Received: 26 October 2023 | Revised: 17 December 2023 | Accepted: 22 December 2023

DOI: 10.1002/pd.6522

REVIEW

PRENATAL DIAGNOSIS WILEY

# Improving prenatal diagnosis through standards and aggregation

Michael H. Duyzend[1,2,3] | Pilar Cacheiro[4] | Julius O. B. Jacobsen[4] | Jessica Giordano[5] | Harrison Brand[1,2,6] | Ronald J. Wapner[5] | Michael E. Talkowski[1,2,6,7,8] | Peter N. Robinson[9,10] | Damian Smedley[4]

[1]Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

[2]Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[3]Division of Genetics and Genomics, Department of Pediatrics, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[4]William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[5]Department of Obstetrics & Gynecology, Columbia University Medical Center, New York, New York, USA

[6]Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA

[7]Program in Biological and Biomedical Sciences, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts, USA

[8]Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, Massachusetts, USA

[9]The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA

[10]Institute for Systems Genomics, University of Connecticut, Farmington, Connecticut, USA

**Correspondence**
Damian Smedley, William Harvey Research Institute Queen Mary University of London Charterhouse Square, London EC1M 6BQ, UK.
Email: d.smedley@qmul.ac.uk

**Funding information**
National Institutes of Health, Grant/Award Numbers: 1F32HD112084, 1R01HD103805, 3R01HD055651-15S1, 5R01HD055651, 5R01HD105266, 5R24OD011883, 5RM1HG010860, UM1HG006370

## Abstract

Advances in sequencing and imaging technologies enable enhanced assessment in the prenatal space, with a goal to diagnose and predict the natural history of disease, to direct targeted therapies, and to implement clinical management, including transfer of care, election of supportive care, and selection of surgical interventions. The current lack of standardization and aggregation stymies variant interpretation and gene discovery, which hinders the provision of prenatal precision medicine, leaving clinicians and patients without an accurate diagnosis. With large amounts of data generated, it is imperative to establish standards for data collection, processing, and aggregation. Aggregated and homogeneously processed genetic and phenotypic data permits dissection of the genomic architecture of prenatal presentations of disease and provides a dataset on which data analysis algorithms can be tuned to the prenatal space. Here we discuss the importance of generating aggregate data sets and how the prenatal space is driving the development of interoperable standards and phenotype-driven tools.

Michael H. Duyzend and Pilar Cacheiro are contributed equally to this work.

**Key points**

**What is already known about this topic?**

- Data organization, homogenous processing, and aggregation are crucial for elucidating the genotype/phenotype relationship.
- Deep phenotyping improves molecular diagnosis.

**What does this review add?**

- Describes efforts and methods to increase prenatal data aggregation and organization.
- Discusses how the prenatal space is driving the development of interoperable standards and phenotype-driven tools.
- Emphasizes the importance and development of a cloud-based prenatal genotype-phenotype repository.

## 1 | INTRODUCTION

Until recently, the primary goal of prenatal diagnosis was to identify severe fetal disorders to provide information that families could use for medical decision making, including providing access to all reproductive options, and preparing for a potentially medically complex child. As newer molecular technologies such as genome and exome sequencing have become available for clinical use, more precise fetal precision medicine has become a reality.

The majority of sequencing studies to date have been performed on postnatal cohorts with specific disease phenotypes, and have begun to disentangle the genetic contributions to disorders such as autism spectrum disorder and developmental delay.[1–6] Little is known regarding phenotype-genotype correlations in the prenatal setting since most variant databases, including ClinVar, Human Gene Mutation Database, and gnomAD, do not have fetal specific annotations. The frequencies of many variants and related phenotypes discovered during fetal development are significantly less understood, and hard to capture and computationally compare to existing knowledge.[7–11] Second, the spectrum of prenatal phenotypes, even of Mendelian disorders with well-described postnatal phenotypes, is not well described and can represent a phenotype expansion.[12–15] Emerging evidence, for example, suggests that *SCN2A*, an epilepsy gene, can present prenatally with central nervous system structural anomalies.[16–18] Third, prenatal phenotypes are largely limited to structural findings with no understanding of functional or cognitive differences. Phenotypes also develop longitudinally during gestation, with certain features appearing, disappearing, or changing during pregnancy.[10,19] Fourth, prenatal phenotypes can be more challenging than postnatal phenotypes to categorize into disease-based cohorts, due to their focus on structural phenotypes or outcomes, including intrauterine fetal demise.[17,20,21] As many prenatal-sequencing studies are now performed on the trio of mother, father, and fetus, there is an additional opportunity to learn about the effects of maternal genetic variants on fetal health.[22–27] Furthermore, "prenatal specific" variation may be identified, that is, variants that only affect a fetus in utero of which phenotypic features are not observed postnatally due to prenatal lethality. Such variation is suggested to exist in an exome sequencing study of stillbirth.[28] Improving our understanding of the genetic and phenotypic architectures of perinatal development is crucial to optimizing patient care. Misinterpretation of the significance of a variant or failure to recognize a genetic condition can lead to misdiagnosis, an incorrect prognosis, inappropriate interventions, missed opportunities for treatment, and/or an unexpected recurrence. Improving understanding of perinatal genomic architecture will provide enhanced counseling on diagnosis, natural history, and potential therapeutic options for families.

## 2 | METHODS: HOW DO WE DO THIS?

Achieving understanding of the genetic and phenotypic architectures of the perinatal continuum will require the systematic organization and aggregation of data, and leveraging algorithms for data interpretation as is achieved for postnatal presentations of rare disease.[29,30] Many techniques, approaches, and tools are already established in the realm of postnatal Mendelian molecular diagnostics that prenatal experts can adopt and are discussed below.

### 2.1 | Data identification, generation, and collection

An initial step is the identification, generation, and collection of genetic and phenotype information. Genetic information can be categorized by its source and laboratory methodology. In prenatal diagnosis, various genetic assessments are used (e.g., chromosomal microarray analysis, sequencing panels, exome and genome sequencing) on various sample types (e.g., chorionic villi, amniotic fluid, cord blood). In conjunction, phenotype data are typically gathered, and a description is sent for analysis. Phenotype data collection is often structured and includes specific features such as gestational age and ultrasound findings, whereas clinical data is found unstructured in clinical notes, or in a short indication for clinical testing given to a laboratory (e.g., intrauterine fetal demise). Recognizing the importance of data aggregation and standardization, the Fetal Genomics Consortium (FGC), a consortium of centers striving to

understand the genetic and phenotypic architectures of prenatal presentations of disease, was established.[31] The FGC focuses on data generation and aggregation from prenatal sequencing consortia and technology development and has established an international repository of prenatal and genetic information. This genotype-phenotype repository is recognized by the Global Alliance for Genomics and Heath (GA4GH), a non-profit international consortium that is developing standards for responsibly collecting, storing, analyzing, and sharing genomic data. Research consortia, such as the Prenatal Genetic Diagnosis by Genomic Sequencing (PrenatalSEQ, NCT03936101) study in the ultrasonography (US) and Prenatal Assessment of Genomes and Exomes study in the UK, have generated a large amount of prenatal sequencing and phenotypic data, allowing for a base set of cases to demonstrate the value of aggregation and standardization of data.[32,33] Notably, data sharing may require data agreements, patient consent and institutional review board approval. Methods such as computerized consent and consent to share with specific prenatal resources can be incorporated into test requisition forms in collaboration with clinics and testing laboratories.[34–39] Given that the agreements can be challenging to arrange due to varied institutional processes, streamlining such processes can drastically reduce the logistical hurdles for data sharing. Finally, through resources such as GenomeConnect,[40] individuals can be directly engaged in collaboration with a sequencing laboratory.

