AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

Development and user evaluation of a rare disease gene prioritization workflow based on cognitive ergonomics

Jessica JY Lee,[1] Clara DM van Karnebeek,[1,2,3,]* and Wyeth W Wasserman[1,4,]*

[1]Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, British Columbia, Canada, [2]Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada, [3]Department of Pediatrics and Clinical Genetics, Emma Children's Hospital, Amsterdam University Medical Centres, Amsterdam, The Netherlands, and [4]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada

*These authors contributed equally to this work.

Corresponding Author: Wyeth W. Wasserman, PhD, Department of Medical Genetics, Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia, Room 3109, 950 West 28th Avenue, Vancouver, BC V5Z 4H4, Canada (wyeth@cmmt.ubc.ca)

ABSTRACT

Objective: The clinical diagnosis of genetic disorders is undergoing transformation, driven by whole exome sequencing and whole genome sequencing (WES/WGS). However, such nucleotide-level resolution technologies create an interpretive challenge. Prior literature suggests that clinicians may employ characteristic cognitive processes during WES/WGS investigations to identify disruptions in genes causal for the observed disease. Based on cognitive ergonomics, we designed and evaluated a gene prioritization workflow that supported these cognitive processes.

Materials and Methods: We designed a novel workflow in which clinicians recalled known genetic diseases with similarity to patient phenotypes to inform WES/WGS data interpretation. This prototype-based workflow was evaluated against the common computational approach based on physician-specified sets of individual patient phenotypes. The evaluation was conducted as a web-based user study, in which 18 clinicians analyzed 2 simulated patient scenarios using a randomly assigned workflow. Data analysis compared the 2 workflows with respect to accuracy and efficiency in diagnostic interpretation, efficacy in collecting detailed phenotypic information, and user satisfaction.

Results: Participants interpreted genetic diagnoses faster using prototype-based workflows. The 2 workflows did not differ in other evaluated aspects.

Discussion: The user study findings indicate that prototype-based approaches, which are designed to model experts' cognitive processes, can expedite gene prioritization and provide utility in synergy with common phenotype-driven variant/gene prioritization approaches. However, further research of the extent of this effect across diverse genetic diseases is required.

Conclusion: The findings demonstrate potential for prototype-based phenotype description to accelerate computer-assisted variant/gene prioritization through complementation of skills and knowledge of clinical experts via human–computer interaction.

Key words: genomics, whole exome sequencing, whole genome sequencing, user–computer interface, cognitive ergonomics

BACKGROUND AND SIGNIFICANCE

Whole exome sequencing (WES) and whole genome sequencing (WGS) are allowing clinicians an unprecedented opportunity to examine human genes *en masse* and to diagnose rare genetic diseases.[1–3] An accurate and efficient analysis of DNA sequence data has become crucial for a timely diagnosis of patients, many of whom might otherwise suffer a long and costly diagnostic odyssey.[4–6] However, identifying causal variants among millions of DNA variations in any individual is challenging.[7] For this reason, computational approaches have been created to improve efficiency in multiple aspects of WES/WGS analyses, from encoding available clinical genetic knowledge into computers,[8–11] to collecting comprehensive phenotype information,[12] to prioritizing potentially pathogenic variants,[13–18] and to matching patients for collaborative investigation of rare, novel genetic diseases.[19]

While the above aspects have been improved, variant (or gene in a wider context) prioritization and interpretation during WES/WGS analyses have remained largely expert-driven tasks with computer assistance, as they require cross-examination of complex evidence (eg, variant/gene/phenotype/population-level) that affect treatment decisions.[7,20] As these informatics/interpretative activities are increasingly dominating the overall cost of genomic analyses,[21] an alternative solution for accelerating WES/WGS analyses may lie in the creation of new computational methods that more efficiently collaborate with highly trained experts (whose skills and knowledge are difficult to fully encode into computers). A recent study in this direction has demonstrated that variant prioritization based on a clinician-generated gene list could outperform purely computational methods in the analysis of singleton WES data.[22] The findings suggest the utility of harnessing clinical expertise, such as a clinician's experience, skills in recognizing clinical gestalt, and their ability to evaluate multifactorial information such as disease onset, family history, and negative findings.[22–24]

In this study, we report the design and evaluation of a gene prioritization workflow based on cognitive ergonomics, the study of understanding human cognitive capabilities in interactive systems, and applying this understanding to support human cognition via human–system interaction for optimized system performance.[25] The word "workflow," within the context of this study, refers to a sequence of interactions between clinical experts and computers during computer-assisted variant/gene prioritization. Using this definition, this study focused on examining 2 different designs of interactions (workflows) and their effect on expert performance regardless of variant/gene prioritization algorithms. These 2 workflows are herein referred to as the prototype-based workflow and feature-based workflow.

