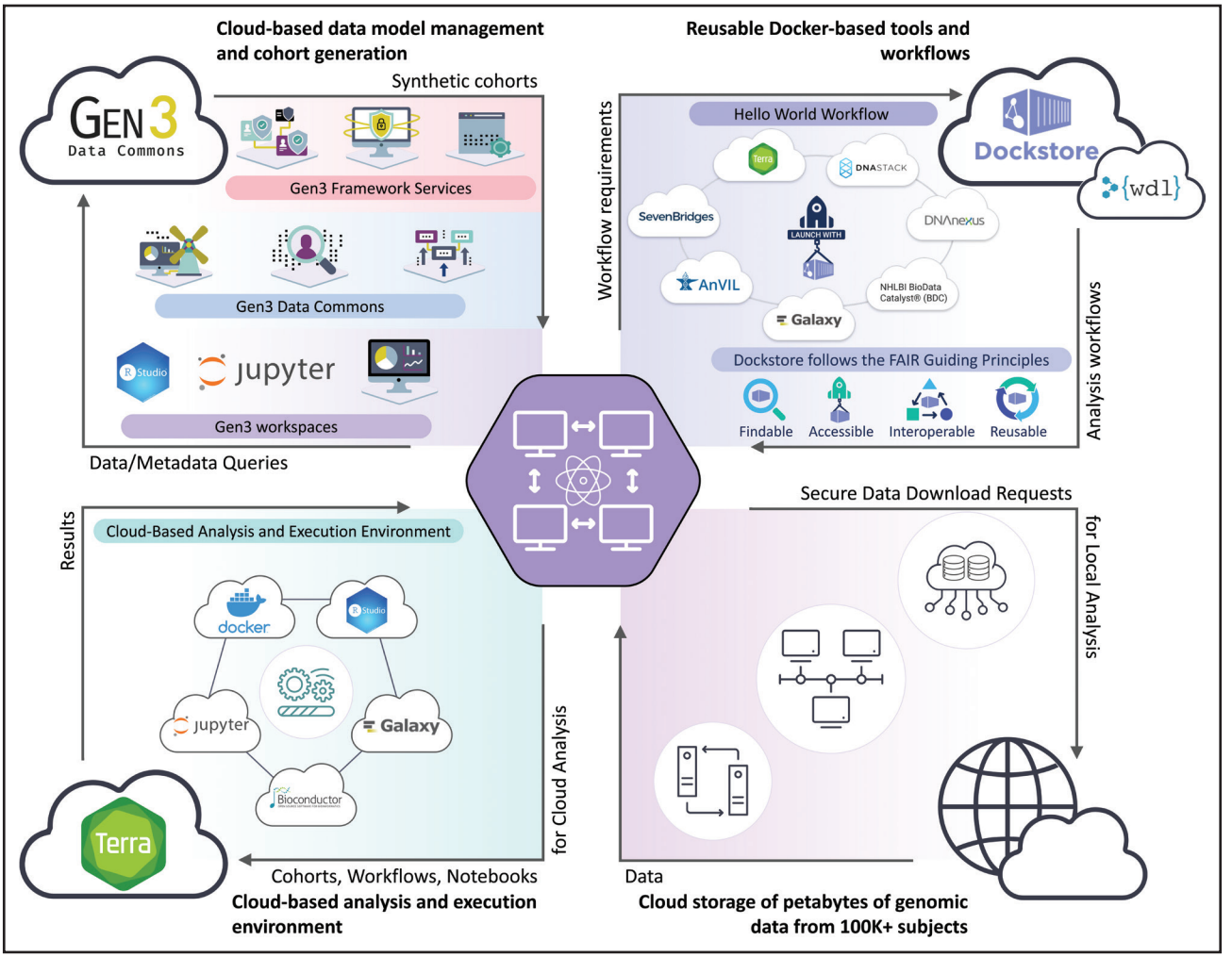

Figure 3. Overview of Gen3 framework services for cloud-based data analysis and execution.
This figure highlights the comprehensive ecosystem supporting cloud-based data analysis and execution. The Gen3 framework, developed to support large-scale biomedical data analysis, leverages cloud infrastructure to provide robust, scalable, and secure services. Components include the following. Synthetic cohorts: the creation and management of synthetic data sets for various research purposes, enabling analysis without compromising real data privacy. Data/metadata queries: tools and services that facilitate efficient querying of data and metadata to support research workflows. Workflow requirements and analysis workflows: specifications and execution pathways necessary for conducting data analyses within the Gen3 environment. Gen3 framework services and data commons: core services and shared resources within the Gen3 ecosystem that supports data storage, management, and accessibility. Gen3 workspaces and hello world workflow: user-friendly interfaces and example workflows provided to help users get started with the Gen3 environment. Findable, accessible, interoperable, and reusable (FAIR) principle implementation: emphasis on ensuring data is FAIR, following the FAIR guiding principles as promoted by initiatives such as FORCE11. Cloud-based analysis and execution environment: integration of secure cloud environments for data analysis, providing scalable resources for handling large data sets. Reusable docker-based tools and workflows: the use of docker containers to encapsulate tools and workflows, promoting reusability and reproducibility of research. Cloud storage and secure data download requests: mechanisms for storing large volumes of data securely and facilitating controlled access for download and analysis. Cohorts, workflows, and notebooks: support for generating and managing cohorts, executing workflows, and using interactive notebooks for data analysis. AnVIL indicates Analysis, Visualization, and Informatics Lab-Space; IRB, institutional review board; OMOP, Observational Medical Outcomes Partnership; PCORnet, patient-centered outcomes research network; and SDV, synthetic data vault.

exist, which are particularly useful for large data, including relational database structures, such as Structured Query Language and network databases, where child records can be linked to multiple parent records.

Data Accessibility and Sharing

EHR originated mainly for making information accurate, current, complete, and more readily available at the point-of-care, coding, and billing purposes, but they were later adapted for improving clinical workflows, enabling healthcare quality innovation, and supporting research endeavors. As such, data sharing is a necessity for achieving equitable healthcare advances. Health data can be deidentified by removing Health Insurance Portability and Accountability Act identifiers,[43] but this may result in the removal of key information needed to optimize model performance. An alternative solution is to maintain

identifiable data but work in secure enclaves. A third solution is to use federated practices to pool data used in analyses or model building without physically bringing the data together. While valuable data often reside in disparate silos across institutions, and sometimes even within institutions, several concerns and policies prevent effective use of these data. Some factors preventing data sharing involve concerns of reidentification of individual patients and risks to privacy, while others involve a lack of motivation and trust by the stakeholders. Nevertheless, data-sharing carries more benefits than risks and can be of paramount importance, as evidenced by the COVID-19 pandemic.[44,45]

Ethics and Compliance

Regulation and management of scientific data involve a range of ethical and data privacy considerations at each stage (ie, at collection, storage, access, and sharing). Research involving human subjects may require review by an institutional review board and informed consent. Data should be deidentified by removing any personal identifying information that could reidentify individuals. Therefore, investigators must be familiar with international and local regulations, such as the General Data Protection Regulation in the European Union or the Health Insurance Portability and Accountability Act in the United States.

Analysis and Reproducibility

Data collection methods should be specified clearly in the study design to ensure equity, integrity, reliability, transparency, traceability, and reproducibility of the research findings.[41,46]

Collaboration with external organizations and experts is needed to validate and benchmark data quality control processes and metrics. Participation in data quality initiatives and consortia provides an opportunity to share and harmonize best practices.

Quality Assurance and Continuous Improvement

Prioritizing an understanding of both the importance of and methodologies in acquiring, analyzing, and maintaining scientific data, is an important principle of medical research. Individuals could be trained in standardized data collection protocols, including training to minimize data entry errors such as double-data entry and validation procedures. Audit trails that track changes and updates to the data provide transparency and help identify any unauthorized modifications.[47] Protocols for data cleaning can address missing data, outliers, and inconsistencies, but these data cleaning protocols must be documented clearly. Quality control checks at various stages of the research may also be implemented to look for anomalies and discrepancies.

Continuous improvement in the methodology of data collection and management should follow advances in technology, evolving best practices, and feedback from interested stakeholders. Periodic reviews of best practices ensure that the most current, relevant, and effective tools and methodologies are being used.

Standards That Exist or Are Being Developed

Imaging

In the domain of cardiovascular imaging, an established array of data standards exists, encompassing multiple facets, such as data storage format, communication protocols, data security, and standardized reporting.[21] These standards are integral to facilitating interoperability and consistency, as well as fostering efficiency and resilience in the management of vast data generated by rapidly advancing imaging techniques.

Data Storage Formats

The most universally recognized image storage standard is stipulated by the DICOM initiative.[48] All major imaging device vendors adhere to the DICOM format, which encapsulates both the metadata pertaining to patient and scan details and the image data (pixel information). DICOM supports a broad range of imaging modalities, including ultrasound, magnetic resonance imaging, and computerized tomography. An alternative image format, Neuroimaging Informatics Technology Initiative, has been developed by the neuroimaging community at the NIH,[49] as a preferred file format within the machine learning community for its compactness, ease of postprocessing, and minimal metadata, which facilitates privacy preservation.