## 2.2 | Data ingestion, structure, and interoperable standards

Clinical electronic health records (EHRs) are the primary source of clinical phenotype data and are used to store hospital data from imaging to clinical diagnoses, test orders, reports, medications, and billing. The data stored in these are extremely heterogeneous and rely on specialized messaging protocols such as HL7 V2 or HL7 Fast Healthcare Interoperability Resources.[41] While these are standards within the medical informatics communities, they are complex and not easily accessible to the research community.

Furthermore, the codified phenotype data within these systems are often unsuited for precision/deep phenotyping as they rely on broader disease-based coding systems such as ICD-10 or Systemized Nomenclature of Medicine. A major advance in the field of deep phenotyping came with the development of the Human Phenotype Ontology (HPO).[42] The HPO assigns each phenotype a unique identifier and defines relationships between phenotypes in a hierarchical manner. For example, the term for macrocephaly has, as parental and grandparental nodes, "increased head circumference" and "abnormality of skull size," respectively. While the HPO defines phenotypic features, not disease states, HPO terms are associated with disease terms in the Orphanet Rare Disease Ontology.[43] As of 2023-10-09, there are 18,052 terms, and 11,988 diseases with associated HPO terms.[44]

To address the community's needs for a standardized clinical case report supporting precision phenotypes, which could be independent of the EHR, the GA4GH developed the phenopacket Schema.[45] The phenopacket Schema specifies a portable, structured file analogous to high-throughput sequencing file formats, such as sequence alignment map and variant call format (VCF),[46] used in the genomics field. A phenopacket can include precision phenotypes, variants, disease diagnoses, measurements, biosample/biopsy data as well as therapeutic treatments and outcomes along with their evolution over time. The schema specifically added fields for prenatal temporality in the latest release, making it especially suited for use in the prenatal space. Phenopackets can interact with other standards for storing genetic alignment and variant level information (Table 1). A detailed practical introduction to the usage of phenopackets in the postnatal space has been previously published.[47]

Teams collaborating through GA4GH are creating a suite of interoperable standards around which an ecosystem of analysis tools can be created. Building on top of the Phenopacket and VCF standards, the Beacon protocol is an application programming interface that provides ways to structure and search on genomic and phenotypic datasets, without requiring access to raw data.[48,49] A specific variant or phenotype-based query can be made on a dataset organized to interact with Beacon, such as, "Is there a C at chromosome 13 at position 32,936,732?" with a "Y/N" answer. Different institutions can implement Beacon and create a Beacon Network, allowing federated search over multiple datasets. A federated network is a model where separate sites (hospital systems, laboratories, internal networks, etc.) are connected via a centralized

**TABLE 1** Selection of global alliance for genomics and heath standards for genotype and phenotype data.

| Standard | Description | Benefits/Limitations |
|---|---|---|
| Sequence alignment map (SAM) | Sequence reads aligned to a reference genome | Larger file size |
| Binary alignment map (BAM) | Sequence reads aligned to a reference genome (compressed) | Compressed |
| Compressed reference-oriented alignment map (CRAM) | Sequence reads aligned to a reference genome (compressed) | Highly compressed/requires reference |
| Variant call format (VCF) | Genomic variants and associated metadata | Easy to annotate/initially developed for small variants |
| Phenopacket | Schema for storing case level phenotype and genotype data | Stores all case information/software ecosystem in development |

management framework, using consistent configurations and policies. Such a network solves the problem of querying over multiple datasets across institutions (i.e., the data can be centrally queried but does not leave the owner's domain). The current version of the Beacon protocol can return data consent type in the returned query, in case additional or raw data would like to be pursued for additional studies. Establishing a prenatal Beacon network is a way to unite prenatal genotype/phenotype data around the globe.

### 2.3 | Cloud based data storage and processing infrastructures

Using standardized algorithms and pipelines allows for robust statistical analyses and reduces the sources of error. Cloud-based computing is a scalable way to process, analyze, and share data.

The three major providers of cloud computing infrastructure are Google (Google Cloud Platform [GCP]), Microsoft (Azure), and Amazon (Amazon Web Services). All provide similar products with different implementations, with efforts in the genomics community for interoperability between these standards.[50]

Recognizing the benefits of cloud computing, the Broad Institute developed Terra, a bioinformatics platform built on GCP and Azure, that facilitates running genomics workflows on the cloud,[51] with relative ease. Terra is part of the National Human Genome Research Institute's Genomic Data Science Analysis, Visualization, and Informatics Lab-space, and is approved data sharing platform through the National Institutes of Health (NIH) Genomic Data Sharing policies.[52] Access controls set through Terra allow for facile data sharing and federation of datasets, and Terra incorporates standards for robust data security and is certified by the Federal Risk and Authorization Management Program.

### 2.4 | Phenotype driven variant prioritization

Given the overwhelming number of detected variants in genome-wide sequencing data, variants must be annotated and filtered for which a host of tools exist.[53,54] Variants can be filtered on their intrinsic characteristics such as consequence (e.g., loss of function), in silico prediction of pathogenicity, absence or low minor allele frequency in population databases (e.g., gnomAD), segregation with disease and known inheritance (e.g., de novo), and classification in variant databases (e.g., ClinVar).[7,8] The remaining 10–100 s of variants are reviewed, requiring significant person-time. Computational methods to prioritize variants are critical, particularly when rapid results are necessary for appropriate care options.

Multiple software tools for phenotype-driven variant prioritization have been developed.[55–57] However, there are substantial differences in terms of accessibility/usability, for example, with regard to whether the software can be downloaded locally or is web-based, input file types and whether the algorithms are proprietary or free to use.[58] Tools that are freely available, that allow for local programmatic access, and accept VCF files and HPO encoded phenotypes include: Exomiser,[59] LIkelihood Ratio Interpretation of Clinical AbnormaLities,[60] PhenIX[61] and Xrare.[62] Several other tools exist, but the majority are either web-based, which can raise privacy concerns, or have not undergone updates in recent years.[58] Exomiser is in use by the UK's Genomic medicine service, and is discussed further here. Exomiser uses standardized phenotypes, for example, HPO terms encoded in a Phenopacket, to prioritize certain variants by combining variant filtering and phenotype profile matching.[57] Semantic similarity calculations are performed between the patient's phenotypes and existing phenotypes described for either: (i) known human gene-disease associations, (ii) model organism (mouse and zebrafish) gene orthologs, and/or (iii) neighbor genes in a protein-protein interaction network to find a possible match. The use of phenotypic associations from model organisms and protein interacting pairs enables phenotype-based strategies not only for known disease-phenotype correlations but also for novel disease gene discovery.[57] The semantic similarity algorithms for ontology-based profile matching incorporate information on the distance between terms in the ontology and their frequency to compare exact or similar phenotypes (Figure 1). Cross-species comparisons are possible by aligning individual disease phenotypes (HPO terms) and individual mouse phenotypes (Mammalian Phenotype [MP] terms).[64] Pairwise phenotype calculations are ultimately aggregated into a single score.[65,66]

Such an approach has already been piloted on large genomic datasets with comprehensive phenotyping, including the 100,000 Genomes Project (100KGP) in the UK and is now used in mainstream healthcare as part of the UK's Genomic Medicine Service.[29,59] In this study, 88% of the diagnoses were detected by Exomiser in the top five prioritized candidates using a fully automated process that took less than 5 min to run. These results were fed back to the clinical geneticists in the diagnostic laboratories, allowing efficient case interpretation. In contrast, restricting analysis to a curated virtual panel of genes known to be associated with the recruited disease category detected 54% of the diagnoses, albeit with higher precision. The precision phenotyping of patients performed in the 100KGP allowed the automated Exomiser approach as well as the inclusion of additional virtual disease gene panels in the panel-based approach that increased detection there to 77% of diagnoses. Indeed, clinically collected phenotype information can be sparse, and the value of algorithms increases with (i) large datasets; and (ii) precision phenotypes, highlighting the need for robust methods to collect detailed phenotypic information.