First, we created the prototype-based workflow (Figure 1A) that aimed to complement the following characteristics of diagnostic reasoning and human cognition reported in literature: (a) clinicians form a gestalt diagnosis from perceived clinical information,[26–28] (b) people tend to make categorizations using an ideal/core representation called the "prototype,"[29] and (c) people tend to focus on deeper structural information when comparing 2 examples, whereas they focus on superficial information when considering an isolated example.[30] The "prototype" in this study refers to a representation that effectively describes patient characteristics, in the form of a specific genetic disease that closely resembles a patient. For example, consider a patient who manifests developmental delay, cleft palate, fusion of the second and third toes, and distinctive facial features (ptosis, narrow forehead, and anteverted nares). Using a prototype,

these individual characteristics can be summarized to describe the patient as "exhibiting the characteristics of Smith-Lemli-Opitz syndrome." In this prototype-based workflow, therefore, clinicians are solicited to provide a prototype disease with resemblance to patient characteristics, prior to initiating gene prioritization. The computer then extracts a set of characteristics described for the selected prototype from an underlying database.

Next, the prototype-based workflow was compared against the feature-based workflow (Figure 1B), which simulated a common process employed by phenotype-driven variant/gene prioritization tools.[13,15–18] This feature-based workflow requires experts to provide a set of individual characteristics observed in the patient before embarking on a computational variant/gene prioritization process. For workflow comparison, a user study was conducted with expert clinical/biochemical geneticists as subjects. The workflows were assessed with respect to accuracy and efficiency in diagnostic interpretation, efficacy in collecting detailed phenotypic information, and user satisfaction. Finally, we created a proof-of-concept mobile application for exploration by interested users.

This study explores an alternative in computer-assisted variant/gene prioritization and interpretation in which computational methods attempt to harness the intellectual power of clinical experts using human–computer interaction. We hope our findings catalyze further interest to explore interactive methods in this domain.

MATERIALS AND METHODS

To help understanding of the upcoming sections, the workflow definitions have been provided in Background and Significance, and in Figure 1. For a detailed explanation of the workflow designs, please refer to the "Workflow designs" section in Results. The upcoming sections also require understanding of an interchangeable use of the words "feature" and "phenotype," both of which will refer to characteristics of patients.

User study participants

Between October 2017 and May 2018, 59 clinicians from specialized (tertiary) healthcare institutions within Canada, the Netherlands, Ireland, Germany, and Switzerland were invited to participate in the user study. Participant inclusion criteria were to (a) hold the title of medical geneticist/biochemical geneticist or specialize in rare genetic diseases, and (b) have prior experience working with WES/WGS data as part of their clinical practice. The invitees were identified by consulting hospital staff directories, a rare disease research network, and collaborators. The invitees were contacted by an email that provided researcher information, explanation on how the contact was obtained, purpose and a brief description of the study, as well as a web link to the user study website. Participation was completely voluntary, and consent to participate was implied by submission of responses. Of the 59 invitees, 18 completed their participation in the study. Power analysis was performed for this study, and its details are provided the Supplementary Material.

The user study was reviewed and approved by the University of British Columbia Research Ethics Board (Certificate: H17-00872).

