



Predicting genes from phenotypes using human phenotype ontology (HPO) terms

Anne Slavotinek¹ · Hannah Prasad² · Tiffany Yip¹ · Shannon Rego¹ · Hannah Hoban¹ · Mark Kvale²

Received: 18 November 2021 / Accepted: 16 March 2022 / Published online: 31 March 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The interpretation of genomic variants following whole exome sequencing (WES) can be aided using human phenotype ontology (HPO) terms to standardize clinical features and predict causative genes. We performed WES on 453 patients diagnosed prior to 18 years of age and identified 114 pathogenic (P) or likely pathogenic (LP) variants in 112 patients. We utilized PhenoDB to extract HPO terms from provider notes and then used Phen2Gene to generate a gene score and gene ranking from each list of HPO terms. We assigned Phen2Gene gene rankings to 6 rank classes, with class 1 covering raw gene rankings of 1 to 10 and class 2 covering rankings from 11 to 50 out of a total of 17,126 possible gene rankings. Phen2Gene ranked causative genes into rank class 1 or 2 in 27.7% of cases and the genes in rank class 1 were all associated with well-characterized phenotypes. We found significant associations between the gene score and the number of years, since the gene was first published, the number of HPO terms with an hierarchical depth greater or equal to 11, and the number of Online Mendelian Inheritance in Man terms associated with the phenotype and gene. We conclude that genes associated with recognizable phenotypes and terms deep in the HPO hierarchy have the best chance of producing a high gene score and ranking in class 1 to 2 using Phen2Gene software with HPO terms. Clinicians and laboratory staff should consider these results when HPO terms are employed to prioritize candidate genes.

Introduction

Standardization of physical findings, including features that are typically covered by a dysmorphology assessment, can be performed using Human Phenotype Ontology (HPO) as a structured, categorical medical terminology that also provides a phenotype–genotype database (Robinson and Mundlos 2010; Koehler et al. 2014, 2017, 2019; Zhao et al. 2020). In addition to documenting phenotypes, HPO terms can also predict causative genes. Numerous computational tools have been developed to utilize HPO terms for prioritization of candidate variants and genes, including Phevor (Singleton et al. 2014), Phen-Gen (Javed et al. 2014), eXtasy (Sifrim et al. 2013), Phenotypic Interpretation of eXomes (PhenIX; Zemojtel et al. 2014), Exomiser (Smedley et al. 2015), Phenomizer (Koehler et al. 2009; Koehler et al. 2018)

and Phenolyzer (Yang et al. 2015). These programs have improved next-generation sequencing analysis pipelines and diagnostic yields compared to undirected variant analysis (Smedley and Robinson, 2015; Son et al. 2018). Many of these tools require manual entry of patient clinical signs and symptoms to generate lists of HPO terms, but more recently, natural language processing (NLP) software has been utilized to compile HPO terms from the electronic medical record (EMR; Son et al. 2018).

Trio-based whole exome sequencing (WES) in patients with neurodevelopmental disorders has identified causative variants in an estimated 36% of patients (Srivastava et al. 2019). However, WES frequently produces negative results or returns variants of unknown significance (VUSs; Wright et al. 2018). In patients for whom one or more VUSs are identified, correlation of the known phenotypic spectrum previously attributed to the gene(s) with the clinical features of the patient can be critical for variant interpretation. A comparison of tools integrating patient phenotypic data, typically in the form of HPO terms, with the corresponding variant data from 21 patients who underwent ES showed that the pathogenic variant was correctly assigned the highest rank in 10–40% of cases utilizing either PhenIX, Exomiser,

✉ Anne Slavotinek
anne.slavotinek@ucsf.edu

¹ Division of Genetics, Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA

² Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA

or eXtasy (Pengelly et al. 2017). More recently, platforms such Phenomizer and Phen2Gene have been developed to establish a priori candidate gene lists from HPO terms without any corresponding variant data (Koehler et al. 2018; Zhao et al. 2020). However, the applicability of software tools and the use of HPO terms for gene prediction and prioritization in diagnostic settings has not been widely investigated. In this study, we used HPO terms manually derived from the EMR together with Phen2Gene (Zhao et al. 2020) to investigate the utility of this gene prediction software to generate gene scores and gene rankings and thus to identify the causative gene for patients with known pathogenic (P) or likely pathogenic (LP) variants after WES. We also examined our results to determine if patient variables, quantitative and qualitative variability in HPO terms, and factors pertaining to the underlying gene were correlated with the gene scores and rankings.

Materials and methods

Patient consent and cohort demographics

Participants were enrolled from the Pediatric arm of the Prenatal and Pediatric Genomic Sequencing (P3EGS) project that was part of the Clinical Sequencing Evidence Generating Research (CSER) Consortium. Written, informed consent was obtained from all patients and family members who underwent genetic testing (IRB number 17–22504). Patients with clinical indications for WES were enrolled from Medical Genetics clinics at the Benioff Children's Hospital Mission Bay, Benioff Children's Hospital Oakland, Children's Medical Center in Fresno and from the Pediatric Neurology clinics at Zuckerberg San Francisco General Hospital. Patients were also enrolled from the inpatient services at the Benioff Children's Hospital Mission Bay and the Benioff Children's Hospital Oakland. In keeping with the recruitment goals of CSER, more than 60% of patients were ascertained from underserved and/or underrepresented populations based on at least one of the following: a home zip code indicating the patient was domiciled in a medically underserved area (MUA) or health professional shortage area (HPSA), state medical insurance rather than private medical insurance, and self-reported ancestry from parents.

Pediatric patients were enrolled from the time of birth up to 25 years of age, but all presented with symptoms prior to 18 years of age. Patients were enrolled under one of 11 diagnostic categories that comprised intellectual disability (ID), ID with multiple congenital anomalies (MCA), metabolic disease without ID, metabolic disease with ID, epilepsy without ID, epilepsy with ID, neurodegenerative disease or cerebral palsy without ID, neurodegenerative disease or cerebral palsy with ID, encephalopathy, MCA

without intellectual disability, and Other. At least one biological parent was available for each patient, and patients and available biological parents were sequenced as trios whenever possible.

Whole exome sequencing

The WES results from 453 patients were utilized. For this study, only the 112 patients with variants that were interpreted as P or as LP and causative of the patient's indication for testing were included. The WES methodology and variant interpretation pipeline have been previously reported (Mendelsohn et al. 2020). Variant classification was performed according to the American College of Medical Genetics and Genomics (ACMG) criteria (Richards et al. 2015). The classification for each case ('definitive positive' or 'probable positive') was also collected at the exome sign-out meeting.

Curation of HPO terms

We used the EMR notes and the online database PhenoDB (Hamosh et al. 2013) to convert clinical information from the EMR into HPO terms. PhenoDB is a freely available website that allows clinical researchers to store standardized phenotypic information, diagnosis and pedigree data to analyze genetic testing results. PhenoDB automatically generates an HPO term or terms from entry of clinical data. In our workflow, one of two clinicians scanned the entire note and selected all clinical information relevant to the phenotype from a single encounter note signed by a Genetics or subspecialty provider, as this methodology has been found to be equivalent to using multiple clinical notes (Son et al. 2018). We utilized the last medical genetics or subspecialty note containing a physical examination prior to results disclosure to maximize the available clinical data and to prevent bias following the return of a diagnostic result. All terms that were encountered in the single clinical note were used, but the clinicians did not add in related terms to supplement the list of HPO terms, for example, we did not include 'intellectual disability' as another term that could be added to 'developmental delay'. We included HPO terms encompassing the current condition of the patient, the past medical history and relevant investigative results in addition to physical examination findings. Clinical findings for which we could not identify a corresponding or appropriate HPO term after searching PhenoDB were excluded.

