UPDATE Prenatal phenotyping

Rebecca M. Reimers, MD
Dr. Reimers is a Clinical Fellow, Maternal-Fetal Medicine and Clinical Genetics, Division of Maternal-Fetal Medicine, Brigham and Women's Hospital and Boston Children's Hospital, Boston, Massachusetts.

Stephanie H. Guseh, MD
Dr. Guseh is a Clinical Instructor, Maternal-Fetal Medicine and Clinical Genetics, Division of Maternal-Fetal Medicine, Brigham and Women's Hospital.

The authors report no financial relationships relevant to this article.

In prenatal phenotyping, understanding standardization of language, specific prenatal descriptions, and artificial intelligence may contribute toward the making of a diagnosis

As prenatal genetic testing and imaging have advanced, the diagnosis of genetic disorders has moved from the postnatal to the prenatal time frame. This has largely been facilitated by the increasing use of exome sequencing (ES) in the prenatal setting. Two landmark trials published in January 2019 highlighted the overall diagnostic yields of prenatal ES as 8.5% and 10% in fetuses with normal karyotype and microarray.[1,2]

Although this is a huge step forward in prenatal diagnosis, ES is currently a manually curated, labor-intensive task. The process involves reviewing thousands of sequence variants for any given sample and prioritizing each variant based on bioinformatic data, prediction models, literature review, and specific patient characteristics. The patient characteristics, or phenotypic information, are critically important in prioritizing candidate variants.

To date, prenatal ES has been limited by the use of inconsistent terminology and the lack of well-understood prenatal phenotypes. In this Update, we highlight how recently published work draws attention to these critical gaps in prenatal diagnosis.

IN THIS ARTICLE

Standardizing phenotyping language in the prenatal setting

Tomar S, Sethi R, Lai PS. Specific phenotype semantics facilitate gene prioritization in clinical exome sequencing. Eur J Hum Genet. 2019;27:1389-1397.

Clinical ES in pediatric and adult populations is enhanced by the use of standardized vocabulary to describe disorders. Standardized language ensures that identified variants are filtered correctly and in a systematic fashion based on the patient characteristics that are provided. One commonly used platform is the Human Phenotype Ontology (HPO).

Tomar and colleagues assessed the impact of HPO-based clinical information on the performance of a gene prioritization tool.[3] Gene prioritization (or simulation) tools are used for interpretation of ES data to help analysts efficiently sort through the thousands of variants in an individual's genetic sequence. The performance, or accuracy, of a prioritization tool can be assessed

by looking at the location of the disease-causing gene in the suggested gene list.

Cohort of diagnosed patients and gene prioritization

In this experimental model, Tomar and colleagues included 50 cases with neuromuscular disorders; all had available sequencing data, fully described phenotypes, and known causal genes. The authors varied the level of available clinical information in the HPO terms used for simulated variant analysis. Using 3 web-based gene prioritization tools on the 50 cases, they varied the HPO input to include a random selection of 10%, 30%, and 50% of HPO terms derived from deep phenotyping.

The 3 prioritization tools ranked input genes based on gene-phenotype associations that were derived from gene-phenotype databases. The authors then assessed the quality of the candidate gene lists by the location of the known causative gene on the generated rank lists. They repeated this analysis 4 times with different randomly selected HPO terms.

Inclusion of more HPO terms allowed for more accurate diagnoses in rare disorders

The authors found that the phenotype input for ES matters. When only 10% and 30% of the HPO terms were used to create a candidate gene list, the causative gene was less likely to be in the top portions of gene lists than when 50% or 100% of the available HPO terms were used.

For well-characterized disorders, use of the top 10% HPO terms performed as well as when all available HPO terms were used. For previously undescribed disease-gene associations, identification of the disease gene suffered with more limited HPO term availability.

WHAT THIS EVIDENCE MEANS FOR PRACTICE

The quantity and quality of phenotype input into ES matters for assessing genetic variants. HPO terms have been developed to represent prenatal sonographic findings, and these have been extended to include gestational age of onset in some cases. Providing as much data as possible about the prenatal phenotype through accepted uniform vocabulary (such as HPO) will increase the likelihood that a prenatal diagnosis can be made.

What this study contributes

This study was a simulation of previously sequenced patients with neuromuscular disorders. It examined a small sample size for a narrow spectrum of disease. However, it clearly illustrated the principle that completeness of phenotypic information for ES pipelines is relevant for interpretation.

FAST TRACK

When only 10% and 30% of the HPO terms were used to create a candidate gene list, the causative gene was less likely to be in the top portions of gene lists than when 50% or 100% of the available HPO terms were used

Detailed description of prenatal findings is essential to diagnosis

Aarabi M, Sniezek O, Jiang H, et al. Importance of complete phenotyping in prenatal whole exome sequencing. Hum Genet. 2018;137:175-181.