Simulated clinical scenarios

Five simulated clinical scenarios were developed for the user study (Table 1). One was dedicated as a tutorial exercise and 4 were for clinical scenario analysis exercises. The latter 4 scenarios were
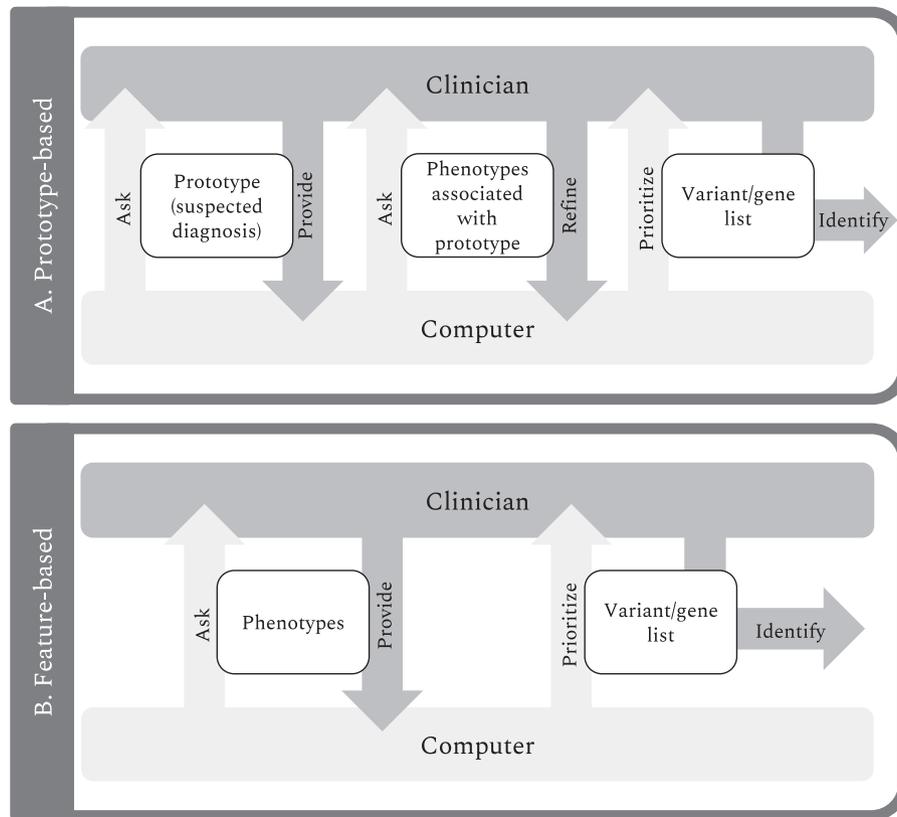
Figure 1. Sequence diagram for prototype-based and feature-based workflows. (A) Illustration of the prototype-based workflow. (B) Illustration of the feature-based workflow. In the prototype-based workflow, clinicians provide a prototype in the form of suspected diagnosis, refine a list of phenotypes that are suggested based on the given prototype, and identify a causal variant/gene from a list of variants/genes that is computationally prioritized by the relevance to given phenotypes. In the feature-based workflow, clinicians provide a list of phenotypes and identify a causal variant/gene from a list of variants/genes that is computationally prioritized by the relevance to given phenotypes. The prototype-based workflow is different from the feature-based workflow in that it explicitly asks clinicians to provide prototypes that they have in mind.

coupled as 2 disease-based pairs, with each pair consisting of a scenario that described a typical presentation and a scenario that described an atypical presentation of a genetic disease. The above arrangements were used for scenario assignment in the study so that each participant analyzed 1 typical scenario from 1 pair and 1 atypical scenario from the other pair, while the order of the scenarios was randomized. This ensured (a) elimination of exposure to the same genetic disease diagnosis during analysis exercises, (b) minimizing ordering bias, and (c) examination of the effect of different disease presentations on workflow performance.

Each scenario consisted of a diagnosis, a patient description, and a gene list, simulating a case involving WES data (equivalent to restricting analysis to exons within WGS data). Normally, WES analyses produce a list of variants at the resolution of nucleotides. In order to limit the time demands on participants, we simplified the results to provide a list of genes impacted by variation. In the user study, participants were explicitly notified that a variant list had been simplified to display only gene-level information, and instructed participants to assume that each gene in the list harbored a variant/variants that was/were rare, potentially pathogenic, and aligned with inheritance models (eg, dominant, recessive).

Each simulated clinical scenario was developed in the following order: diagnosis, patient description, and gene list. Diagnosis selection used the following criteria: (1) the diagnosis was a rare genetic disease that had been described in at least 10 peer-reviewed

publications; (2) it was widely known so that participants could recognize its associated gene by name/symbol during gene list interpretation, thus minimizing time spent looking up gene information using online tools; and (3) the disease was well characterized so that participants could formulate a prototype (or a model presentation of the disease) by reading a text description. After reviewing previously published rare genetic disease annotations,[31] 3 diseases, CHARGE syndrome, Smith-Lemli-Opitz syndrome, and tuberous sclerosis, that fulfilled the above criteria were assigned to each scenario as follows: CHARGE syndrome for the tutorial scenario, Smith-Lemli-Opitz syndrome for 2 analysis scenarios, and tuberous sclerosis for the remaining 2 analysis scenarios.