## 3 | RESULTS

### 3.1 | The prenatal space is driving development of interoperable standards and phenotype-driven tools out of necessity

The prenatal space is ideal for the development and tuning of tools for gene discovery, phenotype driven variant prioritization, and

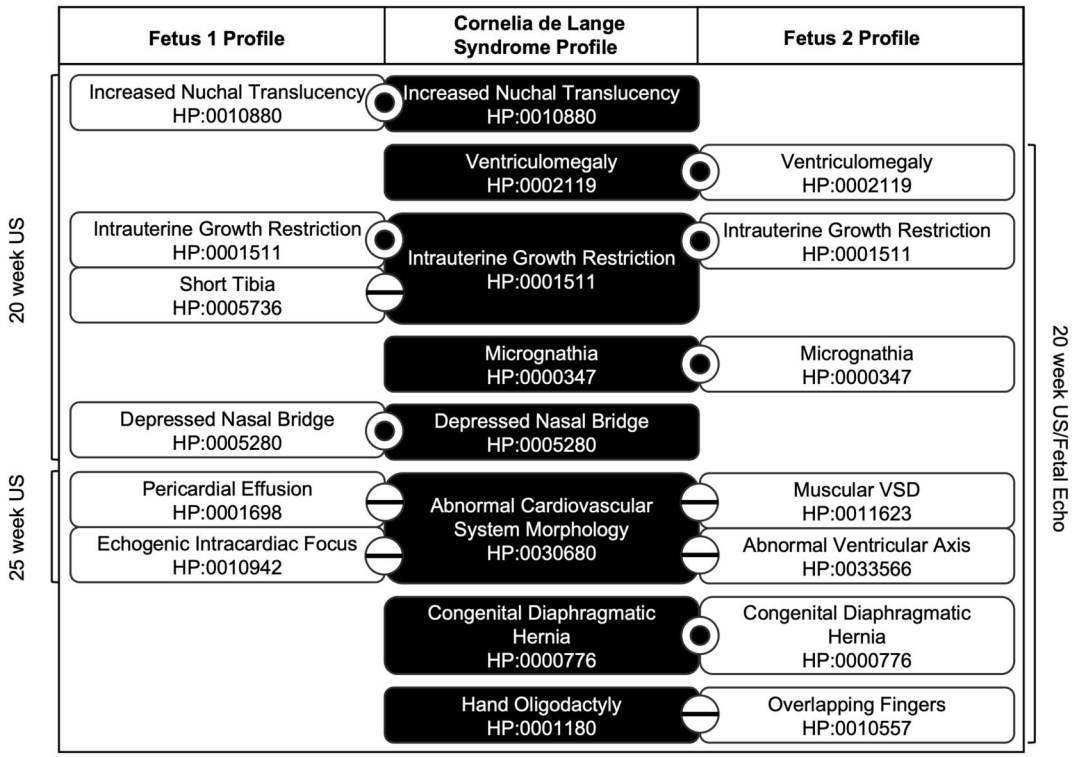


**FIGURE 1** Semantic search algorithms improve with comprehensive phenotypes: Prenatal phenotypes found in two fetuses diagnosed postnatally with Cornelia de Lange syndrome and matching phenotypes associated with the diagnosis. Semantic search algorithms attempt to match phenotypic feature profiles (encoded as terms from ontologies) from patients to phenotypic profiles associated with a particular disease. A growing amount of comprehensive prenatal phenotype-gene-disease correlation data will increase the sensitivity of these algorithms and permit the training of machine learning models. Cases adapted from Clark, et al.[63] Circles with dots represent perfect HPO term matches, and circles with lines represent partial HPO matches. HPO, human phenotype ontology.

developing a suite of tools built on standards. While progression towards this vision is generally accepted in the genomics community, it is essential in the prenatal space for several reasons. First, while large sequencing studies have been performed on postnatal cohorts, the coming years will bring a commensurate deluge of prenatal sequencing data, facilitated by the dropping cost of genomic sequencing, development of robust genome-wide cell-free DNA sequencing methods, and the cohesion of the prenatal community around the importance of genomics in patient care.[67–72] Second, the relationship between genotype and prenatal phenotype is not well understood and there is an opportunity to define the spectrum of prenatal phenotypes that exist for molecular diagnoses. This effort will drive gene discovery, given that molecular diagnoses are not identified in many prenatal cases, likely due to the lack of extensive sequencing to characterize "prenatal-specific" variation not observed postnatally. Third, phenotypes can change during gestation, and recording of phenotypes longitudinally is crucial, for example, using the Phenopackets schema. Fourth, given the rapid development of prenatal sequencing projects and consortia, there is an opportunity to implement and develop standards and pipelines from the beginning. Fifth, the unique nature of the prenatal space gives the opportunity to tune variant interpretation algorithms directly towards prenatal diagnosis.

## 3.2 | Prenatal additions to the HPO

One tangible example is that the prenatal community has contributed to the expansion of the HPO. While there are thousands of terms to describe items that can be seen pre- and postnatally, it was recognized that certain terms describing phenotypes specific to the prenatal period were missing from the HPO, and some existing terms could use revised definitions. Over the course of three years, from 2020 to 2023, the community added 95 terms to the *abnormality of prenatal development or birth* (HP:0001197) sub-hierarchy, and revised definitions, synonyms, and disease associations for the majority of the existing 152 terms, for a total of 247 terms, or a 62% increase.[73] One example of a term added is Lemon Sign (HP: 0032269), referring to the shape of the fetal skull at US, when the frontal bones lose their normal convex contour and appear flattened or inwardly scalloped. Terms were added to specify the time of onset of a particular phenotype, with new terms added to describe embryonal, late first trimester, second trimester, and third trimester onset. An example of HPO terms related to a particular case is shown in Figure 2.[74] It is estimated that the current HPO covers roughly 85% of the terms required for comprehensive annotation, and efforts continue to revise and augment the HPO.