Communication Protocols and Security Standards

Health Level Seven, augmented by the FHIR standard, establishes a broad spectrum of international messaging and communication protocols. These standards are pivotal for the secure transmission, exchange, and retrieval of electronic health information, thereby enhancing interoperability across the radiology IT ecosystem, including picture archiving and communication systems, radiology information systems, and FHIR. FHIR, a highlight of Health Level Seven, introduces a Web-based, application programming interface–driven framework for health data exchange, characterized by its flexible, resource-oriented architecture that is gaining increasing endorsement from key regulatory agencies. These agencies, including the Centers for Medicare & Medicaid Services and the Office of the National Coordinator for Health Information Technology in the United States, are actively encouraging the adoption of FHIR to streamline healthcare interoperability.[21] The Integrating the Healthcare Enterprise initiative[50] further supports the integration of Health Level Seven and FHIR with various healthcare IT assets. Compliance with stringent

security and privacy regulations, such as the Health Insurance Portability and Accountability Act[51] and the General Data Protection Regulation,[52] is integral to these standards, including FHIR. There are specific security challenges pertaining to picture archiving and communication systems and DICOM standards, with vulnerabilities ranging from malware or ransomware and unauthorized access to malicious manipulation of images (injection attacks). DICOM working group 14 coordinates and produces recommendations to enhance DICOM security.[53] The US National Institute of Standards and Technology has also published the best practice guidance to safeguard medical images from cybersecurity threats.[54,55]

Standardized Reporting

Within the realm of cardiovascular imaging, international societies have published modality-specific acquisition and reporting guidelines. These include the Coronary Artery Disease Reporting and Data System[36] for coronary computerized tomography angiography, cardiac magnetic resonance imaging reporting standards as laid out by the Society for Cardiovascular Magnetic Resonance,[37] and the transthoracic echocardiography data set recommended by the American Society of Echocardiography[38] and the British Society of Echocardiography.[51]

Electronic Health Records

Artificial intelligence (AI) research using information from EHR is often challenging because many disparate information technology systems and databases are used to collect, curate, and store the data and do not follow common standards; thus, data are highly heterogeneous.[39] Therefore, there have been 5 proposed themes of EHR data quality: (1) completeness (the presence of data in the EHR); (2) correctness (the truthfulness of data in the EHR); (3) concordance (the agreement between elements within the EHR, between EHR sources, compared with other data sources); (4) plausibility (the extent to which EHR data make sense in a larger medical context); and (5) accuracy (the accuracy of the EHR data for the time at which it was recorded and how up-to-date the data are).

Health information technology providers must meet 3 requirements for their interface to be certified: (1) it must meet certain technical programming standards that ensure interoperability; (2) it must be transparent; and (3) it must be procompetitive or promote efficient exchange, access, and use of health data.[40] Recently, the American Medical Association published Principles for Augmented Intelligence Development, Deployment, and Use.[56]

As previously mentioned, approaches that harmonize data to a common semantic structure, such as CDMs, can enable better-quality linkage of multiple data sources and provide a common platform in which analytical methods can be built and deployed. An example of such an approach is OMOP CDM of the Observational Health Data Sciences and Informatics.[57] The OMOP CDM consists of 23 tables that are organized into 4 top-level domains: (1) clinical, (2) derived elements, (3) health system, and (4) health economics. Clinical data tables hold core data on patient demographics, clinical events (eg, diagnoses, laboratory measurements, medication prescriptions, and surgical procedures), visit occurrences, and observation periods. The health system data tables provide information on healthcare providers associated with the healthcare events held in the clinical data types. The health economics data tables contain cost information and details on the enrollment of patients in health benefit plans. Individual data sets are transformed into the OMOP CDM through an extract, transform, and load process, which maps data fields in each data source to OMOP vocabulary concepts (Athena, https://athena.ohdsi.org/) and to the CDM schema. Once data are transformed, federated studies can be executed on every source containing data in the CDM. OMOP has been successfully used to transform EHR data, clinical registries, and administrative data sets and, furthermore, has been successfully used to harmonize and execute federated analyses on disparate healthcare data sets and provide important insights to policymakers during the pandemic.[57]

Data harmonization can also occur as part of large-scale research studies, that rely on data from health care to enable deeper phenotyping and longer follow-up of individuals. These platforms tend to extract, load, and harmonize data from several EHR sources and link it with research data collected from consented participants. Examples of these types of platforms include the American Heart Association Precision Medicine Platform, Analysis, Visualization, and Informatics LabSpace, the NIH All of Us,[58] the Million Veteran Program and Department of Veteran Affairs and in the UK Biobank, Our Future Health, and Genes and Health studies. The European Union is establishing a European Health Data Space to facilitate the exchange and sharing of health data (eg, health records, genomics, and registries) for purposes such as the delivery of primary care and the development of new treatments, medicines, medical devices, and services while ensuring that people have control of their own health data.[59] Health Data Research UK is an independent, not-for-profit organization of 22 research institutions in the United Kingdom, which enables access to EHR data for research on diseases and ways to prevent, treat, and cure them. Principles of participation have been defined in consultation with policymakers, the National Health Service, the industry, and the public.[60] The Human Colossus Foundation is a Swiss-based not-for-profit organization, which has developed the Dynamic Data Economy architecture and an Overlays Capture Architecture for data harmonization across data ecosystems.[61] The key

strength of this model is decentralized semantics, which describes the separation of semantic (definitional) and pragmatic (contextual) tasks into task-specific objects that, when combined, provide a digital representation of a complex object. The use of the ontology-agnostic Overlays Capture Architecture allows harmonization and interoperability between data models and data representation formats, with a roadmap to resolve privacy-compliant data sharing. The particular strength of this approach lies in separating capture base and overlays, enabling harmonization of non-English language text or within nations with multiple official languages.

OMICS

Omics include, but are not limited to, genomics, epigenomics, transcriptomics, proteomics, metabolomics, nutrigenomics, and microbiomics (Figure 4). The concept of using omic technologies in precision medicine has gained significant attention with the completion of the Human Genome Project in 2003. Today, the term expressed genome is sometimes used to combine epigenetics, RNA transcripts, proteins, and metabolites.[62] The first successful clinical application used genomics to identify cells with KIT and the *PDGFRA* gene mutation, responsible for gastrointestinal stromal tumors, to target them with the use of imatinib mesylate (also known as Gleevec).[63] Since then, large clinical initiatives, such as, but not limited to, National Cancer Institute Molecular Analysis for Therapy Choice (MATCH) trial,[64] Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY 2) trial,[65] MyPathway trial,[66] and National Health Service England's 100 000 Genomes Project,[67] have been pursued to establish and develop a novel healthcare approach under the umbrella of precision medicine.

However, despite >20 years since the first application of omic technologies in precision medicine, we are still far from a state of standard personalized health care. Processing such swathes of generated data is highly dependent on systems network biology to synthesize and curate data and to evaluate the prediction, progression, and outcome

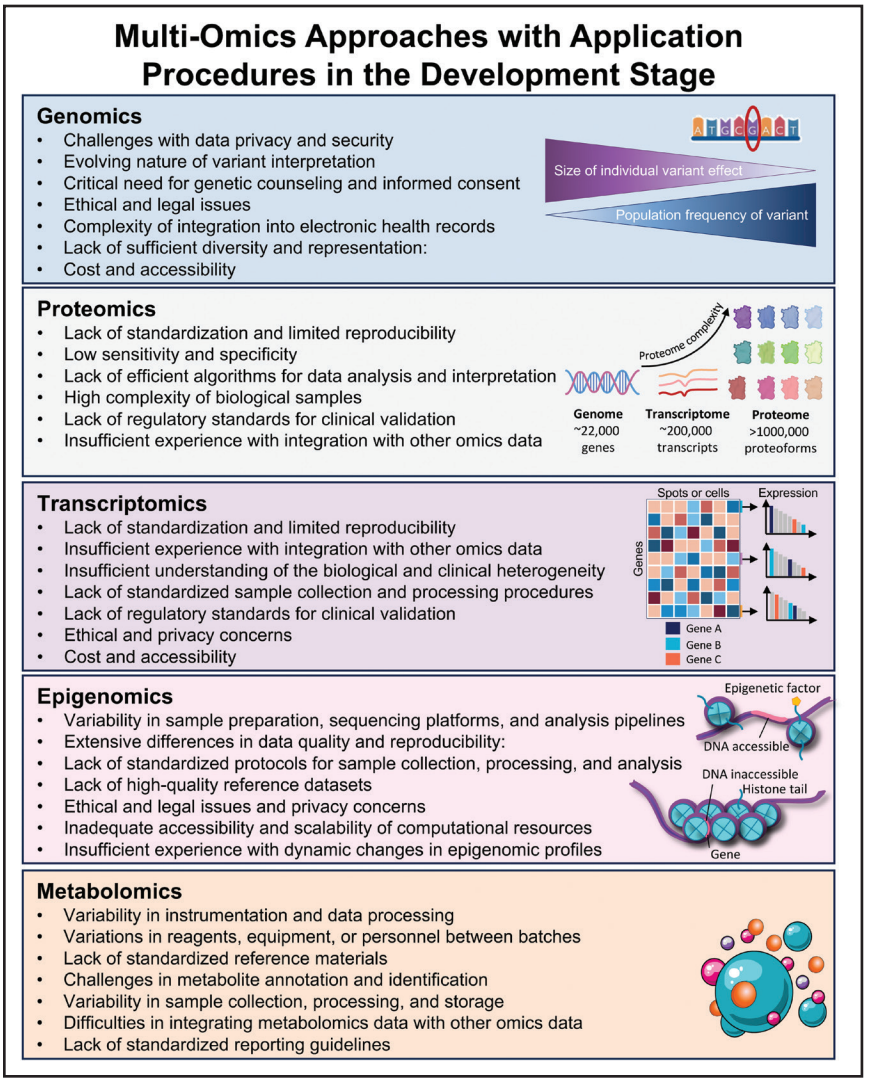

Figure 4. Multiomic approaches with application procedures in the development stages. Different multiomics approaches that are incorporated in healthcare delivery at various extents. Note that the challenges for data acquisition, analysis, harmonization, and reporting of standards and applicability differ between different omics methods.

of disease.[68] In addition, network systems aid with the collation, integration, and prioritization algorithms to guide the cartography and decoding of omic landscapes.