Based on the diagnosis assignment, patient descriptions were then generated by extracting typical/atypical characteristics from published case reports (Supplementary Material) as well as the disease annotations used during the previous step, which contained a list of phenotypes described using the Human Phenotype Ontology (HPO)[10] and their frequency.[31]

After patient descriptions were generated, gene lists were compiled. The gene list for each scenario contained 17 genes, 1 associated with the scenario's diagnosis and the rest associated with diseases that had varying degrees of similarity to the diagnosis. The purpose of such an arrangement was to ensure the investment of thought and time before discerning the diagnosis. The following outlines the steps that determined gene lists. For each scenario, the

Table 1. Simulated scenarios

		Tutorial scenario	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Diagnosis		CHARGE syndrome (MIM 214800)	Smith-Lemli-Opitz syndrome (MIM 270400)	Smith-Lemli-Opitz syndrome (MIM 270400)	Tuberous sclerosis 1 (MIM 191100)	Tuberous sclerosis 1 (MIM 191100)
Gene		*CHD7*	*DHCR7*	*DHCR7*	*TSC1*	*TSC1*
Typical/Atypical		–	Typical	Atypical	Typical	Atypical
Demographic information		5-month-old girl	18-month-old boy	18-month-old boy	6-year-old girl	6-year-old girl
Family information		Parents were nonconsanguineous and of European ancestry	Parents were nonconsanguineous and of European ancestry	Parents were nonconsanguineous and of European ancestry	Parents were nonconsanguineous and of European ancestry	Parents were nonconsanguineous and of European ancestry
Clinical synopsis[a]	Pregnancy and delivery	Born at term following an uneventful pregnancy and delivery	Born at term following an uneventful pregnancy and delivery	Born at term following an uneventful pregnancy and delivery	Born at term following an uneventful pregnancy and delivery	Born at term following an uneventful pregnancy and delivery
	Phenotypic description	Asymmetric facial palsy Bilateral coloboma of the iris Choanal atresia and ventricular septal defect @ birth Developmental delay Missing ear lobes and short, wide ears Swallowing difficulties	2nd-3rd toe syndactyly Anteverted nares Broad nasal bridge Developmental delay Feeding difficulties and failure to thrive @ 3 months Hypotonia Irritable Low-set ears Microcephaly Micrognathia Postaxial polydactyly Ptosis	Brain MRI and MRS: no structural abnormalities Broad nasal bridge Developmental delay Feeding difficulties @ 3 months Finger clinodactyly Micrognathia Mild hypotonia Mild ptosis Minimal cutaneous 2nd-3rd toe syndactyly	Brain MRI: cortical sclerotic tubers Epileptic seizure Hypomelanotic macules on the chest Hypsarrhythmia Renal cysts Skin papules on the side of nose	Brain MRI: normal Epileptic seizure Hypsarrhythmia Intellectual disability Renal cysts Skin papules on the side of nose

[a]Clinical synopses are summarized from a paragraph format for brevity.

patient description was converted into a list of HPO terms. These terms were then used to compute the scenario's similarity against 6946 diseases in Online Mendelian Inheritance in Man (OMIM) that were annotated by HPO [10] (phenotype_annotation.tab downloaded on June 27, 2017). Similarity was computed using a previously published HPO-based disease similarity score[31] and normalized to a range between 0 and 1. OMIM diseases were then ordered and categorized by their similarity [highly similar (0.6-1.0), similar (0.5-0.6), somewhat similar (0.4-0.5), and irrelevant (0-0.4)]. From each category, 4 diseases were randomly selected, and their associated genes were added to the gene list. All components of the simulated clinical scenarios were reviewed by CDMvK.

User study procedure

The user study was formatted as an online survey. Participants were asked to complete the survey as outlined in Figure 2 and were randomly assigned to either prototype-based or feature-based workflows. The survey consisted of 4 sections: introduction, clinical scenario analysis, debriefing, and user satisfaction questionnaire. The introduction section presented 3 questions regarding participants' demographic information/clinical expertise, an orientation video explaining the study purpose and procedure, and a tutorial exercise that walked through a sample clinical scenario to help participants become acquainted with the survey interface.

The clinical scenario analysis section invited participants to diagnose 2 simulated clinical scenarios using their assigned workflow. For each scenario, the analysis exercise proceeded as follows. Participants were presented with a simulated patient description and

asked to input prototypes or patient phenotypes according to their assigned workflow (Figure 1). The order of the sentences within the description was randomized to minimize ordering bias. For prototype selection, participants were restricted to OMIM disease names (provided by OMIM API).[8] For phenotype selection (feature-based workflow) and phenotype refinement (prototype-based workflow), participants were restricted to HPO terms. Such restrictions were imposed to enable accurate comparison of input from different participants. Afterwards, participants were asked to identify a diagnosis within a simulated gene list, which was ordered by the number of phenotypes that overlapped between input and diseases that were associated with each gene. The ordering was performed to mimic the output of common computational variant/gene prioritization tools.[13,15–18] Gene-phenotype-disease associations provided by HPO[10] were used to enable this functionality. Participants could freely modify input phenotypes and reorder the gene list until they identified a diagnosis. Following diagnosis selection, the actual diagnosis was revealed to participants, and they were invited to express their satisfaction with the assigned workflow by completing a modified After-Scenario Questionnaire (ASQ).[32]

During each analysis exercise, the following information was collected: prototype/phenotype selections, changes made to prototype/phenotype selections before making diagnoses, final diagnoses, time elapsed between initial display of the gene list and identification of diagnoses, and ASQ responses.