In a retrospective cohort study, Aarabi and colleagues evaluated the diagnostic utility and limitations of ES in prenatal cases with structural birth defects.[4]

A case series study

The investigators included 20 pregnancies with structural birth defects that were referred to their center for prenatal diagnosis between 12 and 20 weeks' gestation. All pregnancies had normal karyotype and microarray analyses prior to enrollment.

ES was performed on trio samples, which included fetal and parental DNA

WHAT THIS EVIDENCE MEANS FOR PRACTICE

Prenatal genetic diagnosis often is limited by incomplete information about the features seen on ultrasonography. Although not all features are visible prenatally, more diagnoses can be made if laboratories are provided with detailed information about the structural abnormalities that are seen. Furthermore, if ES does not provide a prenatal diagnosis, the data should be reviewed postnatally if more detailed phenotypic information becomes available.

samples (extracted from peripheral blood). Reports provided by the commercial laboratories were normal for all cases and included no pathogenic or likely pathogenic variants. The laboratory provided the investigators with the FASTQ (genetic sequence) files for reanalysis, which was performed using both prenatal and postnatal detailed phenotypic information.

FAST TRACK

Although not all features are visible prenatally, more diagnoses can be made if laboratories are given detailed information about the structural abnormalities that are seen

Use of postnatal information facilitated diagnoses

Reanalysis of ES data using detailed postnatal findings revealed a possible diagnosis in 20% of cases. Each case in which a diagnosis was made, detailed below, highlights an important limitation in our current ability to make prenatal diagnoses.

Case 1. A fetus was diagnosed prenatally with arthrogryposis, plagiocephaly, and club feet. After birth, the infant also was found to have generalized muscle weakness, elevated creatine phosphokinase, and congenital hip dislocation.

Reanalysis of the ES data revealed compound heterozygous missense variants in the nebulin gene (*NEB*). Although classified as variants of uncertain significance (VUS), these are consistent with the phenotype, the authors argued, and with the diagnosis of autosomal recessive nemaline myopathy 2.

Case 2. Prenatal diagnosis was made of a right limb anomaly, tetralogy of Fallot, intrauterine growth restriction, ambiguous genitalia, and dextrocardia. Postnatal evaluation revealed absent pulmonary valve syndrome, right arm dysplasia, pectus carinatum deformity, and failure to thrive.

In this case, ES with the postnatal information revealed a VUS in the *NOTCH1* gene, which has been associated with Adams-Oliver syndrome. Although by strict criteria this variant is also of uncertain significance, Adams-Oliver syndrome is characterized, in part, by transverse limb defects and congenital heart disease, as was found in the proband.

Case 3. Prenatal ultrasonography revealed microcephaly and absence of the septum pellucidum. Postnatal magnetic resonance imaging revealed semi-lobar holoprosencephaly. A holoprosencephaly-specific gene panel revealed a deletion in the *ZIC2* gene, which is known to cause holoprosencephaly.

Careful re-examination of the ES data revealed some abnormality in the *ZIC2* signal, which might have been studied in greater detail and thereby detected if the prenatal diagnosis of holoprosencephaly had been made.

Case 4. An ultrasound evaluation at 12 weeks' gestation revealed a cystic hygroma, short long bones, and possible absent hand and fibula. A postnatal fetal autopsy at 14 weeks showed split-hand and split-foot malformations, which were not appreciated on ultrasonography.

In filtering the ES data with this information, a pathogenic variant in the *PRCN* gene was identified as causal, and the diagnosis of Goltz syndrome was made.

Challenges facing prenatal diagnosis

A case series is inherently limited by its small sample size. Nevertheless, the authors suggest 2 major challenges in our ability to make the above diagnoses in the prenatal setting: 1) the prenatal assessment being limited to major structural abnormalities, and 2) commercial laboratories not having enough experience or volume to interpret the limited information provided by prenatal imaging.

CONTINUED ON PAGE 23

MDedge™ | ObGyn

Ongoing COVID-19 resources for your practice

Did you read these recent clinical and news articles?