While we embrace the many advantages of the approach (Table S4), we still face many hurdles (Table 2) that require disease community-based solutions. Collaborative efforts between researchers, clinicians, policymakers, and the broader healthcare community to ensure continued advancements in technology, data sharing, and regulatory frameworks will play a crucial role in overcoming these hurdles and in the future of precision medicine.[47]

As the omic-based precision medicine application expands worldwide, one of the hurdles that require the big data communities' immediate attention is the challenge of data harmonization. Establishing principles of harmonization, through standardized protocols and computational tools, to data sets from all omic technologies and diseases will be critical in ensuring reliability and reproducibility of the analysis and interpretation process. Such principles are summarized in Table S5.

Genomics and Transcriptomics

Standards for whole genome and whole exome next-generation sequencing, and primary and secondary bioinformatic analyses are well established. Determining variant calls is also reliable. However, the major challenges lie with tertiary bioinformatics and variant annotation, and classification, into the 5 American Council of Medical Genetics and Genomics classes.[69] While there are attempts to evaluate evidence to classify genetic variants, there is a wide variation with reports and often conflicting data within ClinVar. Reports of samples processed by 2 different laboratories can vary widely, which is not unsurprising as all the evidence categories set out by the American Council of Medical Genetics and Genomics cannot be met. As new evidence arises, disease variants can be reclassified, with increased attention to gene-disease mechanisms.[70] From a precision medicine perspective, efforts ensuring genomic data as part of the EHR are being implemented. For example, Epic offers a Precision Medicine module, which allows not only reports but also key structured reporting, including type of gene test (panel or Whole Exome Sequencing/Whole Genome Sequencing), wild-type alleles, and mutations, with their class. There are challenges in updating these data, as new assertions of variants are updated in databases, such as the ClinVar.

Proteomics

Over the years, as proteomics integrated into the precision medicine toolbox, several standards and guidelines have been established to ensure data

Table 2. Challenges With the Use of Omic Technologies in Precision Medicine

Domain	Issues	Challenges
Data harmonization, integration, and interpretation	Harmonizing data from different sources, and integrating and interpreting vast amounts of data in a consistent, reliable, and meaningful way.	Poor integration along with incomplete or inaccurate interpretation can lead to incorrect clinical decisions.
Standardization and quality control	Standardizing protocols for data generation and ensuring data quality and reproducibility across different laboratories and platforms are essential for reliable results.	Inconsistencies in data quality can lead to erroneous conclusions and hinder cross-study comparisons.
Ethical and privacy concerns	The generation and storage of sensitive genetic information require privacy and ethical handling of the data.	Mishandling of genetic data can lead to breaches of privacy and potential discrimination.
Cost and accessibility	Elevated costs may limit the accessibility, particularly in resource-limited settings or for underserved populations.	Limited access can bias the data sets for one population over the other. It can also lead to disparities in health care, as not all individuals may have equal access to the benefits of precision medicine.
Clinical validation and regulatory approval	Demonstrating the clinical validity and utility of omic-based tests or treatments is crucial for gaining regulatory approval and widespread adoption in clinical practice. Because the approaches are in development, the approval process can become a vicious circle.	Without proper validation, there may be uncertainty about the reliability and effectiveness of omic-based approaches.
Lack of comprehensive databases	Comprehensive databases that catalog omic data from diverse populations are essential for understanding genetic variation and disease susceptibility.	Without comprehensive databases, it can be challenging to apply precision medicine approaches to populations with unique genetic backgrounds.
Patient and physician education	Both patients and healthcare providers may have a limited understanding of omic technologies and how they can be applied in clinical practice.	This can lead to hesitancy or reluctance to adopt precision medicine approaches.
Longitudinal data and follow-up	Omic technologies provide snapshots of a patient's molecular profile at a specific point in time. Long-term data collection and follow-up are crucial for understanding disease progression and treatment response over time.	Without longitudinal data, it may be challenging to optimize and adjust treatment plans as needed.
Rare and complex diseases	Identifying relevant genetic or molecular markers and developing targeted treatments for rare or heterogeneous diseases.	Precision medicine approaches may not be as readily applicable for these conditions, due to the low number of cases.

quality, comparability, and reproducibility across different studies and laboratories. A typical pipeline in proteomic profiling includes sample collection, protein isolation, quantification, digestion, peptide fractionation, mass spectrometry acquisition, database search and protein identification, and data processing and validation.[71] Proteomics is one of the omics most vulnerable to variability.[72,73] Therefore, several standards and guidelines have been established to ensure data quality, comparability, and reproducibility across different laboratories (Table S6). However, the lack of application-/disease-specific universal standards makes it difficult to address variability and improve the reliability to make its application a strong presence in precision medicine.

For cardiovascular diseases, we often rely on ontological categorization (biological processes, cellular components, and molecular function), pathway algorithms (protein annotation, canonical pathways, cluster overrepresentation, and network generation), and complex networks (nonstochastic structure, hierarchy/modularity, robustness/criticality, and actionable prognostication) that all have been curated for global applications rather than cardiovascular-specific context. That said, multiple clinical studies have applied proteomics to gain insights into the complexities of cardiovascular diseases to develop and refine treatment plans. Recently, led by the National Cancer Institute, Clinical Proteomic Tumor Analysis Consortium expanded its scope to include cardiovascular diseases. This initiative focuses on in-depth proteomic profiling of cardiovascular tissues and fluids to identify biomarkers and therapeutic targets. Table S7 summarizes challenges and provides possible solutions relevant to omics studies.

Metabolomics

The metabolomics communities in cardiovascular diseases are facing similar challenges to those in transcriptomics- or proteomics-based research.

The metabolomics community has established standards and guidelines to limit the variability and improve the reproducibility of their findings. The main guidelines are included in the Minimum Information About a Metabolomics Experiment, Metabolomics Standards Initiative, and Metabolite Identification Standard.

Microbiome

Microbiomics, the study of microbial communities in various environments, including the human body, also has established standards and guidelines to ensure data quality, comparability, and reproducibility across different studies and laboratories. Some of the key standards for microbiomics are summarized in Table S8.

Application and Implications of Population Descriptors in Genomics Research

The application of population descriptors in omics research is both deliberate and reflective and should adhere to the guidelines recommended by the National Academies' report on using population descriptors in genetics and genomics research.[74] Recognizing the complexity and variability inherent in genetic data, it is important to choose descriptors that accurately represent the genetic and geographic diversity of the study populations. These descriptors can then be used to carefully facilitate comparisons and highlight genetic patterns that help clarify disease prevalence, treatment efficacy, and biological mechanisms. To enhance transparency and reproducibility, studies should define each descriptor in terms of genetic, environmental, social, and cultural factors.

To decrease the risk for omics studies to worsen social biases,[75] it is important to rely on a strict ethical protocol in each study group. This protocol should include ongoing consultations with ethicists and community representatives to ensure that research practices are culturally sensitive and ethically sound. Open communication and adhering to established guidelines are also critical for public oversight and feedback, particularly when sharing methodologies and findings. By adopting and keeping these best practices at the forefront, the potential risks of reinforcing racial, ethnic, and other stereotypes that might occur with genomics research may be mitigated.

Clinical Requirements

Extracting structured data from clinical documentation, diagnoses, medications, imaging studies, laboratory values, and procedures is facilitated by several internationally recognized structured clinical coding sets, as well as standards for exchanging healthcare information electronically.[76] Internationally recognized clinical coding sets, such as the *ICD*, SNOMED CT, the Healthcare Common Procedure Coding System, and DICOM, play key roles in translating clinical data into machine-readable formats (such as eXtensible Markup Language, comma-separated values, and JavaScript Object Notation), with their integration being largely driven by regulatory requirements and the ongoing push toward electronic healthcare record optimization (Table S1). Clinical workflows can be used to help front-line clinicians record data in structured formats.

Clinical Coding Sets

Disease, treatment, or outcome data may be derivable from data required for clinical care or claims billing. Although the integration of healthcare information systems has been accelerated by regulations, including the

EHR Meaningful Use program and the Physician Quality Reporting System, most systems have focused on documentation for billing charges.[77] Thus, billing code classifications may yield the most complete lists of diagnoses and procedures.