Phen2Gene

We used the HPO terms for each patient and Phen2Gene (Zhao et al. 2020) to predict candidate genes in the absence of any variant information. Phen2Gene generates a gene

score and raw gene ranking for each HPO term list. To generate gene scores, each HPO term is assigned a weight representing the granularity of the phenotypic information conveyed by the HPO term. Each candidate gene also has a score that is a quantitative representation of how the gene is associated with the HP. Phen2Gene gives a weighted score to each gene if it is in the HP's candidate gene list (for more details, see Zhao et al. 2020). All of the genes are then sorted by their scores in descending order.

Gene scores ranged from 0 to 1, with higher gene scores indicating a greater probability of the associated gene being causative for the patient's phenotype. The raw gene rankings ranged from 1 to 17,126, including all genes in Phen2Gene. A lower gene ranking indicates a greater probability of the associated gene being causative. We assigned the raw rankings from Phen2Gene to one of six rank classes that were approximately logarithmic—raw rank < 11 was computed as rank class 1, raw rank 11:50 as rank class 2, raw rank 51:100 as rank class 3, raw rank 101:250 as rank class 4, raw rank 251:1000 as rank class 5, and raw rank 1001+ was computed as rank class 6. From a clinician's viewpoint, rank class 1 represents strong evidence of a gene's likely association with the list of HPO terms derived from a patient note and class 2 moderate evidence; the remaining rank classes 3 to 6 suggest weaker evidence of an association.

Statistical analysis

Data on patient variables was extracted from the RED-Cap study database. We used R version 4.0.5 for statistical calculations on the results of the Phen2Gene analysis. In addition to the built-ins, we utilized the *mgcv* library for generalized additive model (GAM) regression and the *ordinal* library for ordinal regression. From the raw data file, a Perl program was written to merge data sources, filter cases, reformat data and run Phen2Gene on the list of HPO terms to generate a gene score and gene rank for HPO term list. We modelled the gene score (nominal gene score), and the gene rank (ordinal gene rank) for the causative gene relative to all the other ranked genes as the dependent responses. We used linear regression analysis to test the following direct, independent variables against Phen2Gene gene score: patient sex, proband age in fractional years, diagnostic category, variant classification, case result interpretation, inheritance pattern of the causative gene, the year in which the gene was first published in association with a clinical phenotype, the total number of clinical terms linked to the phenotype and gene in OMIM, the number of body systems associated with the phenotype and gene in OMIM, and the number of HPO terms. For genes associated with more than one distinct clinical presentation, we used the OMIM data pertaining to the phenotype that most closely resembled the patient's

clinical presentation. We also tested two derived, independent variables as measures of HPO term specificity: HPO term depth, and number of HPO terms at a depth of 11 or more. To compute HPO depth, the HP classes (i.e., the *HP_number* terms) were extracted from the HPO database (see Supplementary Materials and Methods). The HPO terms were then treated as nodes in a graph and the “*subClassOf*” relations then became directed edges, so that these data yielded an adjacency list representation of a directed graph. Topologically sorting the nodes and walking the resulting sorted list starting from the initial root node enabled computation of the maximum depth for every HPO term. As an HPO term may have multiple different paths from the root node with different depths, the maximum depth found across all paths from the HPO term to the root became the maximum depth for that HPO term. An example is provided in Table 1.

After importing the data into a .csv file, the categorical and ordinal variables were cast as discrete factors in R for regression analysis. Phen2Gene weights phenotypic terms according to their specificity for gene prediction, with terms such as ‘cleft palate’ having far fewer gene and disease associations compared to broader descriptions, such as seizures (Zhao et al. 2020). A more skewed gene score distribution is likely to increase the difference between high- and low-ranking genes and hence provide the HPO terms with greater specificity for their associated genes. The data set was, therefore, generated as four copies that corresponded to each of the four Phen2Gene weight models—‘*sk*’ that weights HPO terms by skewness (default), ‘*w*’ that weights HPO terms by information content, ‘*ic*’ that changes the information content weight calculation, and ‘*u*’ that does not weight HPO term gene lists by information content—and thus the statistical tests for the factors had a Bonferroni-corrected significance threshold of $p < 0.05/4 = 0.0125$.

Finally, to determine if a random selection of HPO terms could result in a relevant gene ranking by chance alone, we created a program that, given the number ‘*k*’ of HPO terms, picked *k* HPO terms at random that were then run through Phen2Gene to create an output of the set of scores and ranks for all the genes that were linked to the HPO terms. One thousand runs of random HPO terms were performed for each number of HPO terms for $k = 1, 2, 3, \dots, 34, 35$, thus generating gene scores and gene ranks for 35,000 random Phen2Gene runs. For each of the 112 genes in this study, the number ‘*k*’ of HPO terms that was created was noted and the score and ranks for each of the genes among the 1000 random runs for that value of *k* was extracted. We then compared how often the gene was in the first ten genes in rank from this study versus how often the gene was in the first ten genes from the 1000 ‘random’ runs.

Table 1 Example of human phenotype ontology (HPO) terms and term depth, together with ranking for *NHS* gene

Phenotype	HPO ^a term	HPO depth	Gene ranking/gene score for <i>NHS</i> gene that can be associated with screwdriver-shaped incisors
Screwdriver-shaped incisors	HP:0006346	11	No linked gene
Abnormal incisor morphology	HP:0011063	10	4/0.518067
Abnormality of dental morphology	HP:0006482	9	67/0.170517
Abnormality of the dentition	HP:0000164	8	736/0.15737
Abnormal oral cavity morphology	HP:0000163	7	> 1000
Abnormal oral morphology	HP:0031816	6	> 1000
Abnormality of the mouth	HP:0000153	5	> 1000
Abnormality of the face	HP:0000271	4	> 1000
Abnormality of the head	HP:0000234	3	> 1000
Abnormality of head or neck	HP:0000152	2	> 1000
Phenotypic abnormality	HP:0000118	1	> 1000
All	HP:0000001	0	Not valid as a phenotype

HPO^a human phenotype ontology

Results

Data regarding direct and derived independent variables have been displayed as a single, denormalized table (Supplementary Table 1) comprising Case result type (definitive positive or probable positive; column headed ‘case_result_type’), gene (‘gene’), gene variant interpretation (pathogenic or likely pathogenic; column headed ‘gene_interpretation’), Phen2Gene weight model (‘u’, ‘ic’, ‘sk’, or ‘w’; ‘weight_model’), patient sex (1 = male and 2 = female; ‘patient_sex’), diagnostic category as listed above (‘pediatric_screening_code’), gene inheritance pattern (‘inheritance_pattern’), HPO term depth structure (‘HPO_depth_str’), number of HPO terms (‘num_HPO’), maximum depth of HPO terms (‘max_HPO_depth’), number of HPO terms with depth greater than or equal to 11 (‘num_HPO_depth_ge_11’), number of years ago that the gene was first published, obtained from OMIM (‘years_ago_gene_pub’), total number of clinical terms linked to the phenotype and gene in OMIM (‘omim_num_terms’), number of body systems for phenotype and gene in OMIM (‘omim_num_systems’), gene rank (‘rank’) and log₁₀ of rank (‘log10_rank’), rank class (‘rank_class’), rank class code (1 to 6; ‘rank class code’) and gene score (‘score’). We also used patient age at enrollment, expressed as a fractional number of years and a list of HPO terms in our analyses but have not included this information in the Supplementary Table.