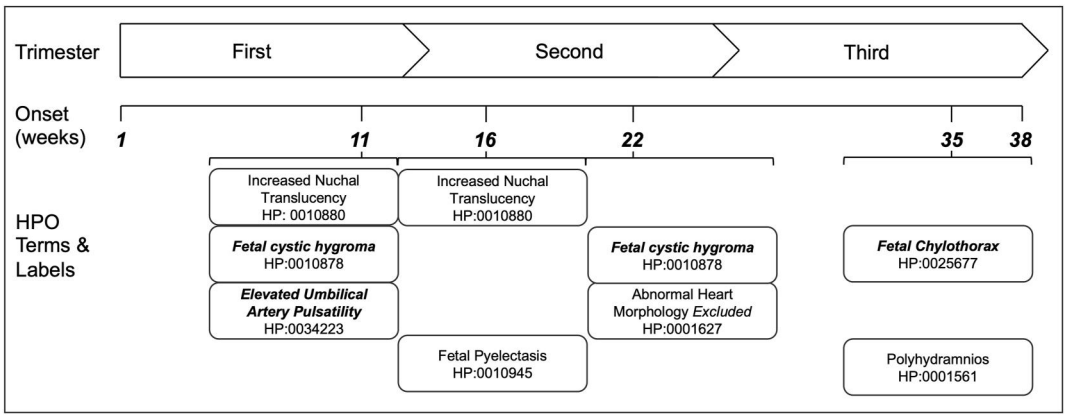


**FIGURE 2** Expansion of the prenatal HPO: Since 2020, over 90 new prenatal-relevant terms have been added to the HPO. This is demonstrated in the evolving fetal phenotype of a fetus with a *PTPN11* variant, adapted from Malniece, et al.[74] Newly added HPO terms, including onset terms, are bolded and italicized. HPO, human phenotype ontology.

### 3.3 | Standardization of prenatal phenotyping

While like any phenotype, there can be variation in prenatal phenotyping; considerable efforts to standardize the collection have been developed and are underway. The International Society of Ultrasound in Obstetrics & Gynecology (ISUOG) has developed practice guidelines for ultrasound and fetal magnetic resonance.[75,76] Most US scans and reports contain similar fields, and work is progressing to extract data directly from US machines and reports to a computable format, such as a phenopacket. Artificial Intelligence approaches for interpretation and collection of US data, while not yet standard of care, have great potential.[77] These approaches can help define optimal ultrasound windows and allow direct identification and interpretation of anomalies. While these efforts will help, the current system of referrals to higher-level centers will identify and define most prenatal structural anomalies.

### 3.4 | Model organisms can help with (human) variant interpretation/gene prioritization in the prenatal space

The wealth of data surrounding phenotypes in model organisms, provides the opportunity to improve variant interpretation and gene prioritization and discovery in the prenatal space. The analogous ontology for mammalian (mainly mouse) phenotype data is called the Mammalian Phenotype Ontology (MPO) and contains 14,273 MP terms (Ref.[78]; release 2023-09-11).[79] Like the HPO "Abnormality of prenatal development or birth (HP:0001197) grouping term, the MPO includes an "embryo phenotype" (MP:0005380) parental term that group terms describing abnormalities of mouse embryo development, embryo morphology and physiology. Phenotypes have been assessed and made available by the International Mouse Phenotyping Consortium (IMPC), a systematic, high-throughput effort to generate and characterize knocked out mice,[80] as well as the Mouse Genome Informatics (MGI) database, a resource providing integrated genetic, genomic, and biological data from literature curation especially.[81] Combining data from the IMPC (DR 19.1) and MGI resources (Data accessed 2023-10-12), 2215 mouse genes have associated abnormal phenotypes under this grouping term. Most of these embryonic phenotypes (2105; 95%) correspond to lines associated with preweaning lethal phenotypes. Additionally, the term "growth/size/body region phenotype" contains terms related to abnormal embryo size and growth. Comparing the prenatal gene-phenotype associations between the MP and HPO, only 402 of these 2215 mouse genes have a human ortholog with prenatal annotations under the grouping terms "Abnormality of prenatal development or birth" and "Intrauterine growth retardation". Despite certain limitations in the ability of mouse models to mimic phenotypes observed in humans, this would imply a large number of human genes with unannotated prenatal phenotypes.[82–84] This has particular implications for understanding the genetics of intrauterine fetal demise, for which few large-scale studies have been performed.[85–87] Importantly, and given the nature of the standardized phenotyping protocol followed by the IMPC, the homozygous knockouts resulting in complete or incomplete penetrance lethal phenotypes and with embryonic abnormalities described, have also undergone an early adult phenotyping screen for the corresponding heterozygous model, which permits establishing correlations between prenatal and postnatal phenotypes.[80] An important limitation, however, is the scarce embryonic data for viable lines, where the phenotypic screens are performed after birth, and hence the difficulty in comparing prenatal and postnatal phenotypes for the same genotype. An example of an IMPC mouse knockout with phenotypic information available for different life stages is shown in Figure 3.

Mendelian genes are significantly overrepresented among knockout mouse orthologs with a lethal phenotype. Consequently, this set of mouse lethal genes with no known association to human disease has been suggested as a compelling source of potential candidate genes[88] and different strategies combining data on mouse viability with other sources of evidence have been successful in identifying novel genes associated with neurodevelopmental

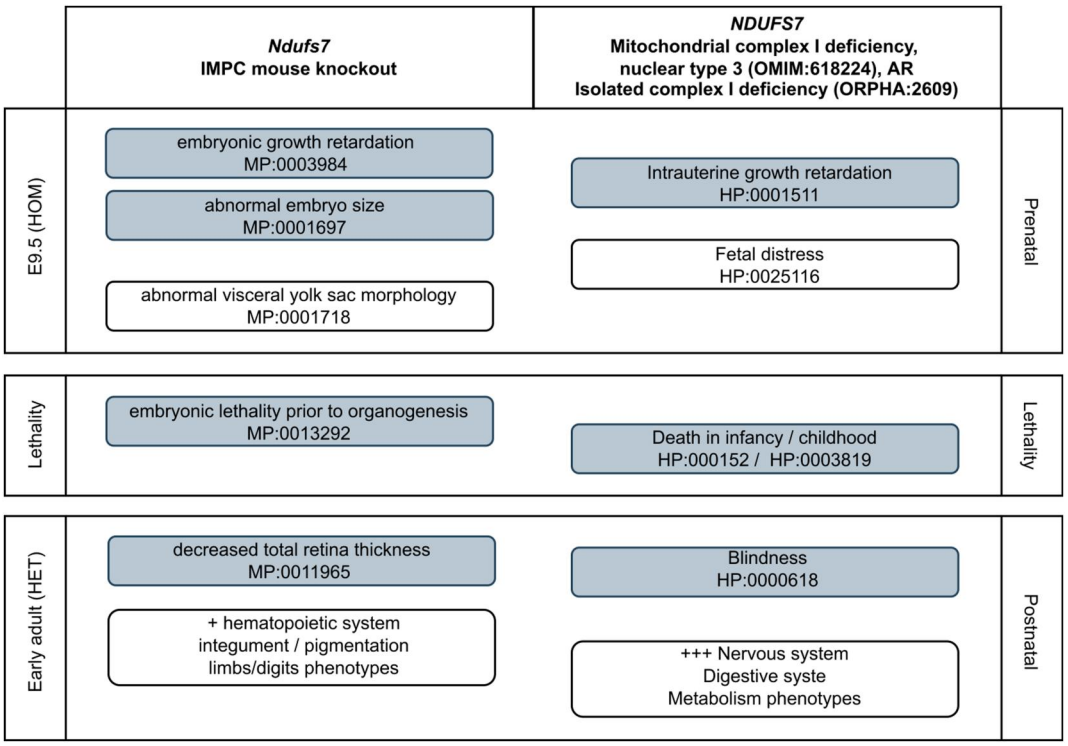


**FIGURE 3** Summary of prenatal and adult mouse and human phenotypes associated with pathogenic variation in the *Ndufs7/NDUFS7* gene. Shown are the abnormal phenotypes associated with the homozygous and heterozygous mouse *Ndufs7* gene knockout from the IMPC at different life stages and some of the overlapping phenotypes/physiological systems reported for the autosomal recessive disorder associated with the corresponding human gene ortholog. HOM: homozygote, HET: heterozygote; AR: autosomal recessive; E: embryonic day; IMPC, International mouse phenotyping consortium. (IMPC DR19.1; HPO v2023-10-09). [Colour figure can be viewed at wileyonlinelibrary.com]

conditions.[89] Leveraging these gene sets in analysis may assist in discovering genes important to perinatal development.