The Healthcare Common Procedure Coding System is a US national code set used for processing claims. HPCS level I codes are part of the current procedural terminology codes, maintained and copyrighted by the American Medical Association, for medical procedures and professional services in outpatient and ambulatory settings, and physician visits to inpatients. The current set (CPT-4) covers services and procedures categorized by Evaluation and Management, Anesthesiology, Surgery, Radiology, Pathology and Laboratory, and Medicine. HPCS level II codes, maintained by the Centers for Medicare & Medicaid Services, except for D codes (dental services) maintained by the American Dental Association, cover products, supplies, and services not included in level I current procedural terminology codes and include durable medical equipment, prosthetics, orthotics, supplies, ambulance services, certain drugs and biologicals, and miscellaneous codes not otherwise classified. SNOMED CT has become a coding standard in the United States for electronic health information exchange and is a standard for interoperability required by the US Health Information Technology Standards Panel.[78] For cardiovascular imaging, DICOM standards have enabled a degree of interoperability and have become a global standard.

Drug databases include the National Drug Code, maintained by the US Food and Drug Administration, and the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use Medical Dictionary for Regulatory Activities, which includes registration, documentation, and safety monitoring of medical products before and after regulatory approval.

Quality

Metrics and outcomes databases have lagged in disease, procedure, drug, and imaging harmonization. Most accessible may be inpatient hospital outcomes using administrative databases, including billing and claim-based sources noted above for individual-level data. The US Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project includes healthcare databases and tools with data elements from inpatient and outpatient discharge records (including inpatient stays, ambulatory surgery and services visits, and emergency department visits) compiled from state, federal, hospital, and private sources. The Healthcare Cost and Utilization Project has been providing patient-level health care and longitudinal hospital care data since 1988, and includes the National (Nationwide) Inpatient Sample, a family of databases and software tools that are the largest publicly available all-payer inpatient healthcare database.[79] Medicare databases, such as the Medicare Provider Analysis and Review, with diagnosis-related groups and billing, as well as cost information from Medicare beneficiaries who use hospital inpatient services can also provide longitudinal patient-level data.[80]

Clinical Workflow

Capturing and using clinical data for precision medicine[47] can be facilitated by designing clinical workflows that guide clinicians and patients with tools embedded directly in the EHR, along with natural language processing, digital health questionnaires in the patient portals to assist patients in tracking outcomes, and integration of data from wearable devices directly into the EHR (Table 1).[81,82]

Digital health questionnaires provide patients with tools such as apps or Web portals that allow them to input their own data before appointments, such as symptoms and family history. Standardized questionnaires can be used to help capture the data in a structured format. Finally, posttreatment digital questionnaires can also be used to monitor patient progress and standardized outcomes.

Phenomics and Applicability to Multiomics

Challenges in Analytical and Modeling Approaches

Phenomics is a rapidly evolving domain, seeking to comprehensively study phenotypes. Despite its potential, many challenges remain, particularly in the existing analytical and modeling approaches, including, but not limited to, the heterogeneous nature of phenotypic data (Figure 5).

Phenotypic data are often linked to other multiomics data, such as genomics, transcriptomics, proteomics, and radiomics, which adds layers of complexity requiring sophisticated tailored approaches for accurate integration, analysis, and interpretation. While the techniques, meta-analysis,[83] Bayesian methods,[84] and machine or deep learning,[85] around the integration and analysis of these data have greatly evolved over time, many obstacles still exist to effectively and meaningfully gain insights from these data. The biggest challenges are 2-fold: integrating/harmonizing the data from different technologies (using a stepwise approach) and consolidating data for the same technology obtained from different sources/centers using different methods.[86]

Several algorithms and analytical tools are available for phenomic data harmonization and analysis, ranging from machine learning models for predictive modeling to network-based approaches for understanding complex interactions within biological systems. These algorithms can uncover latent relationships and provide insights into the underlying biological mechanisms.

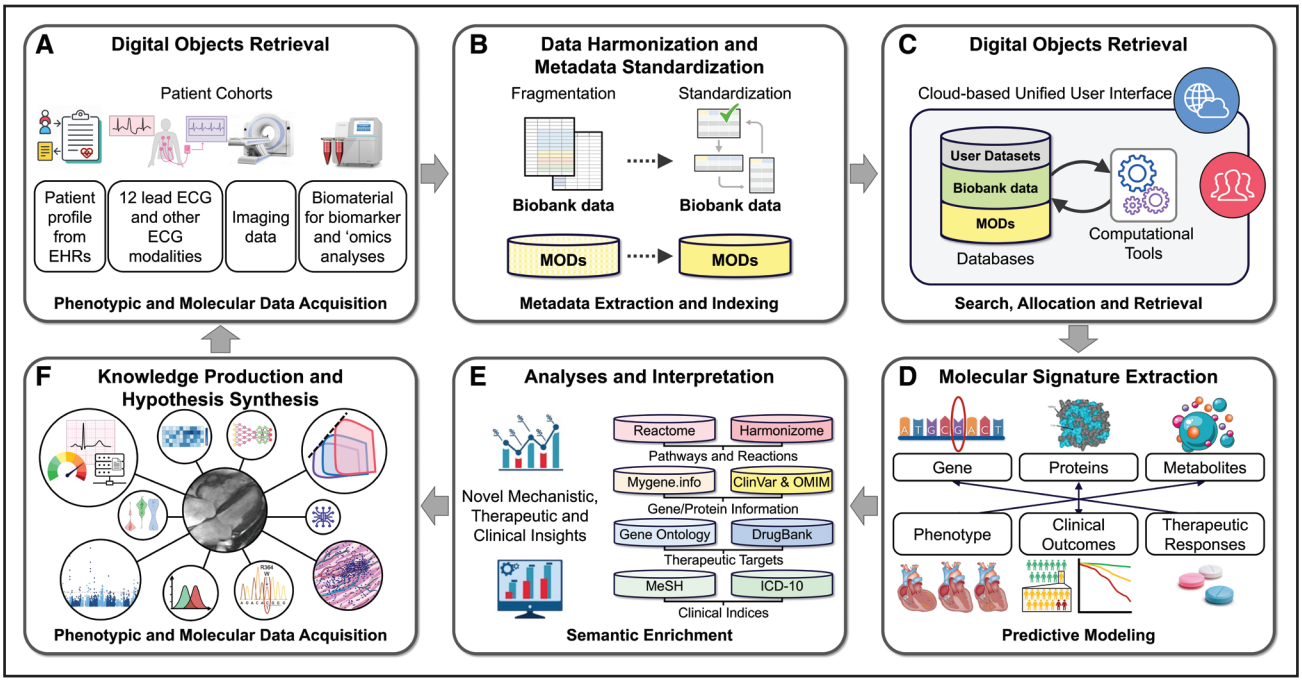

Figure 5. Illustration of the procedural workflow demonstrating the incorporation of data science components within the framework of experimental design in biomedical research or clinical study design.

A, Source data can be obtained from human cohorts or model systems, with tailored techniques and methodologies used to procure phenotypic and molecular data. **B**, Given the diversity of data types and features, such as the sequencing technology applied in transcriptomics data sets, an initial step involves data harmonization, followed by metadata extraction to facilitate indexing and standardization. **C**, Following the transformation of these data into a standardized and accessible format, integration into a unified interface enables investigators to search for and retrieve pertinent digital objects, specifically data sets or computational tools appropriate to the intended study. **D**, These resources are then leveraged to execute cutting-edge analyses, including machine learning and predictive modeling, with the aim of unveiling robust genotype-phenotype associations and delineating molecular signatures for the cohort. **E**, Molecular signatures, thus, obtained undergo subsequent processing and in-depth analysis to derive novel mechanistic, therapeutic, and clinical insights. **F**, Armed with these newfound insights, researchers contribute to the expansive network of biomedical knowledge, thereby propelling cardiovascular research forward. EHR indicates electronic health record; *ICD*, *International Classification of Diseases*; OMIM, Online Mendelian Inheritance in Man; MeSH, Medical Subject Headings; and MOD, Model Organism Database.

An approach expected to promote the harmonization of multiomics data is the FAIR data principles.[87–89] The data FAIRification process supports a metadata schema/method, which captures relations between (omics) measurements and ensures that data structures and concepts are clearly defined and easily interpretable by both humans and computers. Moving forward, the AI-supported analysis of biological systems requires setting the scene to allow growth and expansion of the harmonization methodologies that aim to seamlessly integrate not only different omics sources but also the same omic platform sourced from different centers.[86,90,91]

Differences Across Disease Ontology Systems

Disease ontology systems traditionally operate within predefined categorical boundaries, often limiting the exploration of the vast phenotypic spectrum.[92,93] In contrast, phenomics adopts a more flexible and lateral approach by embracing the variability and nuances inherent in phenotype expressions. This flexibility leads to a broader, more holistic understanding of disease manifestations (with the associated phenotypes) and their potential underlying molecular mechanisms. However, the broader definitions of phenotypes and the increased number of phenotypes can deepen challenges in data harmonization. Mapping and translation frameworks to bridge the terminological and conceptual gaps are often used to condense the observed phenotypes. To be able to maintain the richness of phenotypic data, we need continuous development and testing of integrative approaches to accommodate the depth of clinical phenotypic data while still aligning with established disease ontology frameworks.[94,95]

Managing Patient Phenomics Data

Phenomics data in patient studies can be composed of a comprehensive set of observable traits (physical, physiological, and molecular characteristics). These data are gathered through a variety of methods, including clinical assessments, imaging techniques, genetic sequencing, and omics technologies. Once collected, phenomics data undergo a rigorous analysis process: cleaning and quality control to ensure accuracy and reliability, and statistical and computational analysis to extract meaningful patterns, correlations, and associations within the data.