Data from 112 patients with P or LP variant(s) were included in this study. Two patients (1.7%) had dual diagnoses and each of these patients had two variants relevant to their clinical presentation, making a total of 114 HPO term lists available for Phen2Gene. Among the 112 patients, there were 50 males and 62 females. The commonest diagnostic

category was ID with MCA (41 patients), followed by MCA (28 patients), and ID only (17 patients). 24 causative genes had autosomal dominant inheritance, 81 had autosomal recessive inheritance and 9 were X-linked. Our analyses excluded mitochondrial DNA variants, as we did not have cases that were positive for pathogenic variants in mitochondrial DNA due to limitations in the analysis of the WES data. A summary of data for patient age, number of HPO terms, maximum HPO term depth, number of HPO terms with depth greater than 11, year since the gene was first published, total number of OMIM terms, and total number of body systems associated with each phenotype and gene from OMIM is provided in Table 2. Two of the causative genes, *KIF1BP* and *OTOG*, were not ranked in Phen2Gene as they were not present in the Doc2HPO database and hence were excluded from the analysis of gene score and gene rank. There were no statistically significant differences with the Phen2Gene weight models used, although the ‘sk’ model reduced the gene rank class as defined above by 1.588 and the ‘w’ model reduced the rank class by 0.942 relative to the ‘ic’ model. This suggests that, given the gene score, the ‘sk’ model will give a lower rank class than the ‘u’ or ‘ic’ models and this model was used for the analyses.

From the HPO term lists generated from manual review of the EMR note and PhenoDB, Phen2Gene was able to rank the causative gene previously identified by WES in rank classes 1 and 2, and thus in a list of the top 50 genes, for 27.7% of the causative genes using the ‘sk’ weight model (Table 3). The genes in each rank class are listed in Table 4. Several patients had P or LP variants in the same causative gene, but the rankings for one gene showed considerable variation—for example, *KMT2A* was ranked in the top 50 genes for one patient, but also unranked for another. Most of the

Table 2 Case- and gene-based statistics for 112 patients with 114 pathogenic and likely pathogenic variants

Case-based statistics ($n = 112$)	Range	Median	Mean	SD ^a
Patient age	0–25 years ^b	5 years	6.51	5.64 years
Number of HPO ^c terms	3–35	12	13.63	6.43
Maximum HPO term depth	7–13	11	10.48	1.38
Number of HPO terms with depth greater than 11	0–5	1	1.38	1.30
Gene-based statistics ($n = 114$)				
Years since gene first published	3–32 years	10.5 years	13.02 years	7.68 years
Total number of terms associated with each gene and phenotype in OMIM ^d	3–95	33	34.61	18.44
Total number of body systems associated with each gene and phenotype in OMIM	1–30	14	13.47	7.28

SD^a standard deviation, y^b years, HPO^c human phenotype ontology, OMIM^d online Mendelian inheritance in man

Table 3 Phen2Gene rank classes for genes with pathogenic and likely pathogenic variants

	Causative gene ranked in top 10 Phen2Gene genes	Causative gene ranked in top 11–50 Phen2Gene genes	Causative gene ranked in top 51–100 Phen2Gene genes	Causative gene ranked in top 101–250 Phen2Gene genes	Causative gene ranked in top 251–1000 Phen2Gene genes	Not ranked	Total
Gene rank class	1	2	3	4	5	6	–
Number of variants	14 (12.5%)	17 (15.18%)	5 (4.46%)	11 (9.82%)	20 (17.86%)	45 (40.18%)	112

Rank_class NA = *KIF1BP* and *OTOG*—these genes have been omitted from this table and the results are provided for 112/114 variants; data from Phen2Gene weight model ‘sk’

causative genes that were ranked within the top rank class were pathogenic for established conditions with well characterized and recognizable phenotypes, including *PTPN11* (2 patients), *CDH7*, *TFAP2A*, *RPS6KA3*, *ELN* and *OFD1* (Table 4). These genes were all associated with a clinical phenotype prior to 2009. Single factor scatterplots showed a significant correlation between the gene rank and Phen2Gene gene score ($p = 2e-16$), with a tendency to a lower rank class (i.e., rank class 1 or 2) with increasing gene score (Fig. S1). Rank class 1 had scores ranging from 0.5 to 1.0, whereas for the lower rank classes (3 to 6), almost all genes had scores below 0.4.

A linear regression analysis of the independent and derived variables against the Phen2Gene gene score revealed that the most highly associated variable to gene score was the number of years since the gene was first published, with a positive slope indicating that the longer the time since the gene was first published in association with a phenotype, the higher the gene score ($p < 0.001$) (Fig. 1A, B and Supplementary Table 2). For this variable, there was a strong threshold effect, and with one exception, all genes with a gene score > 0.5 were first reported as causative for a phenotype nine or more years ago. The number of HPO terms with a depth of 11 or more was also significantly associated with gene score ($p < 0.001$). All patients that had terms with HPO depths of 11 or greater had relatively high gene scores,

except for a few gene scores that were > 0.6 for a maximum HPO depth level of 7 (Fig. 2A). A cumulative distribution of maximum depths for HPO terms demonstrated that slightly more than 10% of HPO terms were at a depth of 11 or more (Fig. S2). The number of OMIM terms associated with the gene and phenotype also exhibited a significant linear correlation with gene score ($p = 0.001$; Supplementary Table 2). Genes with autosomal dominant inheritance ($p = 0.015$), the diagnostic category ID with MCA ($p = 0.028$) and patient age ($p = 0.065$) trended towards a significant relationship with increasing gene score, but none of these variables were statistically significant at the Bonferroni-corrected significance threshold of $p = 0.0125$ (Supplementary Table 2).

We then performed univariate (single factor) analyses to examine for interactions between the independent and derived variables and gene score. The results showed that the number of years, since the gene was first published ($p = < 0.001$), number of OMIM terms for the phenotype and gene ($p = < 0.001$), the number of OMIM systems involved for the phenotype and gene ($p = < 0.001$), number of HPO terms with depth 11 or more ($p = < 0.001$), maximum HPO term depth ($p = < 0.001$) and number of HPO terms ($p = < 0.001$) all reached statistical significance (Supplementary Table 3). For gene score versus number of HPO terms, there was no clear linear effect (Fig. 2B). Examining gene score versus diagnostic category showed that patients

Table 4 Phen2Gene rank classes for causative genes with pathogenic and likely pathogenic variants