Furthermore, there is an effort to create a catalog of lethal phenotypes in humans, including curation and collation of the (earliest) age of death reported for known Mendelian disorders.[90] In its current version (2023-09-27), it contains 959 genes associated with prenatal, neonatal, or infancy death based on Online Mendelian Inheritance in Man clinical records. Additional integration of both validated and candidate variants/genes identified through sequencing studies of pregnancy loss, fetal and perinatal lethal manifestations, may reveal novel disease-associated genes along with expansions of the clinical spectrum for known Mendelian genes.

### 3.5 | Building a longitudinal prenatal phenotype/genotype resource

It was recognized that a resource that unites, standardizes, and aggregates genomic and phenotypic studies in the prenatal sphere is essential. The FGC has piloted a cloud-based genotype-phenotype repository built on the bioinformatics platform Terra, called the Repository of the International Fetal Genomics Consortium (RIFGC).

Major goals of the RIFGC are to use GA4GH interoperable standards, create standardized pipelines for genotype/phenotype analysis, and to allow the federation of datasets in the cloud. There is already a host of genotype processing pipelines available on Terra, including those using the Genome Analysis Toolkit[91] for small variant (single nucleotide variant/indel) as well as structural variant discovery.[6,92] Phenopackets have been implemented to allow the integration of phenotype-driven analysis into pipelines. Search over phenotypes and prioritized variants stored in phenopackets has been implemented on Terra using the Beacon v2 protocol (Figure 4).

Governance and data organization are key to this enterprise. For data contributors with data on premise servers, data are uploaded to a cloud storage bucket associated with a particular workspace in Terra. The contributor has full access to their data on the cloud, and access permissions are given to the RIFGC. These raw data are then processed identically between contributors, to create a homogenous dataset for analysis and clinical inquiry. The RIFGC was selected as a Driver Project by GA4GH for the recognition of the enterprise to drive the development and implementation of GA4GH standards important to the prenatal and genomic communities.[93]

## 4 | DISCUSSION

### 4.1 | Bringing it all together and next steps

It is crucial that genomic and phenotypic data utility is maximized because it can translate directly to improved patient care. There will

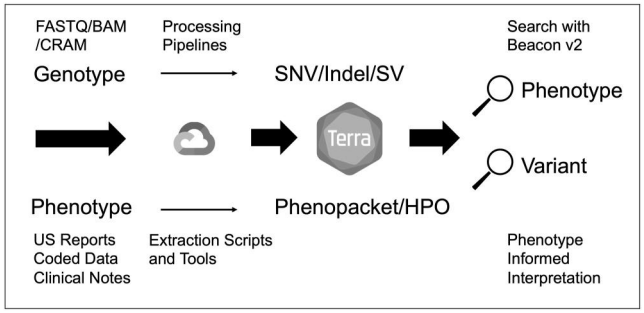


**FIGURE 4** Workflow of the repository of the international fetal genomics consortium. Genotype and phenotype data are uploaded to the cloud and access is given to the repository. For exome and genome sequencing data, standardized genomic pipelines generate small and structural variant callsets. Phenotype information is extracted and used to populate a Phenopacket. Genomic variants are then filtered and interpreted and associated with a Phenopacket. The Beacon v2 application programming interface permits search by phenotype or variant and the possibility to connect to additional prenatal data and other databases through the interface.

soon be a tsunami of additional prenatal genomic data available as sequencing costs decrease and comprehensive prenatal cell-free DNA sequencing approaches develop.[67,69] There are five primary areas of need, summarized below:

*Phenotype Data Ingestion/Standardization of Collection:* While there are established and emerging standards for genotype and phenotype data, a major challenge remains to convert data to a comparable form. This can be achieved by communities agreeing on a set of algorithms for standard processing. For phenotype integration, continued developments to extract data directly from clinical notes and ultrasound machines are needed as is a standard data format (such as a phenopacket). We envision that the community will establish standards for prenatal clinical phenotype collection.

*Longitudinal Genomic Accessibility*: Understanding prenatal presentations of postnatal phenotypes is crucial, though relies on longitudinal data collection. Fetal phenotype information is found in the mother's chart, whereas pediatric data is found in a child's chart. For liveborn individuals, linking this information is essential to understanding developmental trajectory and differences in phenotype. In individuals with genetic disease without a strong prenatal phenotype, understanding the postnatal phenotype can provide clues for features to look for prenatally. This will be increasingly critical as sequencing expands for use in pregnancies without a clear phenotype, allowing for a genotype first approach and requiring ample data about prognosis for counseling. With increased comprehensive sequencing performed prenatally, the ability to use this information across the lifespan is critical. Guidelines must continue to be developed for which genomic testing results are returned and when, as well as when genome reanalysis should occur, for example, due to emerging phenotypes present postnatally that were not present or appreciated prenatally.

*Federation/Aggregation:* Aggregating and federating a large amount of data is important to gain statistical insights into the genetic and phenotypic architecture of prenatal presentation of disease. The RIFGC is a way to unite data across space and time. While the technical hurdles of federating data are surmountable, the logistical aspects can be challenging. First, consent are different by institution and there are different legal frameworks by jurisdiction. Working towards developing computerized consent in collaboration with the sequencing laboratory from the first patient encounter, with the possibility of engaging in data sharing with a resource like the RIFGC on a testing request form, would be steps in this direction. Second, sharing data between institutions requires data use agreements to protect data provenance and to ensure ethical conduct. Easing methods of sharing across institutions without months of agreement negotiation would lower the logistical barrier of federation and build crucial resources more readily. Third, to improve genome accessibility and patient use of data, allowing individuals to share data directly to a central repository increases participant numbers and allows direct individual contribution to understanding their genomic data. It should be noted that in certain jurisdictions, a laboratory may own patient-derived laboratory materials (e.g., sequencing libraries) and genomic data, thus necessitating agreements with entities tasked with generation and preservation of patient data in addition to patient consent.

*Development and Tuning of Algorithms and Resources Specific to the Prenatal Space:* The prenatal space offers numerous opportunities to develop and tune algorithms, given sufficient data. For example, statistical models, such as the logistic regression model used in Exomiser, can be tuned to the prenatal space. However, there is a scarcity of human prenatal genotype to phenotype knowledge, which hinders these approaches. The ongoing curation of this knowledge and use of model organism embryonic data will address this challenge. With large amounts of standardized and comprehensive phenotype information, machine learning algorithms could be utilized to predict prenatal features from known postnatal phenotypes.