Standardization efforts ensure consistency and compatibility across different phenomics data sets. Established guidelines and ontologies are often used to standardize data representation and terminology to target seamless data sharing and collaborative research in the field of phenomics. Finally, and in contrast to the broad scope of phenomics, CDM phenotypes refer specifically to disease definitions or categorizations used within CDMs to ensure standardized data formatting across different sources, as well as ensuring consistency and accuracy.

Harmonizing Data from Different Databases and Internationally

Similar to all domains, data harmonization is a cornerstone in phenomics, especially when assimilating multicentric and international data sets.[96] Increasingly massive amounts of multiomics data are being generated across diverse biomedical domains and geographic locations, with differing data types, sharing requirements, and storage and handling approaches.[97,98] The endeavor to harmonize phenomics data encompasses not only technical challenges but also legal and ethical considerations, particularly when dealing with international data sharing.[99,100] In this regard, there are successful examples, such as the NIH All of Us and the UK Biobank, where EHRs have been mapped to OMOP in cloud-based tools running on different platforms. However, primary care data for UK Biobank have been initially unlinked to participants, briefly linked over the pandemic, with a review underway about long-term linkage. Therefore, moving forward, establishing international guidelines on data interoperability and harmonization can facilitate the pooling of data from different sources, enriching existing and new data sets, increasing the generalizability of findings, allowing for ancestral-specific findings, enhancing statistical power, and fostering international collaborative efforts.[101] It is apparent that the complexity of harmonizing multiple platforms and data sets, in different languages, which have been built on different hosting systems, compounds the implementation complexity of this endeavor into clinical workflows that impact health care. These efforts are particularly important for genomic and precision medicine, where most efforts are vital to include non-European ancestry individuals, a strategy that is expected to improve our knowledge, for example, reclassifying variants, as well as understanding the impact of rare and common variants in those of different ancestry.

Reprocessing and Applying Phenomics Tools for Phenomics Maps

The creation of phenomics maps and visual representations (especially for non-Europeans) of complex phenotype-genotype interactions necessitates the reprocessing of existing data through advanced phenomics tools.[102,103] These maps serve as invaluable resources for understanding the phenotypic landscape of diseases, which is particularly essential given the complexity and high dimensionality of the multiomics data.[104] Robust computational methods can often help with the accuracy and reproducibility of phenomics maps; however, the endeavor to reprocess and map phenomics data is not only technically demanding but also resource-intensive, necessitating substantial computational resources and expertise. Despite these challenges, phenomics maps hold the promise of unveiling novel insights into disease mechanisms and fostering the development of personalized medicine. Finally, although the increasing use of AI/machine learning technologies has helped address many previous challenges inherent in multiomics data integration, it has also added to the complexity of classification.[105]

Phenomics for Biomarkers and Risk Factors: Aligning Genomics and Proteomics to Disease

Phenomics stands at the forefront of biomarker discovery and validation.[106–108] Novel biomarkers that are instrumental for early disease detection, monitoring, and personalized medicine can be identified using the comprehensive understanding facilitated by phenomics.[109] In addition, proper understanding of disease cause, which is fundamental for better risk stratification, diagnostic, and prognostic evaluations, requires accurate evaluation of the phenotype-genotype relationship.[110] By aligning genomic and proteomic data to disease phenotypes, phenomics provides a unique multidimensional view of disease risk factors and its molecular underpinnings. Eventually, this comprehensive and nuanced understanding leads to better risk assessment, early detection, and therapeutic interventions.[110,111]

CHALLENGES

There are several challenges that must be overcome to facilitate national and international efforts[112] in cardiovascular precision medicine.[3] These include governance, consent, availability and access, privacy and security, use of AI, and implementation. These are outlined in Table S9.

The American Heart Association has recently published a policy statement on principles for health information collection, sharing, and use.[76] A challenge faced by many researchers is understanding the differing rules and regulations in data governance based on which country one is operating in, or working with, to share data. The European Union has established the General Data Protection Regulation. The cornerstone of the General Data Protection Regulation is that no organization can collect, store, or use personal data without the explicit consent of the data subject.

The historical model of identifying risk factors for disease and focusing on individual behavioral changes has not been able to resolve health inequities among groups

that have been economically and socially marginalized. Underrepresentation of racial and ethnic minority groups in clinical trials, and even as participants in receiving health care, has resulted in gaps in our understanding of race-related differences in disease pathobiology, diet, lifestyle, and drug responses. Therefore, the existing databases and registries have been influenced by structural racism and many forms of bias, which are likely to be integrated into the AI algorithm development. Consequently, a big step toward achieving health equity is not only the inclusion in clinical trials of culturally and linguistically diverse people but also policymakers, legislators, housing administrators, manufacturers, urban planners, and health insurance stakeholders.[75]

Over the next decade, an exponential rise in the use of AI within the healthcare sector is anticipated.[113] To maximize the benefits of AI while minimizing potential harm, several data and regulatory standards encompassing training data, evaluation metrics, reliability, fairness, generalizability, explainability, and traceability are currently being developed by national and transnational organizations,[47] adopting a risk-based approach.[114]

The use of clinical data in discovery and care hinges on its effective integration with AI and machine learning, which brings a wide range of complex and multifaceted ethical concerns surrounding bias, inequity, and disparities in health care. Examples include, how biased data, subjective decisions in algorithm design, and unequal access to new technological innovations may perpetuate care inequities. Several steps can be taken to minimize the risk of bias or inequity. These include ensuring that the data used to train the AI models come from a diverse and representative group. Statistical techniques can be used to identify disparities in model predictions across groups and algorithms, creating a bias detection mitigation tool. Although not necessary, when the model is transparent about how decisions are made, it can help researchers and clinicians understand the rationale behind AI decisions, while continuous monitoring and evaluation of the model are required. Addressing ethical challenges demands ongoing vigilance, collaboration among stakeholders, and a steadfast commitment to equity and fairness in this new era of health care.

In order for AI technologies to develop a trusting social license of the public, providers, and patients, a continuous effort will be required, as failure at any level of that trust will be perceived as commercially driven hype without the due care, resulting in the social license's total collapse.[75] For that matter, public engagement and dialog will ensure that the use of data in health care meets certain core societal expectations and values, as well as builds and maintains broad trust and acceptance. Thus, public dialog will also attempt to ascertain the views of society, regarding the ethical responsibilities of data usage, as well as its design and uses. Consequently, an essential approach to building trust and facilitating a smooth digital transition of health care is to redesign training programs for the health workforce and improve general education.[115]

PERSPECTIVES AND CONCLUSIONS

Structured data in clinical documentation are essential for capturing comprehensive data that are interoperable,[21] machine-readable, and accessible for research and precision medicine. Despite the breadth of these coding sets, much of the existing data resides in billing data. The importance of user-friendly workflows cannot be overstated; by incorporating features such as standardized EHR fields, voice recognition, and digital health questionnaires, the healthcare system can ensure both clinician and patient participation in data capture. As we move forward, achieving a balance between efficient data capture for billing and ensuring comprehensive care, is paramount. The development of structured data tools and the continued emphasis on integrating them seamlessly into clinical practice will shape the future of patient care, enabling precision medicine[3] and enhancing patient outcomes.[116]

ARTICLE INFORMATION

Affiliations

Department of Heart and Vascular, Center for Inherited Cardiovascular Diseases, WellSpan Health, York, PA (C.A.A.C.). Department of Cardiology, Barts Heart Center, London, United Kingdom (C.A.A.C., N.A.). Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN (C.A.A.C.). Departments of Cardiology & Biomedical Informatics, Biostatistics, and Epidemiology, University of Missouri, Columbia (F.A.). Johns Hopkins University, Baltimore, MD (B.A.). Division of Cardiovascular Medicine, Department of Medicine, Cardio-Oncology Program (D.A.) and Division of Cancer Prevention and Control, Department of Medicine, College of Medicine (D.A.), The Ohio State University, Columbus. The William Harvey Research Institute, London School of Medicine & Dentistry (N.A.) and National Institute for Health and Care Research, Barts Cardiovascular Biomedical Research Centre (N.A.), Queen Mary University of London, United Kingdom. Departments of Cardiovascular Medicine, Heart, Vascular & Thoracic Institute & Cardiovascular and Metabolic Sciences, Lerner Research Institute, Cleveland Clinic, OH (M.K.C.). Institute of Health Informatics, University College London, United Kingdom (S.D.). British Heart Foundation Data Science Center, Health Data Research UK, London (S.D.). Department of Biomedical Engineering, Department of Biostatistics & Bioinformatics, Duke Clinical Research Institute, Duke University, Durham, NC (J.D.). American Heart Association, Dallas, TX (J.L.H.). Center for Preventive Cardiology, Knight Cardiovascular Institute, Oregon Health & Science University, Portland (N.P.). Hofstra School of Medicine, North Shore-Long Island Jewish Health System, New York, NY (D.J.S.). Department of Cardiology, Veterans Affairs Medical Center, Washington, DC (J.D.V.). School of Medicine and Health Sciences, Georgetown University, Washington, DC (J.D.V.). Cardiovascular Research Center, Massachusetts General Hospital, Boston (A.A.A.). Broad Institute, Massachusetts Institute of Technology, Cambridge (A.A.A.).