Upon completion of 2 clinical scenario analyses, participants were debriefed about the alternative (unassigned) workflow. During debriefing, they walked through the alternative workflow using the same scenarios and completed an ASQ at the end of each scenario.
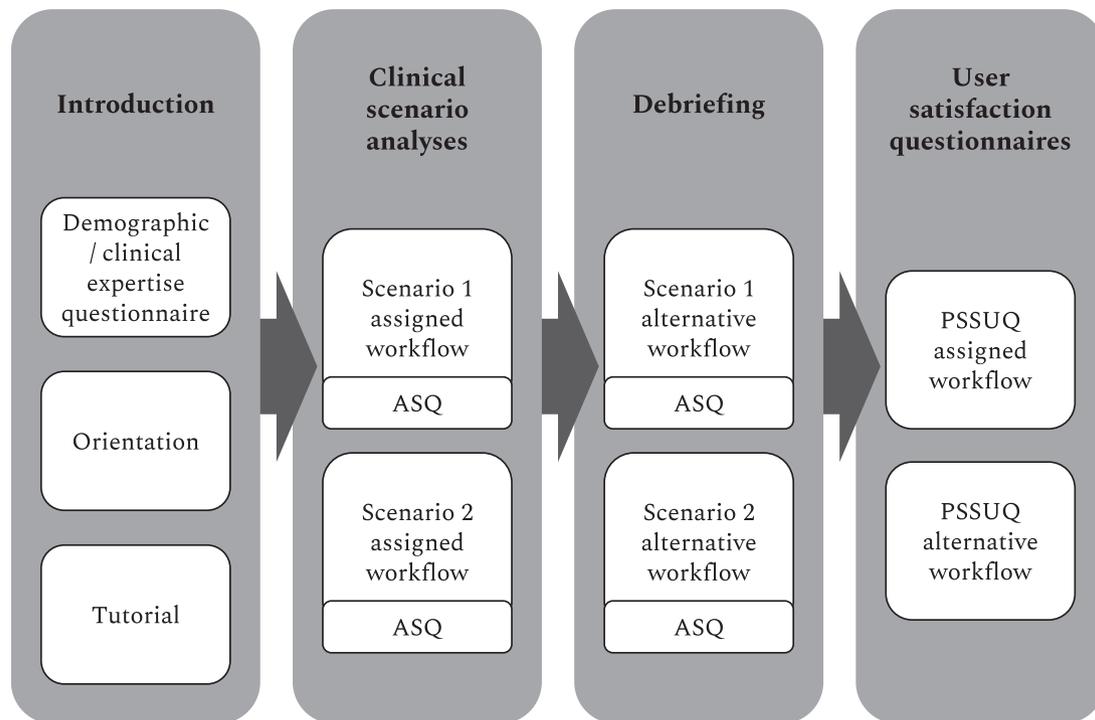
Figure 2. User study structure. The user study consisted of 4 main sections: introduction, clinical scenario analysis, debriefing, and user satisfaction questionnaire. During the introduction, participants answered questions regarding their demographic information and clinical expertise, watched an orientation video, and walked through a sample clinical scenario. Afterwards, participants analyzed 2 simulated clinical scenarios using their assigned workflow. At the end of each scenario, participants completed an After-Scenario Questionnaire (ASQ). Upon completion of clinical scenario analyses, participants were debriefed about the workflow that they were not assigned to and tried out the workflow using the same simulated scenarios. Participants also filled out an ASQ at the end of each scenario. Finally, participants filled out Post-Study System Usability Questionnaires, regarding the assigned workflow and the alternative (unassigned) workflow, respectively.

Only ASQ responses were collected during the walkthrough. Finally, participants were invited to express their overall satisfaction with the workflows by completing 2 modified Post-Study System Usability Questionnaires (PSSUQ)[32] for the assigned workflow and for the alternative workflow, respectively.

For the survey, a custom online interface was developed using Ruby on Rails and React.js in order to implement functionalities required by clinical scenario analysis exercises.