### 4.2 | Provider tools

Ultimately, these resources need to be accessible to providers in the clinic. Efforts are made to develop a web-portal to search over all cases for phenotype (e.g., return all cases at 20 weeks with congenital heart defects) and specific variant information within the RIFGC repository. In conjunction, a growing catalog of lethal phenotypes in humans has been generated, including the collation of the (earliest) age of death reported for known Mendelian disorders. This curation effort will grow as more stillbirth cases are sequenced and incorporated into the RIFGC and as mouse models are generated.

## 5 | CONCLUSION

The rapid advances in sequencing technology, next-generation imaging techniques, and cloud-based computing set the stage for transformative insights into the genomic underpinnings of the

prenatal presentation of disease. This effort relies on the collection, aggregation and homogenization of both comprehensive phenotype and genetic data. Specific to these efforts is the need to capture data longitudinally across the perinatal continuum. Such data will spur the development of data processing algorithms and enable machine learning applications specific to the prenatal space. These community efforts will not only inform how genetics impacts prenatal diagnosis, but also paves the way for improved counseling and intervention and ultimately enhanced patient care.

## ACKNOWLEDGMENT

This work was supported by grants from the National Institutes of Health (1F32HD112084, 5R01HD105266, 5R01HD055651, 3R01 HD055651-15S1, 5RM1HG010860, 5R24OD011883, 1R01HD10 3805, UM1HG006370).

## CONFLICT OF INTEREST STATEMENT

M.E.T. and H.B. received research funding from Microsoft Inc. R.J.W., M.E.T. and H.B. received reagents from Illumina Inc.

## DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID

*Michael H. Duyzend* https://orcid.org/0000-0002-3941-6447
*Pilar Cacheiro* https://orcid.org/0000-0002-6335-8208
*Julius O. B. Jacobsen* https://orcid.org/0000-0002-3265-1591
*Jessica Giordano* https://orcid.org/0000-0002-2265-0691
*Harrison Brand* https://orcid.org/0000-0002-7901-0812
*Ronald J. Wapner* https://orcid.org/0000-0003-0727-1244
*Michael E. Talkowski* https://orcid.org/0000-0003-2889-0992
*Peter N. Robinson* https://orcid.org/0000-0002-0736-9199
*Damian Smedley* https://orcid.org/0000-0002-5836-9850

## REFERENCES

1. Wright CF, Campbell P, Eberhardt RY, et al. Genomic diagnosis of rare pediatric disease in the United Kingdom and Ireland. *N Engl J Med*. 2023;388(17):1559-1571. https://doi.org/10.1056/nejmoa2209046
2. Fu JM, Satterstrom FK, Peng M, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat Genet*. 2022;54(9):1320-1331. https://doi.org/10.1038/s41588-022-01104-0
3. Zhou X, Feliciano P, Shu C, et al. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat Genet*. 2022;54(9):1305-1319. https://doi.org/10.1038/s41588-022-01148-2
4. Wilfert AB, Turner TN, Murali SC, et al. Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat Genet*. 2021;53(8):1125-1134. https://doi.org/10.1038/s41588-021-00899-8
5. Warrier V, Zhang X, Reed P, et al. Genetic correlates of phenotypic heterogeneity in autism. *Nat Genet*. 2022;54(9):1293-1304. https://doi.org/10.1038/s41588-022-01072-5
6. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature*. 2020;581(7809):444-451. https://doi.org/10.1038/s41586-020-2287-8
7. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. https://doi.org/10.1038/s41586-020-2308-7
8. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067. https://doi.org/10.1093/nar/gkx1153
9. Stenson PD, Mort M, Ball EV, et al. The human gene mutation database (HGMD®): optimizing its use in a clinical diagnostic or research setting. *Hum Genet*. 2020;139(10):1197-1207. https://doi.org/10.1007/s00439-020-02199-3
10. Mone F, Abu Subieh H, Doyle S, et al. Evolving fetal phenotypes and clinical impact of progressive prenatal exome sequencing pathways: cohort study. *Ultrasound Obstet Gynecol*. 2022;59(6):723-730. https://doi.org/10.1002/uog.24842
11. Gray KJ, Wilkins-Haug LE, Herrig NJ, Vora NL. Fetal phenotypes emerge as genetic technologies become robust. *Prenat Diagn*. 2019;39(9):811-817. https://doi.org/10.1002/pd.5532
12. Brabbing-Goldstein D, Kozlova D, Bazak L, et al. Unique prenatal manifestations of biallelic NDUFAF5 variants: expansion of the phenotype. *Ultrasound Obstet Gynecol*. 2023. [Internet]. https://doi.org/10.1002/uog.27482
13. Fu F, Li R, Yu Q, et al. Application of exome sequencing for prenatal diagnosis of fetal structural anomalies: clinical experience and lessons learned from a cohort of 1618 fetuses. *Genome Med*. 2022;14(1):123. https://doi.org/10.1186/s13073-022-01130-x
14. Saini N, Venkatapuram VS, Vineeth VS, et al. Fetal phenotypes of Mendelian disorders: a descriptive study from India. *Prenat Diagn*. 2022;42(7):911-926. https://doi.org/10.1002/pd.6172
15. Giuffrida MG, Mastromoro G, Guida V, et al. A new case of SMABF2 diagnosed in stillbirth expands the prenatal presentation and mutational spectrum of ASCC1. *Am J Med Genet*. 2020;182(3):508-512. https://doi.org/10.1002/ajmg.a.61431
16. Bernardo S, Marchionni E, Prudente S, et al. Unusual association of SCN2A epileptic encephalopathy with severe cortical dysplasia detected by prenatal MRI. *Eur J Paediatr Neurol*. 2017;21(3):587-590. https://doi.org/10.1016/j.ejpn.2017.01.014
17. Petrovski S, Aggarwal V, Giordano JL, et al. Whole-exome sequencing in the evaluation of fetal structural anomalies: a prospective cohort study. *Lancet*. 2019;393(10173):758-767. https://doi.org/10.1016/s0140-6736(18)32042-7
18. Lei TY, She Q, Fu F, et al. Prenatal exome sequencing in fetuses with callosal anomalies. *Prenat Diagn*. 2022;42(6):744-752. https://doi.org/10.1002/pd.6107
19. Chandler NJ, Scotchman E, Mellis R, Ramachandran V, Roberts R, Chitty LS. Lessons learnt from prenatal exome sequencing. *Prenat Diagn*. 2022;42(7):831-844. https://doi.org/10.1002/pd.6165
20. Mellis R, Oprych K, Scotchman E, Hill M, Chitty LS. Diagnostic yield of exome sequencing for prenatal diagnosis of fetal structural anomalies: a systematic review and meta-analysis. *Prenat Diagn*. 2022;42(6):662-685. https://doi.org/10.1002/pd.6115
21. Lowther C, Valkanas E, Giordano JL, et al. Systematic evaluation of genome sequencing for the diagnostic assessment of autism spectrum disorder and fetal structural anomalies. *Am J Hum Genet*. 2023;110(9):1454-1469. https://doi.org/10.1016/j.ajhg.2023.07.010
22. Solé-Navais P, Flatley C, Steinthorsdottir V, et al. Genetic effects on the timing of parturition and links to fetal birth weight. *Nat Genet*. 2023;55(4):559-567. https://doi.org/10.1038/s41588-023-01343-9
23. Chen J, Bacelis J, Sole-Navais P, et al. Dissecting maternal and fetal genetic effects underlying the associations between maternal phenotypes, birth outcomes, and adult phenotypes: a mendelian-