	Causative gene ranked in top 10 Phen2Gene genes	Causative gene ranked in top 11–50 Phen2Gene genes	Causative gene ranked in top 51–100 Phen2Gene genes	Causative gene ranked in top 101–250 Phen2Gene genes	Causative gene ranked in top 251–1000 Phen2Gene genes	Not ranked
Gene rank class	1	2	3	4	5	6
Gene name	<i>ARID1A</i> <i>BMP2</i> <i>BRAF</i> <i>CHD7</i> <i>CYP11B1</i> <i>ELN</i> <i>GLI2</i> <i>IDUA</i> <i>NFIX</i> <i>OFD1</i> <i>PTPN11</i> <i>RPS6KA3</i> <i>TFAP2A</i> <i>TMEM237</i>	<i>ACTG1</i> <i>ALS2</i> <i>CHD7</i> <i>COL6A2</i> <i>EP300</i> <i>EP300</i> <i>GLI2</i> <i>KMT2A</i> <i>KMT2D</i> <i>KMT2D</i> <i>KMT2D</i> <i>KMT2D</i> <i>KMT2D</i> <i>NR3C2</i> <i>PTEN</i> <i>PTPN11</i> <i>RAD21</i> <i>SCN2A</i> <i>SMAD4</i>	<i>DGAT1</i> <i>KMT2D</i> <i>PDHA1</i> <i>PDHA1</i> <i>PTPN11</i>	<i>COL6A1</i> <i>FOXP2</i> <i>GRIN2D</i> <i>KAT6A</i> <i>KAT6A</i> <i>KMT2A</i> <i>KMT2A</i> <i>LAMA1</i> <i>LAMA2</i> <i>MECP2</i> <i>TUBA1A</i>	<i>AQP2</i> <i>EFTUD2</i> <i>EXT2</i> <i>FOXF1</i> <i>GNAS</i> <i>KCNK4</i> <i>KIF1A</i> <i>KMT2D</i> <i>MAGEL2</i> <i>MCPH1</i> <i>MECP2</i> <i>PMM2</i> <i>PPP2R5D</i> <i>SIN3A</i> <i>SLC16A2</i> <i>SLC17A5</i> <i>SLC19A3</i> <i>SRD5A2</i> <i>SYNGAP1</i> <i>TANGO2</i>	<i>AGA</i> <i>AHDC1</i> <i>AHDC1</i> <i>ANKRD11</i> <i>ARHGEF9</i> <i>ASH1L</i> <i>BCS1L</i> <i>CACNA1A</i> <i>CDK10</i> <i>CDK13</i> <i>CDK13</i> <i>CHAMP1</i> <i>CHD3</i> <i>CHD3</i> <i>DDX3X</i> <i>DYRK1A</i> <i>ECEL1</i> <i>HNRNP2</i> <i>IFT74</i> <i>INTS1</i> <i>KAT6A</i> <i>KCNA2</i> <i>KMT2A</i> <i>KMT2A</i> <i>KMT2A</i> <i>KMT2A</i> <i>KRIT1</i> <i>MAGEL2</i> <i>MECP2</i> <i>MSL3</i> <i>NALCN</i> <i>NFIB</i> <i>PBX1</i> <i>PDCD10</i> <i>POGZ</i> <i>PUS7</i> <i>SATB2</i> <i>SATB2</i> <i>SLC6A1</i> <i>SLC6A1</i> <i>SLC9A6</i> <i>TCF12</i> <i>TRAF7</i> <i>WDR45</i> <i>ZBTB18</i> <i>ZC4H2</i>

Rank_class NA = *KIF1BP* and *OTOG*—these genes have been omitted from this table and the results are provided for 112/114 variants; data from Phen2Gene weight model ‘sk’

enrolled under the category of MCA (category 12) were highly represented in high-scoring genes (Fig. 3), and only three other diagnostic categories—ID and MCA (category 3), ID only (category 4) and Other (category 13)—had genes with Phen2Gene scores above 0.5 (Fig. 3).

We also modeled each rank class from 1 to 6 using ordinal regression. The gene score dominated this regression

analysis, with a Z-score of -15.9 (data not shown). Other significant variables contributing to rank class were the number of years, since the gene was first published ($p = <0.001$) and the total number of OMIM terms ($p = 0.004$; Supplementary Table 4). Testing for factors associated with rank class beyond the gene score to determine the factors contributing to the spread of ranks and different rank classes

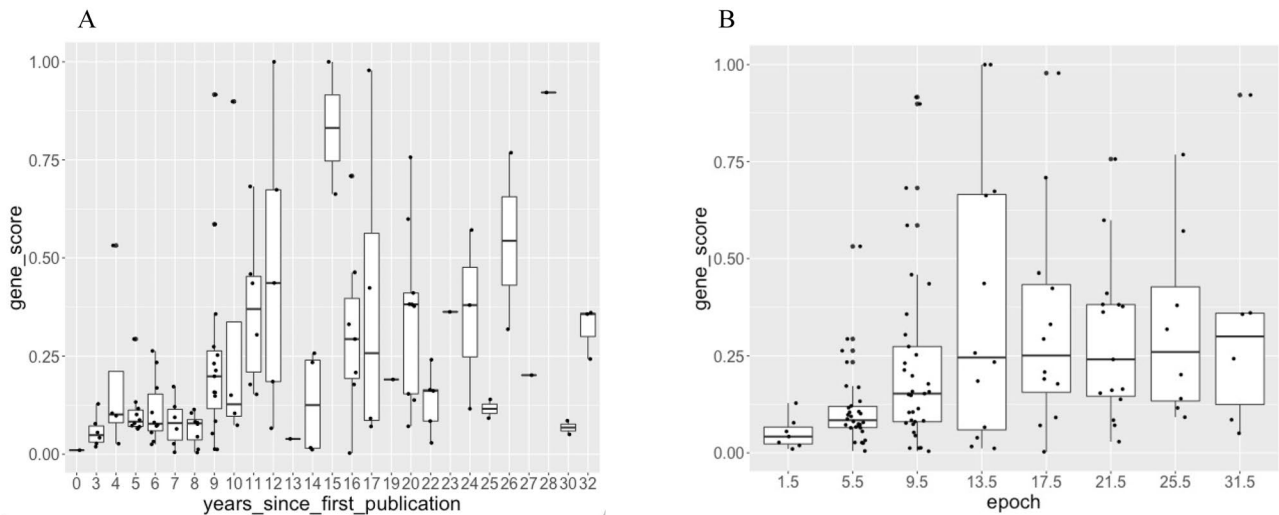


Fig. 1 Relationships between year since the causative gene was first published in yearly and 4-yearly intervals with gene score from Phen2Gene. All graphs are shown as boxplots with minimum and maximum datapoints, median, and boxes representing second and third quartiles. **A** Graphs showing the year, since the causative gene was first published plotted against gene score from Phen2Gene. The datapoints demonstrate a positive slope, indicating that the longer the time, since the gene was first published, the higher the Phen2Gene score. All genes with a gene score > 0.5 except one were first reported

9 or more years ago. The year that the gene was first published was obtained from Online Mendelian Inheritance in Man (OMIM) and the plot was generated in year 2021. **B.** Graphs showing the years since the causative gene was first published in 4–6-yearly intervals, or epochs, plotted against gene score from Phen2Gene. The datapoints again demonstrate a positive slope until a plateau is reached at 13.5 years. The year that the gene was first published was obtained from Online Mendelian Inheritance in Man (OMIM) and the plot was generated in year 2021