Data analysis

All data analyses were performed using R version 3.4.4. The 2 workflows were compared with respect to (a) diagnostic accuracy (measured as the number of correctly diagnosed scenarios), (b) efficiency in gene list interpretation (measured as the time elapsed between when the gene list was presented and when participants selected a causal gene from the list; Supplementary Material), (c) efficacy in phenotype collection (measured as the number of participant-provided phenotypes), and (d) user satisfaction (measured as ASQ and PSSUQ scores). All comparisons except the PSSUQ score comparison were performed using a 2 x 2 analysis of variance (ANOVA) (afex package) with workflow assignments (prototype-based/feature-based) as a between-subject variable, disease presentations (atypical/typical) as a within-subject variable, and each measurement as a response variable. The primary focus of ANOVA was on the main effect of workflow assignments. PSSUQ scores were compared using the Mann–Whitney U test (wilcox.test). To account for multiple comparisons within (d), the Bonferroni correction (p.adjust) was applied to ASQ and PSSUQ comparisons. Participant-provided prototypes

and phenotypes were analyzed for common and workflow-specific information patterns. Optional written comments provided in ASQ and PSSUQ were reviewed to extract common participant opinions.

RESULTS

Workflow designs

We present the 2 workflow designs investigated in this study as follows. The prototype-based workflow (Figure 1A) was designed to augment the following properties of clinical reasoning and human cognition during WES/WGS investigations: (a) an ability to form gestalt diagnosis,[26–28] (b) a tendency to categorize using an ideal/core representation called the "prototype,"[29] and (c) a tendency to focus on deeper structural information when comparing 2 examples.[30] The specific steps of this prototype-based workflow follow: (1) the computer solicits the clinician to provide a prototype in the form of suspected diagnosis; (2) the computer presents a list of key phenotypes of the given prototype; (3) the clinician refines the presented list by adding/excluding phenotypes; (4) the computer prioritizes genes based on their overlap with the phenotypes; and (5) the clinician specifies a causal gene (diagnosis) from the prioritized list.

The rationale behind this prototype-based workflow design was that articulating gestalt diagnosis in the form of prototype (suspected diagnosis) and using this prototype as an aggregate representation of patient phenotypes would simultaneously (a) relieve the requirement to frequently recall granular details of the patient and (b) engender focus on structural information. To embody this concept, step (3) of the prototype-based workflow was implemented to
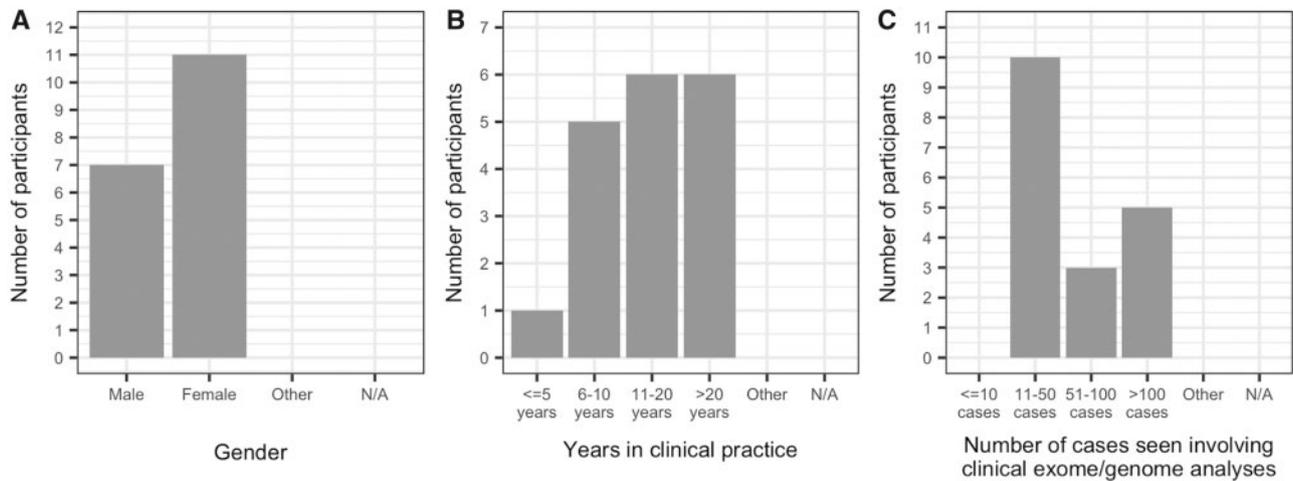
Figure 3. Participant characteristics. (A) Gender of participants; (B) participants' level of clinical expertise, measured as years in clinical practice; and (C) participants' experience with exome or genome sequencing data, measured as the number of cases involving exome or genome analyses.

encourage clinicians to compare patient characteristics with respect to key presentations of the selected prototype. Based on cognitive principles, this comparison was anticipated to bring focus on deeper structural differences, resulting in more concrete and detailed description of patient phenotypes, in contrast to asking clinicians to describe patients *de novo*. The same concept was also applied to step (5) of the prototype-based workflow. In this step, the prototype was to be used as a proxy against which candidate genes and their associated diseases would be assessed for clinical relevance to patient phenotypes.