- Wuhan data link COVID-19 with myocardial damage
- Physicians and health systems can reduce fear around COVID-19
- Cardiac symptoms can be first sign of COVID-19
- COVID-19 during pregnancy: How would you proceed in this case of a novel and ominous emerging pathogen?
- 20% of U.S. COVID-19 deaths were aged 20-64 years
- COVID-19 critical care guideline offers support for frontline clinicians
- Three COVID-19 rapid diagnostic tests get FDA thumbs-up
- How long is it safe to delay gynecologic cancer surgery?
- AMA offers resources for frontline physicians
- Hand washing and hand sanitizer on the skin and COVID-19 infection risk
- 7 tips for running your practice in the coronavirus crisis
- Should patients with COVID-19 avoid ibuprofen or RAAS antagonists?
- FDA okays serum AMH assay to determine menopause status

Bookmark the CORONAVIRUS UPDATES SPECIALTY page at **mdedge.com/obgyn**

CONTINUED FROM PAGE 20

Can AI technology be incorporated to make a genetic diagnosis?

Hsieh TC, Mensah MA, Pantel JT, et al. PEDIA: prioritization of exome data by image analysis. Genet Med. 2019;21:2807-2814.

Increasingly, ES is used in all types of undiagnosed, rare genetic diseases. Although there is a high diagnostic yield in many populations, ES's clinical utility is limited by the labor-intensive process of interpreting each variant in the context of detailed phenotypic information. The widespread use of HPO would be one step toward standardizing the information that is entered into the analysis of ES data, but even HPO cannot capture certain visual clues.

Hsieh and colleagues attempted to use artificial intelligence (AI) for "next-generation phenotyping" to assess facial dysmorphology and integrate the information into variant classification.[5] The authors described their approach of incorporating AI as "prioritization of exome data by image analysis" (PEDIA).

Designing dysmorphology machine learning

The cohort included 679 individuals with 105 different genetic disorders. All individuals had a previously confirmed molecular diagnosis that would be detected by ES. Each individual had a frontal facial photograph analyzed and detailed clinical features documented in HPO terms extracted by 2 clinicians.

A facial analysis software called DeepGestalt, trained on 17,000 patient images, was used to create a Gestalt score. Each individual had 4 different predicted gene scoring approaches:

- a molecular deleteriousness score
- facial analysis with the Gestalt score
- a combination of molecular deleteriousness score and HPO-based gene-prioritization tool (termed semantic similarity score)
- the PEDIA score, which included all 3 prior approaches.

A type of machine learning algorithm (support vector machine, or SVM) was applied, validated, and used to prioritize genes based on the combined scores.

AI seemed to improve diagnostic accuracy

Utilizing the combination of machine learning, HPO terms, and facial analysis software greatly improved the accuracy of variant classification predictions over any approach alone.

Using only the sequence variant and molecular deleteriousness score, the causative variant was ranked in the top 10 of all identified variants in less than 45% of cases. Adding the HPO-based gene prioritization tools increased the accuracy to 63% to 94%. Use of the PEDIA score, which incorporated all 3, increased the accuracy to 99% for the top 10 ranking.

Even more impressive improvements were made in the top 1 ranking accuracy rate, which went from 36% to 74% without facial image information to 86% to 89% with inclusion of DeepGestalt scores.

Study strengths and limitations

This study's innovative application of facial analysis and machine learning, combined with

FAST TRACK

The combination of machine learning, HPO terms, and facial analysis software greatly improved the accuracy of variant classification predictions over any approach alone

WHAT THIS EVIDENCE MEANS FOR PRACTICE

The accuracy of gene prediction in pediatric and adult populations is enhanced by the use of computer-assisted image analysis and machine-learning algorithms. These computational methods may be employed to automate variant classification, making it more accurate, efficient, and less laborious. Detailed descriptions or characteristic images of prenatal findings also may allow this technology to be introduced in the prenatal setting.

HPO-driven variant classification, showed added benefit. To achieve this with available patient photographs and thorough phenotyping, previously diagnosed patients were used. Because complete ES information was not available for those patients, their known pathogenic variant was inserted into randomly selected exomes from the 1000 Genomes Project (healthy individuals). The authors additionally noted that the PEDIA score performance was diminished for rare disorders in which limited data were available. ●

References

1. Lord J, McMullan DJ, Eberhardt RY, et al; for the Prenatal Assessment of Genomes and Exomes Consortium. Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. *Lancet.* 2019;393:747-757.
2. Petrovski S, Aggarwal V, Giordano JL, et al. Whole-exome sequencing in the evaluation of fetal structural anomalies: a prospective cohort study. *Lancet.* 2019;393:758-767.
3. Tomar S, Sethi R, Lai PS. Specific phenotype semantics facilitate gene prioritization in clinical exome sequencing. *Eur J Hum Genet.* 2019;27:1389-1397.
4. Aarabi M, Sniezek O, Jiang H, et al. Importance of complete phenotyping in prenatal whole exome sequencing. *Hum Genet.* 2018;137:175-181.
5. Hsieh TC, Mensah MA, Pantel JT, et al. PEDIA: prioritization of exome data by image analysis. *Genet Med.* 2019;21:2807-2814.