randomization and haplotype-based genetic score analysis in 10,734 mother-infant pairs. *PLoS Med*. 2020;17(8):e1003305. https://doi.org/10.1371/journal.pmed.1003305
24. Mardy AH, Chetty SP, Norton ME. Maternal genetic disorders and fetal development. *Prenat Diagn*. 2020;40(9):1056-1065. https://doi.org/10.1002/pd.5659
25. Zhang G, Feenstra B, Bacelis J, et al. Genetic associations with gestational duration and spontaneous preterm birth. *N Engl J Med*. 2017;377(12):1156-1167. https://doi.org/10.1056/nejmoa1612665
26. Warrington NM, Beaumont RN, Horikoshi M, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet*. 2019;51(5):804-814. https://doi.org/10.1038/s41588-019-0403-1
27. Beaumont RN, Flatley C, Vaudel M, et al. Genome-wide association study of placental weight identifies distinct and shared genetic influences between placental and fetal growth. *Nat Genet*. 2023;55(11):1807-1819. [Internet]. https://doi.org/10.1038/s41588-023-01520-w
28. Stanley KE, Giordano J, Thorsten V, et al. Causal genetic variants in stillbirth. *N Engl J Med*. 2020;383(12):1107-1116. https://doi.org/10.1056/nejmoa1908753
29. Smedley D, Smedley D, Smith KR, et al. 100,000 genomes pilot on rare-disease diagnosis in health care - preliminary report. *N Engl J Med*. 2021;385(20):1868-1880. https://doi.org/10.1056/nejmoa2035790
30. Splinter K, Adams DR, Bacino CA, et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. *N Engl J Med*. 2018;379(22):2131-2139. https://doi.org/10.1056/nejmoa1714458
31. Fetal genomics consortium [Internet]. Accessed 2023 Oct 26. https://www.fetalgenomics.org/
32. Lord J, McMullan DJ, Eberhardt RY, et al. Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. *Lancet*. 2019;393(10173):747-757. https://doi.org/10.1530/ey.16.14.14
33. Kernie CG, Wynn J, Rosenbaum A, et al. Information is power: the experiences, attitudes and needs of individuals who chose to have prenatal genomic sequencing for fetal anomalies. *Prenat Diagn*. 2022;42(7):947-954. https://doi.org/10.1002/pd.6153
34. Chimonas S, Lipitz-Snyderman A, Matsoukas K, Kuperman G. Electronic consent in clinical care: an international scoping review. *BMJ Health Care Inf*. 2023;30(1):e100726. [Internet]. https://doi.org/10.1136/bmjhci-2022-100726
35. Gainotti S, Turner C, Woods S, et al. Improving the informed consent process in international collaborative rare disease research: effective consent for effective research. *Eur J Hum Genet*. 2016;24(9):1248-1254. https://doi.org/10.1038/ejhg.2016.2
36. Marshall PA, Adebamowo CA, Adeyemo AA, et al. Voluntary participation and informed consent to international genetic research. *Am J Publ Health*. 2006;96(11):1989-1995. https://doi.org/10.2105/ajph.2005.076232
37. Staunton C, Slokenberga S, Mascalzoni D. The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks. *Eur J Hum Genet*. 2019;27(8):1159-1167. https://doi.org/10.1038/s41431-019-0386-5
38. Courbier S, Dimond R, Bros-Facer V. Share and protect our health data: an evidence based approach to rare disease patients' perspectives on data sharing and data protection - quantitative survey and recommendations. *Orphanet J Rare Dis*. 2019;14(1):175. https://doi.org/10.1186/s13023-019-1123-4
39. Hartman AL, Hechtelt Jonker A, Parisi MA, Julkowska D, Lockhart N, Isasi R. Ethical, legal, and social issues (ELSI) in rare diseases: a landscape analysis from funders. *Eur J Hum Genet*. 2020;28(2):174-181. https://doi.org/10.1038/s41431-019-0513-3
40. Savatt JM, Azzariti DR, Faucett WA, et al. ClinGen's GenomeConnect registry enables patient-centered data sharing. *Hum Mutat*. 2018;39(11):1668-1676. https://doi.org/10.1002/humu.23633
41. Index - FHIR v5.0.0 [Internet]. Accessed 2023 Oct 26. https://www.hl7.org/fhir/
42. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83(5):610-615. https://doi.org/10.1016/j.ajhg.2008.09.017
43. Pavan S, Rommel K, Mateo Marquina ME, Höhn S, Lanneau V, Rath A. Clinical practice guidelines for rare diseases: the Orphanet database. *PLoS One*. 2017;12(1):e0170365. https://doi.org/10.1371/journal.pone.0170365
44. Human phenotype ontology [Internet]. Accessed 2023 Oct 26. https://hpo.jax.org/app/data/ontology
45. Jacobsen JOB, Baudis M, Baynam GS, et al. The GA4GH Phenopacket schema defines a computable representation of clinical data. *Nat Biotechnol*. 2022;40(6):817-820. https://doi.org/10.1038/s41587-022-01357-4
46. HTS format specifications [Internet]. Accessed 2023 Oct 26. https://samtools.github.io/hts-specs/
47. Ladewig MS, Jacobsen JOB, Wagner AH, et al. GA4GH phenopackets: a practical introduction. *Adv Genet*. 2023;4(1):2200016. https://doi.org/10.1002/ggn2.202200016
48. Rueda M, Ariosa R, Moldes M, Rambla J. Beacon V2 Reference Implementation: a Toolkit to enable federated sharing of genomic and phenotypic data. *Bioinformatics*. 2022:btac568.
49. Rambla J, Baudis M, Ariosa R, et al. Beacon v2 and Beacon networks: a "lingua franca" for federated data discovery in biomedical genomics, and beyond. *Hum Mutat*. 2022;43(6):791-799. https://doi.org/10.1002/humu.24369
50. NIH cloud platform interoperability (NCPI) effort [Internet]. Accessed 2023 Oct 26. https://www.ga4gh.org/driver_project/nih-cloud-platform-interoperability-ncpi-effort/
51. Terra [Internet]. Accessed 2023 Oct 26. https://app.terra.bio/
52. Schatz MC, Philippakis AA, Afgan E, et al. Inverting the model of genomics data sharing with the NHGRI genomic data science analysis. *Visualization, and Informatics Lab-space. Cell Genom [Internet]*. 2022;2(1):100085. https://doi.org/10.1016/j.xgen.2021.100085
53. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17(1):122. https://doi.org/10.1186/s13059-016-0974-4
54. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164. https://doi.org/10.1093/nar/gkq603
55. Yuan X, Wang J, Dai B, et al. Evaluation of phenotype-driven gene prioritization methods for Mendelian diseases. *Briefings Bioinf*. 2022;23(2):1-12. [Internet]. https://doi.org/10.1093/bib/bbac019
56. Jacobsen JOB, Kelly C, Cipriani V, Robinson PN, Smedley D. Evaluation of phenotype-driven gene prioritization methods for Mendelian diseases. *Brief Bioinform [Internet*. 2022;23(5):1-3. https://doi.org/10.1093/bib/bbac188
57. Jacobsen JOB, Kelly C, Cipriani V, et al. Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease. *Hum Mutat*. 2022;43(8):1071-1081. https://doi.org/10.1002/humu.24380
58. Kelly C, Szabo A, Pontikos N, et al. Phenotype-aware prioritisation of rare Mendelian disease variants. *Trends Genet*. 2022;38(12):1271-1283. https://doi.org/10.1016/j.tig.2022.07.002
59. Smedley D, Jacobsen JOB, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015;10(12):2004-2015. https://doi.org/10.1038/nprot.2015.124
60. Robinson PN, Ravanmehr V, Jacobsen JOB, et al. Interpretable clinical genomics with a likelihood ratio paradigm. *Am J Hum Genet*. 2020;107(3):403-417. https://doi.org/10.1016/j.ajhg.2020.06.021
61. Zemojtel T, Köhler S, Mackenroth L, et al. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med*. 2014;6(252):252ra123. https://doi.org/10.1126/scitranslmed.3009262