The feature-based workflow (Figure 1B) was designed for comparison with the prototype-based workflow. This feature-based workflow modeled common phenotype-driven variant/gene prioritization tools.[13,15–18] Specific steps of the feature-based workflow follow: (1) the computer solicits the clinician to provide a list of patient phenotypes; (2) the computer prioritizes genes based on their overlap with the phenotypes; and (3) the clinician identifies a causal gene (diagnosis) from the prioritized list.

The difference between the 2 workflows was that the prototype-based workflow explicitly asked for a prototype (to populate the set of phenotypes, which the user can refine by eliminating/adding terms), whereas the feature-based workflow required the user to serially specify individual patient phenotypes.

User study participant characteristics

Characteristics of the 18 participants are summarized in Figure 3. Ninety-four percent of participants have practiced more than 5 years. All participants had experience with cases involving clinical WES/WGS data.

Workflow performance evaluation

Figure 4 summarizes the performance of the prototype-based workflow and the feature-based workflow. There was no difference in diagnostic accuracy between the two workflows [($F(1, 16) = 1.0$, $P = .33$, $\eta_p^2 = .059$]. Almost all participants, except 1, correctly diagnosed assigned scenarios. The participant who incorrectly diagnosed 1 scenario explained via optional comments that a general diagnosis (tuberous sclerosis) was correctly anticipated and the correct genetic diagnosis (*TSC1*) was considered during gene list interpretation. However, the participant determined that the presented scenario

was more compatible with a different genetic diagnosis (*TSC2*) and thus did not select any diagnosis.

Participants who were assigned to prototype-based workflows identified diagnoses significantly faster than those assigned to feature-based workflows [$F(1, 16) = 6.04$, $P = .026$, $\eta_p^2 = .27$]. In addition, participants identified diagnoses faster for scenarios with typical presentations than atypical presentations [$F(1, 16) = 18.1$, $P = .0006$, $\eta_p^2 = .53$], while no significant interaction between workflow assignment and disease presentation was observed [$F(1, 16) = 3.26$, $P = .090$, $\eta_p^2 = .17$].

No difference was observed in the number of phenotypes collected by either workflow [$F(1, 16) = 2.71$, $P = .12$, $\eta_p^2 = .14$]. Three outliers were observed in the number of phenotypes collected using prototype-based workflows. Examination of individual responses revealed that at least 2 participants who were assigned to prototype-based workflows selected almost all of the phenotypes that were suggested based on participant-specified prototypes, regardless of their presence/absence in simulated scenarios (ie, they chose not to eliminate phenotypes that were not reported in the scenarios). Lastly, there was no difference in user satisfaction between the 2 workflows [ASQ: $F(1, 16) = 1.50$, $P = .48$ (uncorrected $P = .24$), $\eta_p^2 = .086$; PSSUQ: $W = 37$, $P = 1.0$ (uncorrected $P = .79$), $r = 0.19$].