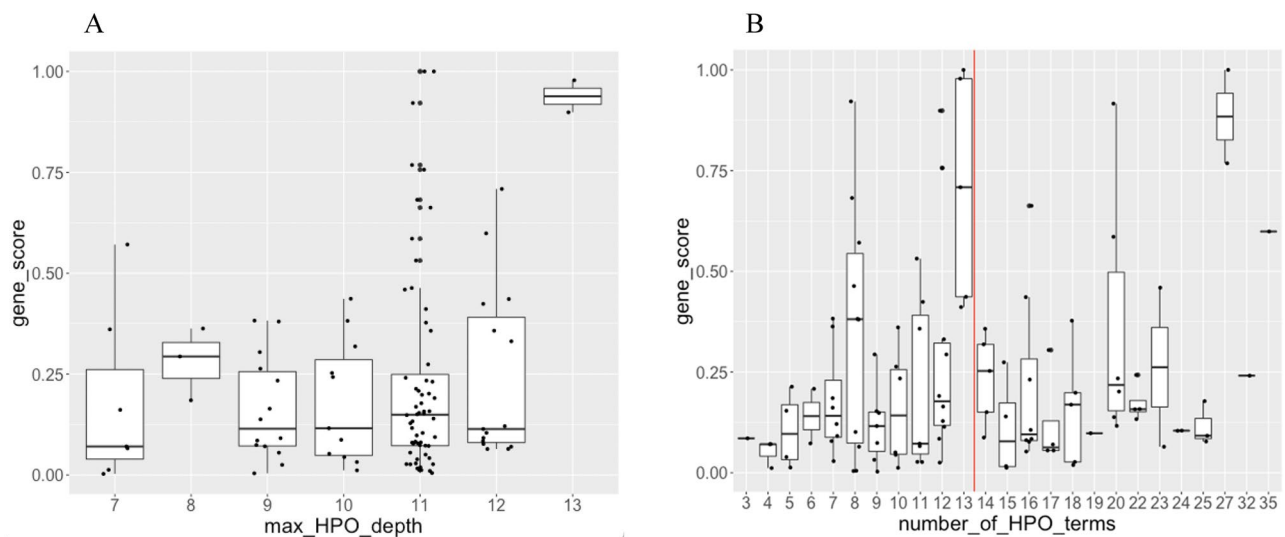


Fig. 2 Relationships between maximum depth of Human Phenotype Ontology (HPO) score and number of HPO terms with gene score from Phen2Gene. All graphs are shown as boxplots with minimum and maximum datapoints, median, and boxes representing second and third quartiles. **A** Graph showing maximum depth of Human Phenotype Ontology (HPO) score plotted against gene score from Phen-

2Gene. The datapoints demonstrate a positive correlation between high gene score and increasing depth of HPO terms. **B** Graph showing the relationship between the number of Human Phenotype Ontology (HPO) terms versus gene score with Phen2Gene. The red line illustrates the mean number of HPO terms

across all genes with a similar gene score revealed that the number of HPO terms ($p = < 0.001$) and weight models ‘sk’ ($p = < 0.001$) and ‘w’ ($p = < 0.001$) were significantly

associated with rank class (Supplementary Table 5). Using a logistic regression analysis to examine the probability that a gene would be present in rank class 1 (raw ranking

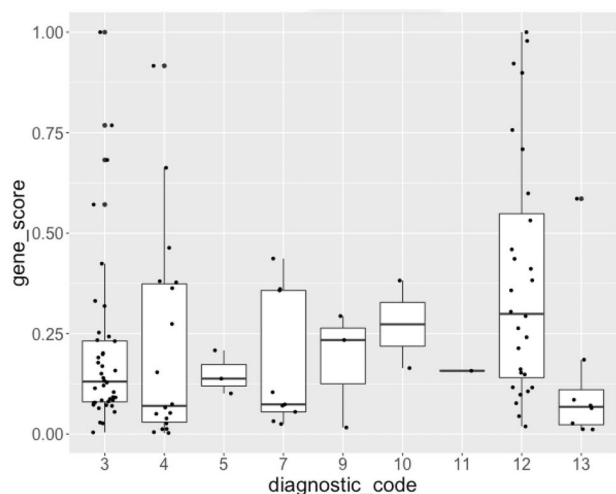


Fig. 3 Graph showing the relationship between diagnostic category/diagnostic code versus gene score from Phen2Gene. Examining gene score versus diagnostic category/diagnostic code showed that patients enrolled under the category of multiple congenital anomalies (MCA; category 12) dominated in high-scoring genes. Only three other diagnostic categories—intellectual disability (ID) and MCA (category 3), ID only (category 4) and Other (category 13)—had genes with Phen2Gene scores above 0.6. The diagnostic categories used were: 3 = ID with MCA; 4 = ID only; 5 = metabolic disease with ID; 6 = metabolic disease without ID; 7 = epilepsy with ID; 8 = epilepsy without ID; 9 = neurodegenerative disease or cerebral palsy with ID; 10 = neurodegenerative disease or cerebral palsy without ID; 11 = encephalopathy; 12 = MCA; 13 = other

1–10) showed a significant relationship with the number of years since the gene was first published in association with a phenotype ($p < 0.001$) and the number of HPO terms with depth greater or equal to 11 ($p < 0.001$; Supplementary Table 6). Proband age was also significant ($p = 0.004$) and the age of the proband had a negative slope, indicating that the probability of the gene being in the highest rank decreased with increasing patient age (Supplementary Table 6). A similar analysis for the probability that a gene would be in the top two rank classes (raw ranking 1 to 50) demonstrated that the significant variables were the number of years ago that the gene was first published in association with a phenotype ($p < 0.001$), diagnostic categories ID and MCA ($p < 0.001$) and ID ($p = 0.009$), number of HPO terms with a depth greater or equal to 11 ($p = 0.001$), and autosomal dominant inheritance ($p = 0.001$; Supplementary Table 7).

Finally, to determine if a random selection of HPO terms could result in a low gene ranking by chance alone, we compared how often a gene was in the first ten genes in rank from this study versus how often the gene was in the first ten genes from 1,000 ‘random’ runs. Our results showed that, when the clinical terms selected from the EMR in this study were converted to HPO terms and run through Phen2Gene, 14/112 (12.5%) of genes were ranked in the first 10

and 31/112 (27.7%) of genes were ranked within the top 50 gene rankings (Table 2). For the random runs performed as above, 775/112,000 (0.69%) were ranked within the top 10 and 3,188/114,000 (2.85%) were ranked within the top 50 gene rankings. This analysis demonstrated that using a relevant set of HPO terms improves the chances of the gene being in the top ten by around 18-fold and improves the chances of being in the top 50 by around tenfold.

Discussion

We studied 112 patients who received a definite positive or probable positive case diagnosis due to P and/or LP variant(s) identified with WES from the Pediatric arm of the P3EGS study. For these patients, we used clinical terms from an EMR note with PhenoDB to generate HPO terms and then used the phenotypic information represented by the HPO terms with Phen2Gene to determine if the previously known, causative gene could be identified due to a high gene score. Our results showed that Phen2Gene was able to rank the causative gene in the two top rank classes, corresponding to a list of the top 50 genes, in 27.7% of cases. At this ranking level, such information could be clinically meaningful, although these results imply that clinicians and laboratorians may need to scan through a list of the top 50 genes to identify candidate genes for further consideration in situations, where WES did not reveal a clear diagnostic variant. However, we identified significant relationships between gene score and the number of years since the gene was first published in association with a phenotype, the number of HPO terms with a depth of 11 or more, and the number of OMIM terms associated with a causative gene. These relationships provide guidance for best practices using HPO terms, implying that the use of specific HPO terms may be more effective for gene rankings with software prediction programs. In addition, clinicians should be attentive to the higher scores and lower rankings that are more likely to accompany genes that are well understood and connected to pleiotropic phenotypes. Although there is variation in the numbers of clinical findings that are associated with different genes, the importance of complete phenotypic delineation and documentation in the medical literature and corresponding databases should be stressed.

A prior comparison of Phen2Gene with Phenolyzer (Yang et al. 2015), AMELIE 2 (Birgmeier et al. 2020), and GeneNetwork Assisted Diagnostic Optimization (GADO; Deelen et al. 2019) showed that Phen2Gene was able to rank the causative gene in the top 50 genes, or top two rank class codes, in 47.4–66.7% patients for three different case sets ranging from 27 to 85 patients (Zhao et al. 2020). Phen2Gene incorporates HPO-gene annotations from the Jackson Laboratory (Kohler et al. 2019), gene-disease annotations from OMIM (McKusick, 2007), ClinVar (Landrum

et al. 2014), Orphanet (Rath et al. 2012) and GeneReviews (Adam 1993) amongst other data, and, although this list of HPO-gene annotations is comprehensive, it is possible that it does not include more recently reported genes that are first described in the medical literature and that accounted for the diagnosis in at least some of the patients in this study. Our lower rate of gene ranking in the top two rank classes may also simply be due to case selection, as the numbers in this study and previous work using Phen2Gene are small (Zhao et al. 2020). As expected, there was a highly significant correlation between gene score and gene rank class for Phen2Gene (Fig. S1).