62. Li Q, Zhao K, Bustamante CD, Ma X, Wong WH. Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. *Genet Med*. 2019;21(9):2126-2134. https://doi.org/10.1038/s41436-019-0439-8
63. Clark DM, Sherer I, Deardorff MA, et al. Identification of a prenatal profile of Cornelia de Lange syndrome (CdLS): a review of 53 CdLS pregnancies. *Am J Med Genet*. 2012;158A(8):1848-1856.
64. Smith CL, Eppig JT. The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data. *Mamm Genome*. 2012;23(9-10):653-668.
65. Smedley D, Oellrich A, Köhler S, et al. *PhenoDigm: Analyzing Curated Annotations to Associate Animal Models with Human Diseases*. Database.; 2013:bat025.
66. Bone WP, Washington NL, Buske OJ, et al. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. *Genet Med*. 2016;18(6):608-617.
67. Brand H, Whelan CW, Duyzend M, et al. High-resolution and noninvasive fetal exome screening. *N Engl J Med*. 2023;389(21):2014-2016.
68. Fan HC, Gu W, Wang J, Blumenfeld YJ, El-Sayed YY, Quake SR. Noninvasive prenatal measurement of the fetal genome. *Nature*. 2012;487(7407):320-324.
69. Miceikaitė I, Hao Q, Brasch-Andersen C, et al. Comprehensive noninvasive fetal screening by deep trio-exome sequencing. *N Engl J Med*. 2023;389(21):2017-2019.
70. Xu C, Li J, Chen S, et al. Genetic deconvolution of fetal and maternal cell-free DNA in maternal plasma enables next-generation noninvasive prenatal screening. *Cell Discov*. 2022;8(1):109.
71. Kitzman JO, Snyder MW, Ventura M, et al. Noninvasive whole-genome sequencing of a human fetus. *Sci Transl Med*. 2012;4(137):137ra76.
72. Schmitz D, Henn W. The fetus in the age of the genome. *Hum Genet*. 2022;141(5):1017-1026.
73. Dhombres F, Morgan P, Chaudhari BP, et al. Prenatal phenotyping: a community effort to enhance the Human Phenotype Ontology. *Am J Med Genet C Semin Med Genet*. 2022. [Internet]. https://doi.org/10.1002/ajmg.c.31989
74. Malniece I, Grasmane A, Inashkina I, et al. The fetal phenotype of noonan syndrome caused by severe, cancer-related PTPN11 variants. *Am J Case Rep*. 2020;21:e922468.
75. Prayer D, Malinger G, De Catte L, et al. ISUOG Practice Guidelines (updated): performance of fetal magnetic resonance imaging. *Ultrasound Obstet Gynecol*. 2023;61(2):278-287.
76. Bilardo CM, Chaoui R, Hyett JA, Kagan KO, Karim JN, et al. ISUOG Practice Guidelines (updated): performance of 11-14-week ultrasound scan. *Ultrasound Obstet Gynecol*. 2023;61(1):127-143.
77. Horgan R, Nehme L, Abuhamad A. Artificial intelligence in obstetric ultrasound: a scoping review. *Prenat Diagn*. 2023;43(9):1176-1219.
78. OBO foundry [Internet]. Accessed 2023 Oct 26. http://obofoundry.org/ontology/mp.html
79. Smith CL, Eppig JT. Expanding the mammalian phenotype ontology to support automated exchange of high throughput mouse phenotyping data generated by large-scale mouse knockout screens. *J Biomed Semant*. 2015;6:11.
80. Groza T, Gomez FL, Mashhadi HH, et al. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res*. 2023;51(D1):D1038-D1045.
81. Blake JA, Baldarelli R, Kadin JA, et al. Mouse genome database (MGD): knowledgebase for mouse-human comparative biology. *Nucleic Acids Res*. 2021;49(D1):D981-D987.
82. Cardoso-Moreira M, Sarropoulos I, Velten B, et al. Developmental gene expression differences between humans and mammalian models. *Cell Rep*. 2020;33(4):108308.
83. Liao BY, Zhang J. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc Natl Acad Sci USA*. 2008;105(19):6987-6992.
84. Cacheiro P, Haendel MA, Smedley D. International mouse phenotyping consortium and the monarch initiative. New models for human disease from the international mouse phenotyping consortium. *Mamm Genome*. 2019;30(5-6):143-150.
85. Byrne AB, Arts P, Ha TT, et al. Genomic autopsy to identify underlying causes of pregnancy loss and perinatal death. *Nat Med*. 2023;29(1):180-189.
86. Yates CL, Monaghan KG, Copenheaver D, et al. Whole-exome sequencing on deceased fetuses with ultrasound anomalies: expanding our knowledge of genetic disease during fetal development. *Genet Med*. 2017;19(10):1171-1178.
87. Filges I, Friedman JM. Exome sequencing for gene discovery in lethal fetal disorders--harnessing the value of extreme phenotypes. *Prenat Diagn*. 2015;35(10):1005-1009.
88. Dawes R, Lek M, Cooper ST. Gene discovery informatics toolkit defines candidate genes for unexplained infertility and prenatal or infantile mortality. *NPJ Genom Med*. 2019;4:8.
89. Cacheiro P, Muñoz-Fuentes V, Murray SA, et al. Human and mouse essentiality screens as a resource for disease gene discovery. *Nat Commun*. 2020;11(1):655.
90. Cacheiro P, Lawson S, Van den Veyver IB, et al. Lethal phenotypes in Mendelian disorders. *medRxiv*. 2024. https://doi.org/10.1101/2024.01.12.24301168
91. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.
92. Babadi M, Fu JM, Lee SK, et al. GATK-gCNV enables the discovery of rare copy number variants from exome sequencing data. *Nat Genet*. 2023;55(9):1589-1597.
93. Repository of the international fetal genomics consortium (RIFGC) [Internet]. Accessed 2023 Oct 26. https://www.ga4gh.org/driver_project/repository-of-the-international-fetal-genomics-consortium-rifgc/


**How to cite this article:** Duyzend MH, Cacheiro P, Jacobsen JOB, et al. Improving prenatal diagnosis through standards and aggregation. *Prenat Diagn*. 2024;44(4):454-464. https://doi.org/10.1002/pd.6522