Sources of Funding

Dr Armoundas is funded by the Institute of Precision Medicine (grant 17UNPG33840017) of the American Heart Association, the RICBAC Foundation, and the National Institutes of Health grants 1 R01 HL135335-01, 1 R01 HL161008-01, 1 R21 HL137870-01, 1 R21EB026164-01, and 3R21EB026164-02S1. Dr Addison is funded by grants K23-HL155890 and R01HL170038. Dr Chahal is funded by NextGen. The NextGen project received funding from the European Union's Horizon Europe under grant 101136962. The British-associated partners of NextGen were funded by UK Research and Innovation (UKRI) under the UK Government's Horizon Europe Funding Guarantee (grants 10098097 and 10104323). The European Union NextGen project was funded by the Swiss Confederation, Schweizerische Eidgenossenschaft, Confédération Suisse, Confederazione Svizzera, Confederaziunsvizra, the Federal Department of Economic Affairs, Education and Research, and the State Secretariat for Education, Research and Innovation.

Disclosures

All authors declare no conflicts of interest, no relevant relationships to disclose, and no funding received from any companies or organizations related to the subject of this article.

Acknowledgments

The authors thank Mohamad Alkhayat, MD, and Kristian Galanti, MD, for assistance in proof-reading and submission.

Supplemental Material

Tables S1–S9

REFERENCES

1. Armoundas AA, Narayan SM, Arnett DK, Spector-Bagdady K, Bennett DA, Celi LA, Friedman PA, Gollob MH, Hall JL, Kwitek AE, et al; on behalf of the American Heart Association Institute for Precision Cardiovascular Medicine; Council on Cardiovascular and Stroke Nursing; Council on Lifelong Congenital Heart Disease and Heart Health in the Young; Council on Cardiovascular Radiology and Intervention; Council on Hypertension; Council on the Kidney in Cardiovascular Disease; and Stroke Council. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. *Circulation*. 2024;149:e1028. doi: 10.1161/CIR.0000000000001201
2. NIH. *NIH broadens genomic data-sharing policy. Cancer Discov*. 2014;4:OF4. doi: 10.1158/2159-8290
3. Bota P, Thambiraj G, Bollepalli SC, Armoundas AA. Artificial intelligence algorithms in cardiovascular medicine: an attainable promise to improve patient outcomes or an inaccessible investment? *Curr Cardiol Rep*. 2024;26:1477–1485. doi: 10.1007/s11886-024-02146-y
4. Donnelly WJ. Viewpoint: patient-centered medical care requires a patient-centered medical record. *Acad Med*. 2005;80:33–38. doi: 10.1097/00001888-200501000-00009
5. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, Lehmann HP, Hripcsak G, Hartzog TH, Cimino JJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care*. 2013;51:S30–S37. doi: 10.1097/MLR.0b013e31829b1dbd
6. Lee RY, Kross EK, Torrence J, Li KS, Sibley J, Cohen T, Lober WB, Engelberg RA, Curtis JR. Assessment of natural language processing of electronic health records to measure goals-of-care discussions as a clinical trial outcome. *JAMA Netw Open*. 2023;6:e231204. doi: 10.1001/jamanetworkopen.2023.1204
7. Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. *J Am Med Inform Assoc*. 2011;18:539. doi: 10.1136/amiajnl-2011-000501
8. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform*. 2008;17:67–79. doi: 10.1055/s-0038-1638585
9. HITECH Act Enforcement Interim Final Rule. In: (OCR) HaHSHotUSGOfCR, ed. 2009. Accessed January 8, 2025. https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html
10. Hirsch JA, Nicola G, McGinty G, Liu RW, Barr RM, Chittle MD, Manchikanti L. ICD-10: history and context. *AJNR Am J Neuroradiol*. 2016;37:596–599. doi: 10.3174/ajnr.A4696
11. Feinstein JA, Gill PJ, Anderson BR. Preparing for the International Classification of Diseases, 11th Revision (ICD-11) in the US Health Care System. *JAMA Health Forum*. 2023;4:e232253. doi: 10.1001/jamahealthforum.2023.2253
12. International Classification of Diseases (ICD-9). In: (WHO) WHO, ed. 2009. Accessed January 8, 2025. www.who.int
13. Healthcare Common Procedure Coding System (HCPCS). In: (CMS) CfMaMS, ed. 2015. Accessed January 8, 2025. www.cms.gov
14. ICD-10 Procedure Coding System (ICD-10-PCS). In: (CMS) CfMaMS, ed. 2015. Accessed January 8, 2025. www.cms.hhs.gov
15. Classification of Surgical Operations and Procedures 4th version. In: Surveys NHSNOoPCOa, ed. 2023. Accessed January 8, 2025. https://isd.digital.nhs.uk/trud/user/guest/group/0/pack/10
16. SNOMED. Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT). In: International S, ed. Accessed September 16, 2023. www.snomed.org
17. Spackman K. SNOMED RT and SNOMEDCT. Promise of an international clinical terminology. *MD Comput*. 2000;17:29. PMID: 11189756.
18. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, Mozaffar H, Mason K, Eason S, Hinder S, et al. Reconceptualising the digital maturity of health systems. *Lancet Digit Health*. 2019;1:e200–e201. doi: 10.1016/S2589-7500(19)30083-4
19. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag*. 2013;10:1c. PMID: 24159271
20. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. *Mayo Clin Proc*. 2006;81:741–748. doi: 10.4065/81.6.741
21. Armoundas AA, Ahmad FS, Bennett DA, Chung MK, Davis LL, Dunn J, Narayan SM, Slotwiner DJ, Wiley KK Jr, Khera R; American Heart Association Data Science and Precision Medicine Committee of the Council on Genomic and Precision Medicine and Council on Clinical Cardiology; Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation; Council on Cardiovascular and Stroke Nursing; Council on Lifestyle and Cardiometabolic Health; Council on Peripheral Vascular Disease; Council on Quality of Care and Outcomes Research; and Stroke Council. Data interoperability for ambulatory monitoring of cardiovascular disease: a scientific statement from the American Heart Association. *Circ Genom Precis Med*. 2024;17:e000095. doi: 10.1161/HCG.0000000000000095
22. Utter GH, Cox GL, Atolagbe OO, Owens PL, Romano PS. Conversion of the agency for healthcare research and quality's quality indicators from ICD-9-CM to ICD-10-CM/PCS: the process, results, and implications for users. *Health Serv Res*. 2018;53:3704–3727. doi: 10.1111/1475-6773.12981
23. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, Ghali WA, Investigators I. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 2008;43:1424–1441. doi: 10.1111/j.1475-6773.2007.00822.x
24. Rosenbloom ST, Carroll RJ, Warner JL, Matheny ME, Denny JC. Representing knowledge consistently across health systems. *Yearb Med Inform*. 2017;26:139–147. doi: 10.15265/IY-2017-018
25. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574–578. doi: 10.1038/psp.2013.52
26. Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform*. 2016;64:333–341. doi: 10.1016/j.jbi.2016.10.016
27. Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, van Thiel GJM, Cronin M, Brobert G, Vardas P, et al; Innovative Medicines Initiative 2nd Programme, Big Data for Better Outcomes, BigData@Heart Consortium of 20 Academic and Industry Partners including ESC. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J*. 2018;39:1481–1495. doi: 10.1093/eurheartj/ehx487
28. FDA. Real-world evidence. Accessed October 5, 2024. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
29. Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. Paper/Poster presented at: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013;326–331. Doi: 10.1109/cbms.2013.6627810
30. Mantri M, Taran S, Sunder G. DICOM integration libraries for medical image interoperability: a technical review. *IEEE Rev Biomed Eng*. 2020;15:247–259. doi: 10.1109/rbme.2020.3042642
31. Harron K. Data linkage in medical research. *BMJ Med*. 2022;1:e000087. doi: 10.1136/bmjmed-2021-000087
32. Petersen SJ, Lieberthal RD, Miller KJ, Vakil NH. Privacy preserving record linkage (PPRL) strategy and recommendations. 2023. Accessed January 8, 2025. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.nia.nih.gov/sites/default/files/2023-08/pprl-linkage-strategies-preliminary-report.pdf
33. Pcornet. Privacy: five ways patients win with PPRL in PCORnet. March 2023. Accessed October 5, 2024. https://pcornet.org/news/tag/privacy
34. Harron K, Doidge JC, Goldstein H. Assessing data linkage quality in cohort studies. *Ann Hum Biol*. 2020;47:218–226. doi: 10.1080/03014460.2020.1742379
35. Antunes RS, André da Costa C, Küderle A, Yari IA, Eskofier B. Federated learning for healthcare: systematic review and architecture proposal. *ACM Trans Intell Syst Technol*. 2022;13:1–23. doi: 10.1145/3501813