We examined the effects of patient attributes, diagnostic codes, gene inheritance pattern, number of HPO terms and depth of HPO terms and variables relating to the causative genes, such as the year in which the gene was first published in association with a clinical phenotype, number of clinical terms linked to the phenotype and gene, and the number of body systems associated with the phenotype and gene, to affect the gene score and ranking. We noted a significant relationship between a high gene score and the number of years, since the gene was first published in association with a phenotype. In support of this finding, when the years since a gene was first associated with a phenotype are subdivided into 4–6-year increments, termed epochs, gene score increases with each epoch until around 13.5 years, after which a plateau is reached (Fig. 1B). The phenotypes associated with ‘older’ genes are likely to be better characterized than for ‘newer’ genes and this result thus suggests that the amount of data regarding the gene and its corresponding phenotype can be important for Phen2Gene and similar programs to rank a given gene highly. Supporting this conclusion, the causative genes ranked within the top rank class in this work have all been associated with phenotypes that have been extensively characterized (Table 3). It is also possible that these ‘older’ genes are associated with phenotypes that are more easily recognizable due to distinctive clinical features and thus are easier to identify using specific HPO terms. However, this finding may also reflect the methodology for data ascertainment for Phen2Gene, in which ‘older’ genes are more likely to be represented in databases or reviews and it remains to be determined if the inclusion of recently discovered genes in ‘real-time’ will improve gene prioritization from HPO terms.

A second finding was that the number of HPO terms with a depth of 11 or more was also significantly associated with gene score, implying that specific terms that are deepest in the HPO hierarchy have the best chance of producing a high gene score and a low gene ranking (Table 1). However, in the hypothetical example provided, the most specific HPO term, screwdriver-shaped incisors, with an HPO depth of 11, was not linked to a gene (Table 1), although the prior terms abnormality of the dentition (gene ranking 736 and gene score

0.157), abnormality of dental morphology (gene ranking 67 and gene score 0.171) and abnormal incisor morphology (gene ranking 4 and gene score 0.518) show the improvement in gene identification with increasing term specificity.

In addition to HPO term depth, there was a statistically significant association between gene score and the number of OMIM terms associated with a causative gene, suggesting that genes associated with complex and highly delineated phenotypes are more likely to achieve a higher score. These two variables retained their statistical significance with multiple methods of analyses. The importance of phenotypic specificity and complexity is further supported by statistical significance for the number of OMIM systems involved for the phenotype and gene ($p < 0.001$), maximum HPO term depth ($p < 0.001$) and number of HPO terms ($p < 0.001$) in single factor analyses.

The median number of HPO terms generated from each clinical note was 12, a number that is likely to be higher than typical for clinical scenarios that rely on health professionals to enter medical terms or a physical description. However, the relationship between number of HPO terms and clinical utility is still controversial. Several studies have noted a correlation between a higher number of HPO terms and increased diagnostic yield after genetic testing (Trujillano et al. 2017; Thompson et al. 2019; Pearson et al. 2021). One study found that diagnostic yield was 0.26 with one to five HPO terms, 0.33 with six to fifteen terms, and 0.39 with more than fifteen terms and that this relationship was independent of phenotype, family structure, or inheritance pattern (Trujillano et al. 2017). A meta-analysis of hospital laboratory testing with WES or whole genome sequencing that was associated with greater accessibility of patient information was shown to have had a higher diagnostic utility (0.42, 95% CI 0.38–0.45) compared to reference laboratory testing (0.29, 95% CI 0.27–0.31), leading to a recommendation to include the maximum amount of clinical information with send out testing (Clark et al. 2018). Other studies have suggested that there is an optimal number of phenotypic terms. The gene-ranking tools Phrank (Jagadeesh et al. 2019), hiPhive (Smedley et al. 2015), and PhenIX (Zemojtel et al. 2014) ranked causative genes higher at less than ten phenotypic terms and Phrank yielded the best causative gene rankings at a maximum of three phenotypic terms (Deisseroth et al. 2019). In the same study that had an average of 291 candidate genes per patient, Phrank ranked the causative gene at an average position of 13.4 with unfiltered phenotypic terms generated by ClinPhen, and at an average rank of 9.5 with ClinPhen’s three, top-priority phenotypes (Deisseroth et al. 2019). Prioritization of the phenotype by information content based on the number of genes known to cause the phenotype did not improve rankings (Deisseroth et al. 2019). A recent study on Likelihood Ratio Interpretation of Clinical Abnormalities (LIRICAL) with a collection of 384 case

reports derived from the medical literature utilized a mean number of HPO terms per case of 11.1, with a median number of 9, and a mean number of negated HPO terms per case of 2.71, with a median of 0 (Robinson et al. 2020). The importance of specific phenotypic information for recognition of a condition is well known by health professionals and reflects current clinical practice (Kernohan et al. 2018).

Finally, there were associations between specific diagnostic categories, such as ID with MCA, and autosomal dominant inheritance with gene ranking. Patients with MCA are more likely to have a complex and pleiotropic phenotype that may be associated with specific clinical findings and this finding is indirectly supported by the significant association of number of HPO terms with a depth > 11 and rank class. It is also possible that ‘de novo’ variants that are associated with autosomal dominant inheritance are more likely to be designated as P or LP after variant annotation.

There is still little information on best practices for using HPO terms and phenotypic information in the interpretation of genetic variants. Although some of the clinical information from the EMR did not have an appropriate HPO term, all HPO terms that were generated for each patient were included in the analysis. Manual extraction of clinical information for the generation of HPO terms by a clinician reviewing a note, although in some ways still seen as a gold standard, remains with challenges in addition to the time consuming and labor-intensive nature of the task. Natural language processing programs are faster at generating HPO terms, which may be critical for ill patients (Clark et al. 2018). However, annual examination of notes does not reliably capture phenotypic weighting or specificity as shown to be important in this work, as some features may be more relevant to the clinical presentation than others and more likely to trigger recognition of a particular condition. In addition, numerous terms can be used to describe common clinical occurrences, such as feeding difficulties with a g-tube (for example, g-tube dependent, oral aversion, feeding intolerance, and/or poor feeding), and it is unclear whether all or just one descriptor should be included. Best practices for determining which HPO terms should be utilized are still uncertain and although this work suggests that specific terms deep in the HPO hierarchy and increasing numbers of HPO terms are associated with higher gene scores, clinical documentation can be contradictory. A single note frequently does not capture the temporal nature of a patient’s symptomatology, in terms of age of onset, resolution of symptoms, or fluctuations of the clinical course. Many terms used in medical records are also non-specific, such as ‘pain’, and it can be unclear if terms such as ‘sleep disorder’ or ‘frequent ear infections’ describe normal variation or are indicative of an underlying medical condition. Some NLP programs have developed strategies to compensate for these shortcomings and ClinPhen ignores phenotypes that

are commonly found in the patient population (Deisseroth et al. 2019). We were unable to determine if there were different rankings generated from the notes written by Genetics professionals or by subspecialty providers, as almost all of the notes were from Clinical Genetics providers.