36. Cury RC, Abbara S, Achenbach S, Agatston A, Berman DS, Budoff MJ, Dill KE, Jacobs JE, Maroules CD, Rubin GD, et al. Coronary Artery Disease - Reporting and Data System (CAD-RADS): an expert consensus document of SCCT, ACR and NASCI: endorsed by the ACC. *J Cardiovasc Comput Tomogr*. 2016;9:1099–1113. doi: 10.1016/j.jcmg.2016.05.005
37. Hundley WG, Bluemke DA, Bogaert J, Flamm SD, Fontana M, Friedrich MG, Grosse-Wortmann L, Karamitsos TD, Kramer CM, Kwong RY, et al. Society for Cardiovascular Magnetic Resonance (SCMR) guidelines for reporting cardiovascular magnetic resonance examinations. *J Cardiovasc Magn Reson*. 2022;24:29. doi: 10.1186/s12968-021-00827-z
38. Mitchell C, Rahko PS, Blauwet LA, Canaday B, Finstuen JA, Foster MC, Horton K, Ogunyankin KO, Palma RA, Velazquez EJ. Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the American Society of Echocardiography. *J Am Soc Echocardiogr*. 2019;32:1–64. doi: 10.1016/j.echo.2018.06.004
39. Lewis AE, Weiskopf N, Abrams ZB, Foraker R, Lai AM, Payne PRO, Gupta A. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. 2023;30:1730–1740. doi: 10.1093/jamia/ocad120
40. HHS. *HHS Finalizes Historic Rules to Provide Patients With More Control of Their Patient Data*. Department of Health and Human Services; 2020. Accessed May 28, 2025. https://www.cms.gov/newsroom/press-releases/hhs-finalizes-historic-rules-provide-patients-more-control-their-health-data
41. Stevens LM, de Lemos JA, Das SR, Rutan C, Alger HM, Elkind MSV, Zhao J, Iyer K, Figueroa CA, Hall JL. American Heart Association precision medicine platform addresses challenges in data sharing. *Circ Cardiovasc Qual Outcomes*. 2021;14:e007949. doi: 10.1161/CIRCOUTCOMES.121.007949
42. Schatz MC, Philippakis AA, Afgan E, Banks E, Carey VJ, Carroll RJ, Culotti A, Ellrott K, Goecks J, Grossman RL, et al. Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-Space. *Cell Genom*. 2022;2:100085. doi: 10.1016/j.xgen.2021.100085
43. HHS. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. 2012. Accessed January 8, 2025. https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveredentities/De-identification/hhs_deid_guidance.pdf
44. Seastedt KP, Schwab P, O'Brien Z, Wakida E, Herrera K, Marcelo PGF, Agha-Mir-Salim L, Frigola XB, Ndulue EB, Marcelo A, et al. Global healthcare fairness: we should be sharing more, not less, data. *PLOS Digit Health*. 2022;1:e0000102. doi: 10.1371/journal.pdig.0000102
45. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS. A systematic review of barriers to data sharing in public health. *BMC Public Health*. 2014;14:1144. doi: 10.1186/1471-2458-14-1144
46. National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Board on Research Data and Information; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Division on Earth and Life Studies; Nuclear and Radiation Studies Board; Division of Behavioral and Social Sciences and Education; Committee on National Statistics; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Reproducibility and Replicability in Science. Reproducibility and replicability in science. 2019. Washington: National Academic Press; 2019.
47. Bazoukis G, Hall J, Loscalzo J, Antman EM, Fuster V, Armoundas AA. The inclusion of augmented intelligence in medicine: a framework for successful implementation. *Cell Rep Med*. 2022;3:100485. doi: 10.1016/j.xcrm.2021.100485
48. DICOM. DICOM. Accessed January 8, 2025. https://www.dicomstandard.org
49. Larobina M, Murino L. Medical image file formats. *J Digit Imaging*. 2014;27:200–206. doi: 10.1007/s10278-013-9657-9
50. IHE. Integrating the Healthcare Enterprise (IHE). In: International I, ed. Accessed January 8, 2025. https://www.ihe.net/
51. Gostin LO. National Health Information privacy regulations under the Health Insurance Portability and Accountability Act. *JAMA*. 2001;285:3015–3021. doi: 10.1001/jama.285.23.3015
52. European Society of Radiology. The new EU General Data Protection Regulation: what the radiologist should know. *Insights Imaging*. 2017;8:295–299. doi: 10.1007/s13244-017-0552-7
53. Desjardins B, Mirsky Y, Ortiz MP, Glozman Z, Tarbox L, Horn R, Horii SC. DICOM images have been hacked! Now what? *AJR Am J Roentgenol*. 2020;214:727–735. doi: 10.2214/AJR.19.21958
54. NIST. In: Excellence NCCo, ed. 2020. Accessed January 8, 2025. https://www.nccoe.nist.gov/healthcare/securing-picture-archiving-and-communication-system
55. NCCoE. Securing picture archiving and communication system. 2020. Accessed January 8, 2025. https://www.nccoe.nist.gov/healthcare/securing-picture-archiving-and-communication-system
56. Amaral-Zettler L. AMA issues new principles for AI development, deployment & use. Chicago, IL; 2023. Accessed January 8, 2025. https://www.ama-assn.org/press-center/press-releases/ama-issues-new-principles-ai-development-deployment-use
57. Jean-Baptiste L, Mouazer A, Sedki K, Tsopra R. Translating the Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) electronic health records to an OWL ontology. *Stud Health Technol Inform*. 2022;290:76-80. doi: 10.3233/SHTI220035.
58. ALL_OF_US. All of Us, research program. 2020. Accessed January 8, 2025. https://allofusnihgov
59. EU. *European Health Data Space*. 2020. Accessed September 16, 2023. https://eceuropaeu/health/ehealth/dataspace
60. HDR_UK. *Digital Innovation Hub Programme Prospectus. Appendix: Principles for Participation*. 2019. Accessed September 16, 2023. https://www.hdruk.ac.uk/wp-content/uploads/2019/07/Digital-Innovation-Hub-Programme-Prospectus-Appendix-Principles-for-Participation.pdf
61. Yang L. *The Human Colossus Foundation Releases "Overlays Capture Architecture."* The Human Colossus Foundation; 2018. Accessed January 8, 2025. https://humancolossus.foundation/blog/oca-v1-launch
62. Musunuru K, Ingelsson E, Fornage M, Liu P, Murphy AM, Newby LK, Newton-Cheh C, Perez MV, Voora D, Woo D. The expressed genome in cardiovascular diseases and stroke: refinement, diagnosis, and prediction: a scientific statement from the American Heart Association. *Circ Cardiovasc Genet*. 2017;10:e1-e24. doi: 10.1161/HCG.0000000000000037
63. Demetri GD, von Mehren M, Blanke CD, Van den Abbeele AD, Eisenberg B, Roberts PJ, Heinrich MC, Tuveson DA, Singer S, Janicek M, et al. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N Engl J Med*. 2002;347:472–480. doi: 10.1056/NEJMoa020461
64. Flaherty KT, Gray R, Chen A, Li S, Patton D, Hamilton SR, Williams PM, Mitchell EP, Lafrate AJ, Sklar J, et al. The Molecular Analysis for Therapy Choice (NCI-MATCH) Trial: lessons for genomic trial design. *J Natl Cancer Inst*. 2020;112:1021–1029.
65. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther*. 2009;86:97–100. doi: 10.1038/clpt.2009.68
66. Janne PA, Shaw AT, Pereira JR, Jeannin G, Vansteenkiste J, Barrios C, Franke FA, Grinsted L, Zazulina V, Smith P, et al. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol*. 2013;14:38–47. doi: 10.1016/S1470-2045(12)70489-8
67. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, Halai D, Baple E, Craig C, Hamblin A, et al; 100 000 Genomes Project. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ*. 2018;361:k1687. doi: 10.1136/bmj.k1687
68. Arrell DK, Terzic A. Systems proteomics for translational network medicine. *Circ Cardiovasc Genet*. 2012;5:478. doi: 10.1161/CIRCGENETICS.110.958991
69. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405–424. doi: 10.1038/gim.2015.30
70. Josephs KS, Roberts AM, Theotokis P, Walsh R, Ostrowski PJ, Edwards M, Fleming A, Thaxton C, Roberts JD, Care M, et al. Beyond gene-disease validity: capturing structured data on inheritance, allelic requirement, disease-relevant variant classes, and disease mechanism for inherited cardiac conditions. *Genome Med*. 2023;15:86. doi: 10.1186/s13073-023-01246-8
71. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK Jr, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, et al. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol*. 2007;25:887–893. doi: 10.1038/nbt1329
72. Piehowski PD, Petyuk VA, Orton DJ, Xie F, Moore RJ, Ramirez-Restrepo M, Engel A, Lieberman AP, Albin RL, Camp DG, et al. Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J Proteome Res*. 2013;12:2128–2137. doi: 10.1021/pr301146m
73. Valikangas T, Suomi T, Chandler CE, Scott AJ, Tran BQ, Ernst RK, Goodlett DR, Elo LL. Benchmarking tools for detecting longitudinal differential expression in proteomics data allows establishing a

robust reproducibility optimization regression approach. *Nat Commun*. 2022;13:7877. doi: 10.1038/s41467-022-35564-z
74. National Academies of Sciences E, Medicine. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. The National Academies Press; 2023.
75. Bazoukis G, Loscalzo J, Hall JL, Bollepalli SC, Singh JP, Armoundas AA. Impact of social determinants of health on cardiovascular disease. *J Am Heart Assoc*. 2025;14:e039031. doi: 10.1161/JAHA.124.039031
76. Spector-Bagdady K, Armoundas AA, Arnaout R, Hall JL, Yeager McSwain B, Knowles JW, Price WN 2nd, Rawat DB, Riegel B, Wang TY, et al; American Heart Association Advocacy Coordinating Committee. Principles for health information collection, sharing, and use: a policy statement from the American Heart Association. *Circulation*. 2023;148:1061–1069. doi: 10.1161/CIR.0000000000001173
77. Windle JR, Katz AS, Dow JP Jr, Fry ET, Keller AM, Lamp T, Lippitt A Jr, Paruche MP, Resnic FS, Serwer GA, et al. 2016 ACC/ASE/ASNC/HRS/SCAI Health Policy Statement on integrating the healthcare enterprise. *J Am Coll Cardiol*. 2016;68:1348–1364. doi: 10.1016/j.jacc.2016.04.017
78. HITSP. Document number: HITSP 05 N 21. In: Panel ANSIHITS, ed.; 2005. Accessed January 8, 2025. https://www.consortiuminfo.org/list/health-information-technology-standards-panel-hitsp
79. AHRQ. Overview of the National (Nationwide) Inpatient Sample (NIS). In: Quality AfHRa, ed.; 2012. Accessed January 8, 2025. https://hcup-us.ahrq.gov/nisoverview.jsp
80. CMS. Medicare fee for service for parts A & B. In: Services CfMM, ed.; 1983. Accessed January 8, 2025. https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-fee-for-service-parts-a-b/medpar
81. Kulkarni K, Sevakula RK, Kassab MB, Nichols J, Roberts JD Jr, Isselbacher EM, Armoundas AA. Ambulatory monitoring promises equitable personalized healthcare delivery in underrepresented patients. *Eur Heart J Digit Health*. 2021;2:494–510. doi: 10.1093/ehjdh/ztab047
82. Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. *J Am Coll Cardiol*. 2020;75:1582–1592. doi: 10.1016/j.jacc.2020.01.046
83. Kim S, Jhong JH, Lee J, Koo JY. Meta-analytic support vector machine for integrating multiple omics data. *BioData Min*. 2017;10:2. doi: 10.1186/s13040-017-0126-8
84. Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics*. 2018;19:71–86. doi: 10.1093/biostatistics/kxx017
85. Ma T, Zhang A. Integrate multi-omics data with biological interaction networks using Multi-view Factorization AutoEncoder (MAE). *BMC Genomics*. 2019;20:944. doi: 10.1186/s12864-019-6285-x
86. Kuhn Cuellar L, Friedrich A, Gabernet G, de la Garza L, Fillinger S, Seyboldt A, Koch T, Zur Oven-Krockhaus S, Wanke F, Richter S, et al. A data management infrastructure for the integration of imaging and omics data in life sciences. *BMC Bioinf*. 2022;23:61. doi: 10.1186/s12859-022-04584-3
87. Inau ET, Sack J, Waltemath D, Zeleke AA. Initiatives, concepts, and implementation practices of FAIR (findable, accessible, interoperable, and reusable) data principles in health data stewardship practice: protocol for a scoping review. *JMIR Res Protoc*. 2021;10:e22505. doi: 10.2196/22505
88. Wilkinson MD, Dumontier M, Jan Aalbersberg I, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. Addendum: the FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2019;6:6. doi: 10.1038/s41597-019-0009-6
89. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. doi: 10.1038/sdata.2016.18
90. Yu CT, Chao BN, Barajas R, Haznadar M, Maruvada P, Nicastro HL, Ross SA, Verma M, Rogers S, Zanetti KA. An evaluation of the National Institutes of Health grants portfolio: identifying opportunities and challenges for multi-omics research that leverage metabolomics data. *Metabolomics*. 2022;18:29. doi: 10.1007/s11306-022-01878-8
91. Holzinger A, Haibe-Kains B, Jurisica I. Why imaging data alone is not enough: AI-based integration of imaging, omics, and clinical data. *Eur J Nucl Med Mol Imaging*. 2019;46:2722–2730. doi: 10.1007/s00259-019-04382-9
92. Kohler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Ayme S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, et al. The human phenotype ontology in 2017. *Nucleic Acids Res*. 2017;45:D865–D876. doi: 10.1093/nar/gkw1039
93. Galeota E, Kishore K, Pelizzola M. Ontology-driven integrative analysis of omics data through Onassis. *Sci Rep*. 2020;10:703. doi: 10.1038/s41598-020-57716-1
94. Wang Z, He Y. Precision omics data integration and analysis with interoperable ontologies and their application for COVID-19 research. *Brief Funct Genomics*. 2021;20:235–248. doi: 10.1093/bfgp/elab029
95. Ghandikota S, Hershey GKK, Mersha TB. GENEASE: real time bioinformatics tool for multi-omics and disease ontology exploration, analysis and visualization. *Bioinformatics*. 2018;34:3160–3168. doi: 10.1093/bioinformatics/bty182
96. Oldoni E, Saunders G, Bietrix F, Garcia Bermejo ML, Niehues A, t Hoen PAC, Nordlund J, Hajduch M, Scherer A, Kivinen K, et al. Tackling the translational challenges of multi-omics research in the realm of European personalised medicine: a workshop report. *Front Mol Biosci*. 2022;9:974799. doi: 10.3389/fmolb.2022.974799
97. Palsson B, Zengler K. The challenges of integrating multi-omic data sets. *Nat Chem Biol*. 2010;6:787–789. doi: 10.1038/nchembio.462
98. Krassowski M, Das V, Sahu SK, Misra BB. State of the field in multi-omics research: from computational needs to data mining and sharing. *Front Genet*. 2020;11:610798. doi: 10.3389/fgene.2020.610798
99. Abou-Khalil B, Alldredge B, Bautista J, Berkovic S, Bluvstein J, Boro A, Cascino G, Consalvo D, Cristofaro S, Crumrine P, et al; EPGP Collaborative. The epilepsy phenome/genome project. *Clin Trials*. 2013;10:568–586. doi: 10.1177/1740774513484392
100. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Worheide MA, Oerton E, Cook J, Stewart ID, Kerrison ND, et al. Mapping the proteogenomic convergence of human diseases. *Science*. 2021;374:eabj1541. doi: 10.1126/science.abj1541
101. Ying W. Phenomic studies on diseases: potential and challenges. *Phenomics*. 2023;3:285–299. doi: 10.1007/s43657-022-00089-4
102. Lussier YA, Liu Y. Computational approaches to phenotyping: high-throughput phenomics. *Proc Am Thorac Soc*. 2007;4:18–25. doi: 10.1513/pats.200607-142JG
103. Zhou Y, Kamruzzaman M, Schnable P, Krishnamoorthy B, Kalyanaraman A, Wang B. Pheno-mapper: an interactive toolbox for the visual exploration of phenomics data. *arXiv*. 2021:1–10. doi: 10.1145/3459930.3469511
104. Schlotter F, Halu A, Goto S, Blaser MC, Body SC, Lee LH, Higashi H, DeLaughter DM, Hutcheson JD, Vyas P, et al. Spatiotemporal multi-omics mapping generates a molecular atlas of the aortic valve and reveals networks driving disease. *Circulation*. 2018;138:377–393. doi: 10.1161/CIRCULATIONAHA.117.032291
105. Nabwire S, Suh HK, Kim MS, Baek I, Cho BK. Review: application of artificial intelligence in phenomics. *Sensors (Basel)*. 2021;21:4363. doi: 10.3390/s21134363
106. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18:696–705. doi: 10.1038/s41568-018-0060-1
107. Xiao L, Xiao T, Wang ZM, Cho WC, Xiao ZQ. Biomarker discovery of nasopharyngeal carcinoma by proteomics. *Expert Rev Proteomics*. 2014;11:215–225. doi: 10.1586/14789450.2014.897613
108. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. *J Pharm Biomed Anal*. 2014;87:1–11. doi: 10.1016/j.jpba.2013.08.041
109. Hristova VA, Chan DW. Cancer biomarker discovery and translation: proteomics and beyond. *Expert Rev Proteomics*. 2019;16:93–103. doi: 10.1080/14789450.2019.1559062
110. Karczewski KJ, Snyder MP. Integrative omics for health and disease. *Nat Rev Genet*. 2018;19:299–310. doi: 10.1038/nrg.2018.4
111. Borrebaeck CA. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nat Rev Cancer*. 2017;17:199–204. doi: 10.1038/nrc.2016.153
112. Whitsel LP, Wilbanks J, Huffman MD, Hall JL. The role of government in precision medicine, precision public health and the intersection with healthy living. *Prog Cardiovasc Dis*. 2019;62:50–54. doi: 10.1016/j.pcad.2018.12.002
113. Sevakula RK, Au-Yeung WM, Singh JP, Heist EK, Isselbacher EM, Armoundas AA. State-of-the-art machine learning techniques aiming to improve patient outcomes pertaining to the cardiovascular system. *J Am Heart Assoc*. 2020;9:e013924. doi: 10.1161/JAHA.119.013924
114. Meszaros J, Minari J, Huys I. The future regulation of artificial intelligence systems in healthcare services and medical research in the European Union. *Front Genet*. 2022;13:927721. doi: 10.3389/fgene.2022.927721
115. Matheny M, Israni ST, Ahmed M, Whicher D. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. In: Matheny M, Israni ST, Ahmed M, Whicher D, eds. National Academy of Medicine; 2019.
116. Pearson TA, Vitalis D, Pratt C, Campo R, Armoundas AA, Au D, Beech B, Brazhnik O, Chute CG, Davidson KW, et al. The science of precision prevention: research opportunities and clinical applications to reduce cardiovascular health disparities. *JACC Adv*. 2024;3:100759. doi: 10.1016/j.jacadv.2023.100759