Examination of the list of genes that were not highly ranked by Phen2Gene reveals several plausible explanations for the failure of a given gene to score highly after entry of HPO terms. Recently described or emerging genes and rare genes may be associated with a paucity of phenotypic and clinical information and one limitation of Phen2Gene is that newer genes, for which there may also be less variant information, may not be sufficiently represented in the HPO2Gene Knowledgebase (H2GKB) data set, from which Phen2Gene derives gene rankings from HPO terms. Phen2Gene incorporates HPO-gene annotations from several public sources as above, but the utility of these sources depends on the accuracy of their curation and the frequency with which they are updated and there is a time lag between publication of novel genes in the medical literature and incorporation of relevant information into data sources. It is uncertain if a more nimble or broader pipeline could improve gene prediction accuracy. Genes associated with few specific clinical findings or non-specific phenotypes, such as those linked to predominantly neurocognitive presentations without MCA or facial anomalies, could all lower rankings. With increasingly widespread testing, it is also possible that WES can be performed prior to the development of characteristic physical findings relating to the underlying diagnosis in the patient, thus leading to missing HPO terms that would be critical for gene prediction or identification. Alternatively, ‘phenotypic noise’, with inclusion of clinical terms unrelated or potentially distracting to the phenotype, or a more complex presentation due to dual diagnoses could also prevent ranking. Greater inclusivity of all known genes in the database may also improve gene prediction rankings from Phen2Gene. In addition, we have not studied the effectiveness of this software for using HPO terms to rank and interpret VUSs, a function that is critical for genetic test interpretation.

Further studies are still needed to determine if Phen2Gene and similar programs will provide advantages compared to traditional methods of variant analysis. In particular, it is important to ensure that computational tools are accessible to users from different disciplines. It is also unclear how generalizable our results will be across the different software programs and algorithms that are currently used for gene prediction from phenotypic terms.

Conclusion

After obtaining HPO terms from a single provider note, we used Phen2Gene with the HPO terms to predict and rank the genes containing 112 P or LP variant(s). Phen2Gene was able

to rank the causative gene into the two highest rank classes in 27.7% of cases using the ‘sk’ weight model and high rankings were associated with well characterized genes. We found significant associations between gene score and the number of years, since the gene was first published in association with a clinical phenotype, a high number of OMIM terms associated with the phenotype and gene, and the number of terms with an HPO hierarchical depth greater or equal to 11, implying that well-delineated genes, complex phenotypes and terms deep in the HPO hierarchy had the best chance of producing a high-scoring gene. This information is relevant to establishing best clinical practices for HPO term use in variant interpretation.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00439-022-02449-6>.

Acknowledgements We are grateful to the patients and families and the clinicians who participated in the P3EGS study.

Author contributions All authors contributed to the study conception and design. Data collection was performed by AS, HP, SR, TY and HH and statistical analysis was performed by MK. The first draft of the manuscript was written by Anne Slavotinek and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by Grant Number U01HG009599 from the National Human Genome Research Institute, National Institutes of Health.

Data availability The data sets analyzed during the current study are available from the corresponding author.

Declarations

Competing Interests None of the authors have any competing financial interests.

Ethics Approval Ethics approval was granted by the Ethics Committee of University of California, San Francisco, Institutional Review Board number 17–22504.

Consent to participate Informed consent was obtained from parents and all individual participants included in the study.

Consent to publish No individual data has been provided.

References

Adam MP (ed). 1993 GeneReviews. University of Washington, Seattle. Birgmeier J, Haeussler M, Deisseroth CA, Steinberg EH, Jagadeesh KA, Ratner AJ, Guturu H, Wenger AM, Diekhans ME, Stenson PD, Cooper DN, Ré C, Beggs AH, Bernstein JA, Bejerano G (2020) AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature. *Sci Transl Med* May 20;12(544):eaau9113

Clark MM, Stark Z, Farnaes L, Tan TY, White SM, Dimmock D, Kingsmore SF (2018) Meta-analysis of the diagnostic and clinical utility of genome and whole exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom Med* 3:16. <https://doi.org/10.1038/s41525-018-0053-8>

Deelen P, van Dam S, Herkert JC, Karjalainen JM, Brugge H, Abbott KM, van Diemen CC, van der Zwaag PA, Gerkes EH, Zonneveld-Huijssoon E, Boer-Bergsma JJ, Folkertsma P, Gillett T, van der Velde KJ, Kanninga R, van den Akker PC, Jan SZ, Hoorntje ET, Te Rijdt WP, Vos YJ, Jongbloed JDH, van Ravenswaaij-Arts CMA, Sinke R, Sikkema-Raddatz B, Kerstjens-Frederikse WS, Swertz MA, Franke L (2019) Improving the diagnostic yield of exome-sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. *Nat Commun* 10(1):2837. <https://doi.org/10.1038/s41467-019-10649-4>

Deisseroth CA, Birgmeier J, Bodle EE, Kohler JN, Matalon DR, Nazarenko Y, Genetti CA, Brownstein CA, Schmitz-Abe K, Schoch K, Cope H, Signer R, Network UD, Martinez-Agosto JA, Shashi V, Beggs AH, Wheeler MT, Bernstein JA, Bejerano G (2019) ClinPhen extracts and prioritizes patient phenotypes directly from medical records to expedite genetic disease diagnosis. *Genet Med* 21(7):1585–1593

Hamosh A, Sobreira N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, Valle D (2013) PhenoDB: a new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum Mutat* 34(4):566–571

Jagadeesh KA, Birgmeier J, Guturu H, Deisseroth CA, Wenger AM, Bernstein JA, Bejerano G (2019) Phrank measures phenotype sets similarity to greatly improve Mendelian diagnostic disease prioritization. *Genet Med* 21(2):464–470

Javed A, Agrawal S, Ng PC (2014) Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nat Methods* 11(9):935–937. <https://doi.org/10.1038/nmeth.3046>

Kernohan KD, Hartley T, Alirezaie N, Care4Rare Canada Consortium, Robinson PN, Dymont DA, Boycott KM (2018) Evaluation of exome filtering techniques for the analysis of clinically relevant genes. *Hum Mutat* 39(2):197–201. <https://doi.org/10.1002/humu.23374>

Köhler S (2018) Vom Symptom zum Syndrom mit moderner Softwareunterstützung [From symptom to syndrome using modern software support]. *Internist (Berl)* 59(8):766–775

Köhler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, Mundlos C, Horn D, Mundlos S, Robinson PN (2009) Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* 85(4):457–464. <https://doi.org/10.1016/j.ajhg.2009.09.003>

Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Baillieu-Forrestier I, Black GC, Brown DL, Brudno M, Campbell J (2014) The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res* 42:D966–D974

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, Brudno M, Buske OJ, Chinnery PF, Cipriani V, Connell LE, Dawkins HJ, Demare LE, Devereau AD, de Vries BB, Firth HV, Freson K, Greene D, Hamosh A, Helbig I, Hum C, Jähn JA, James R, Krause R, Laulederkind SJ, Lochmüller H, Lyon GJ, Ogishima S, Olry A, Ouwehand WH, Pontikos N, Rath A, Schaefer F, Scott RH, Segal M, Sergouniotis PI, Sever R, Smith CL, Straub V, Thompson R, Turner C, Turro E, Veltman MW, Vulliamy T, Yu J, von Ziegenweid J, Zankl A, Züchner S, Zemojtel T, Jacobsen JO, Groza T, Smedley D, Mungall CJ, Haendel M, Robinson PN (2017) The human phenotype ontology in 2017. *Nucleic Acids Res* 45(D1):D865–D876

Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gouridine JP, Gargano M, Harris NL, Matentzoglou N, McMurry JA, Osumi-Sutherland D, Cipriani V, Balhoff JP, Conlin T, Blau H, Baynam

- G, Palmer R, Gratian D, Dawkins H, Segal M, Jansen AC, Muaz A, Chang WH, Bergerson J, Laulederkind SJF, Yüksel Z, Beltran S, Freeman AF, Sergouniotis PI, Durkin D, Storm AL, Hanauer M, Brudno M, Bello SM, Sincan M, Rageth K, Wheeler MT, Oegema R, Lourghi H, Della Rocca MG, Thompson R, Castellanos F, Priest J, Cunningham-Rundles C, Hegde A, Lovering RC, Hajek C, Olry A, Notarangelo L, Similuk M, Zhang XA, Gómez-Andrés D, Lochmüller H, Dollfus H, Rosenzweig S, Marwaha S, Rath A, Sullivan K, Smith C, Milner JD, Leroux D, Boerkoel CF, Klion A, Carter MC, Groza T, Smedley D, Haendel MA, Mungall C, Robinson PN (2019) Expansion of the human phenotype ontology (HPO) knowledge base and resources. *Nucleic Acids Res* 47(D1):D1018–D1027. <https://doi.org/10.1093/nar/gky1105>
- Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 42(Database issue):D980–5. <https://doi.org/10.1093/nar/gkt1113>
- McKusick VA (2007) Mendelian Inheritance in Man and its online version. *OMIM Am J Hum Genet* 80(4):588–604. <https://doi.org/10.1086/514346>
- Mendelsohn BA, Belefod DT, Abu-El-Hajja A, Alsaleh NS, Rahbeeni Z, Martin PM, Rego S, Huang A, Capodanno G, Shieh JT, Van Ziffle J, Risch N, Alkuraya FS, Slavotinek AM (2020) A novel truncating variant in ring finger protein 113A (RNF113A) confirms the association of this gene with X-linked trichothiodystrophy. *Am J Med Genet A* 182(3):513–520
- Pearson NM, Stolte C, Shi K, Beren F, Abul-Husn NS, Bertier G, Brown K, Diaz GA, Odgis JA, Suckiel SA, Horowitz CR, Wasserstein M, Gelb BD, Kenny EE, Gagnon C, Jobanputra V, Bloom T, Grealley JM (2021) GenomeDiver: a platform for phenotype-guided medical genomic diagnosis. *Genet Med* 23(10):1998–2002. <https://doi.org/10.1038/s41436-021-01219-5>
- Pengelly RJ, Alom T, Zhang Z, Hunt D, Ennis S, Collins A (2017) Evaluating phenotype-driven approaches for genetic diagnoses from exomes in a clinical setting. *Sci Rep* 7(1):13509. <https://doi.org/10.1038/s41598-017-13841-y>
- Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S (2012) Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum Mutat* 33(5):803–808. <https://doi.org/10.1002/humu.22078>
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehms HL, Laboratory Quality Assurance Committee ACMG (2015) 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* 17(5):405–424
- Robinson PN, Mundlos S (2010) The human phenotype ontology. *Clin Genet* 77(6):525–534. <https://doi.org/10.1111/j.1399-0004.2010.01436.x>
- Robinson PN, Ravanmehr V, Jacobsen JOB, Danis D, Zhang XA, Carmody LC, Gargano MA, Thaxton CL, Biocuration Core UNC, Karlebach G, Reese J, Holtgrewe M, Köhler S, McMurry JA, Haendel MA, Smedley D (2020) Interpretable clinical genomics with a likelihood ratio paradigm. *Am J Hum Genet* 107(3):403–417
- Sifrim A, Popovic D, Tranchevent LC, Ardeshirdavani A, Sakai R, Konings P, Vermeesch JR, Aerts J, De Moor B, Moreau Y (2013) eXtasy: variant prioritization by genomic data fusion. *Nat Methods* 10(11):1083–1084. <https://doi.org/10.1038/nmeth.2656>
- Singleton MV, Guthery SL, Voelkerding KV, Chen K, Kennedy B, Margraf RL, Durtschi J, Eilbeck K, Reese MG, Jorde LB, Huff CD, Yandell M (2014) Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet* 94(4):599–610. <https://doi.org/10.1016/j.ajhg.2014.03.010>
- Smedley D, Robinson PN (2015) Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. *Genome Med* 7(1):81. <https://doi.org/10.1186/s13073-015-0199-2>
- Smedley D, Jacobsen JO, Jäger M, Köhler S, Holtgrewe M, Schubach M, Siragusa E, Zemojtel T, Buske OJ, Washington NL, Bone WP, Haendel MA, Robinson PN (2015) Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* 10(12):2004–2015
- Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, Huang L, Wang L, Shen F, Liu H, Mehl K, Groopman EE, Marasa M, Kiryluk K, Gharavi AG, Chung WK, Hripacsak G, Friedman C, Weng C, Wang K (2018) Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. *Am J Hum Genet* 103(1):58–73. <https://doi.org/10.1016/j.ajhg.2018.05.010>
- Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, Firth HV, Frazier T, Hansen RL, Prock L, Brunner H, Hoang N, Scherer SW, Sahin M, Miller DT (2019) Meta-analysis and multidisciplinary consensus statement: whole exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med* 21(11):2413–2421
- Thompson R, Papakonstantinou A, Beltran S, Töpf A, de Paula EE, Polavarapu K, t Hoen PAC, Missier P, Lochmüller H (2019) Increasing phenotypic annotation improves the diagnostic rate of whole exome sequencing in a rare neuromuscular disorder. *Hum Mutat* 40(10):1797–1812
- Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Köster J, Marais A, Paknia O, Schröder R, Garcia-Aznar JM, Werber M, Brandau O, Calvo Del Castillo M, Baldi C, Wessel K, Kishore S, Nahavandi N, Eyaid W, Al Rifai MT, Al-Rumayyan A, Al-Twajiri W, Alothaim A, Alhashem A, Al-Sanna N, Al-Balwi M, Alfadhel M, Rolfs A, Abou Jamra R (2017) Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet* 25(2):176–182
- Wright CF, Mcrae JF, Clayton S, Gallone G, Aitken S, FitzGerald TW, Jones P, Prigmore E, Rajan D, Lord J, Sifrim A, Kelsell R, Parker MJ, Barrett JC, Hurler ME, FitzPatrick DR, Firth HV (2018) Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genet Med* 20(10):1216–1223. <https://doi.org/10.1038/gim.2017.246>
- Yang H, Robinson PN, Wang K (2015) Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. *Nat Methods* 12(9):841–843. <https://doi.org/10.1038/nmeth.3484>
- Zemojtel T, Köhler S, Mackenroth L, Jäger M, Hecht J, Krawitz P, Graul-Neumann L, Doelken S, Ehmke N, Spielmann M, Oien NC, Schweiger MR, Krüger U, Frommer G, Fischer B, Kornak U, Flöttmann R, Ardeshirdavani A, Moreau Y, Lewis SE, Haendel M, Smedley D, Horn D, Mundlos S, Robinson PN (2014) Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci Transl Med* 6(252):252ra123. <https://doi.org/10.1126/scitranslmed.3009262>
- Zhao M, Havrilla JM, Fang L, Chen Y, Peng J, Liu C, Wu C, Sarmady M, Botas P, Isla J, Lyon GJ, Weng C, Wang K (2020) Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genom Bioinform* 2(2):lqaa032. <https://doi.org/10.1093/nargab/lqaa032>

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