Summary of prototype and phenotype selection

Nine participants who were assigned to prototype-based workflows selected the actual or very close diagnoses as prototypes prior to interpreting gene lists (Table 2). Phenotypes that were collected by the 2 workflows are summarized in Figure 5 and detailed in the Supplementary Material. Phenotypes provided by 3 outlier individuals assigned to prototype-based workflows were excluded from this comparison, as those phenotype lists likely did not involve a conscious assessment of patient phenotypes. Participants who were assigned to feature-based workflows had a tendency to input close synonyms of a phenotype. For example, hypotonia in the atypical Smith-Lemli-Opitz scenario was captured in 3 different terms: generalized hypotonia, central hypotonia, and muscular hypotonia. Meanwhile, synonyms were rarely present in phenotypes captured by prototype-based workflows because participants were offered to select/unselect suggested phenotypes that were associated with the
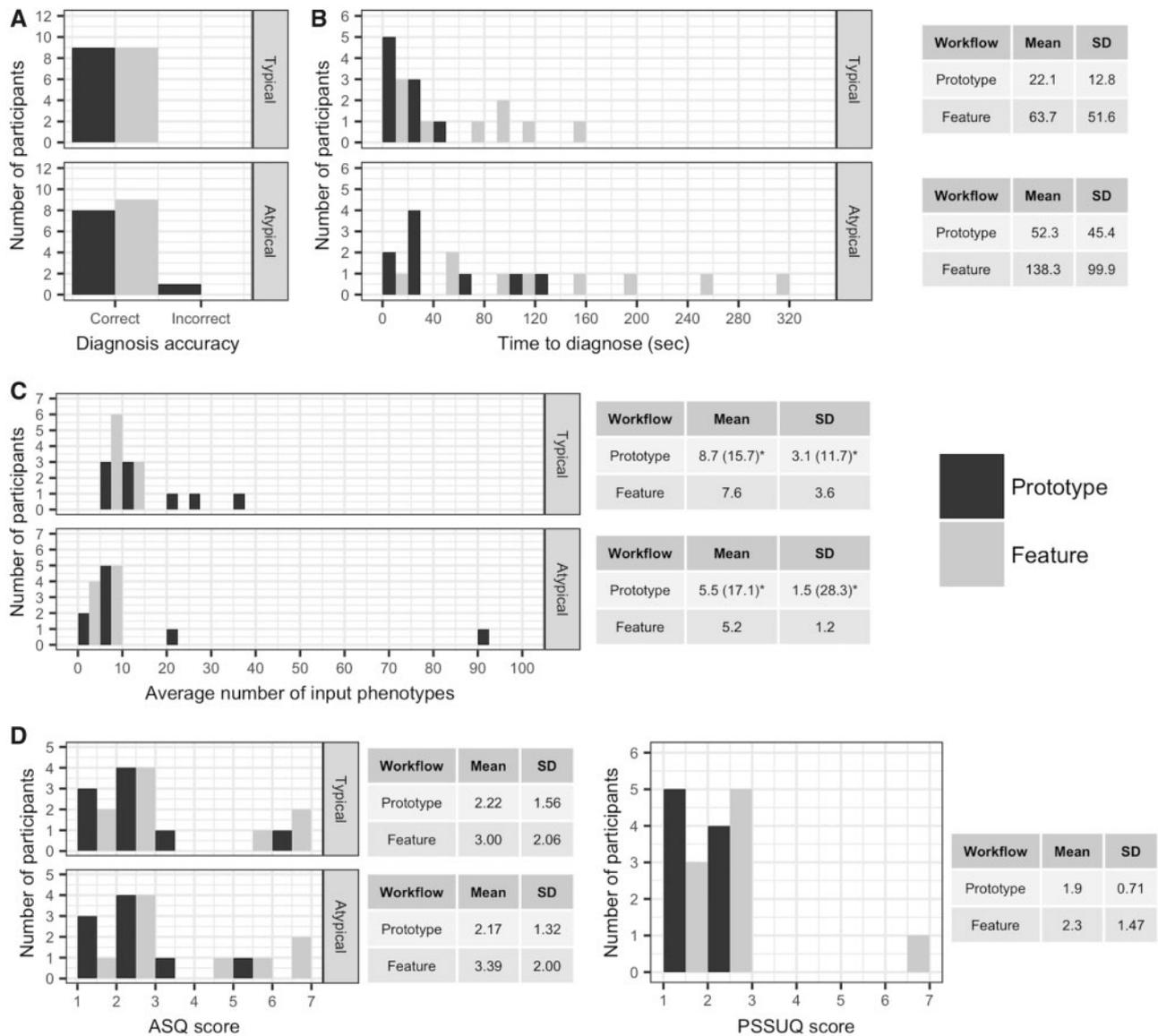
Figure 4. Summary of workflow performance evaluation. Evaluation results are shown in histograms or bar-plots for categorical variables. In (B), (C), and (D), the tables next to histograms summarize descriptive statistics for each corresponding histogram. SD = standard deviation. (A) Diagnostic accuracy, measured as the number of correctly diagnosed scenarios. (B) Interpretation time, measured as the time elapsed between when the gene list was presented and when participants selected causal gene from the list. (C) Number of participant-provided phenotypes. Values denoted by * represent mean or standard deviation including (within brackets) or excluding (without brackets) 3 outlier individuals assigned to the prototype-based workflow. (D) User satisfaction, measured as After-Scenario Questionnaire (ASQ) and Post-Study System Usability Questionnaire (PSSUQ) scores.

prototype of their choice. Furthermore, the prototype-based suggestions seem to have encouraged participants to enter additional phenotypes that were not collected by feature-based workflows. For example, terms such as vomiting, gastroesophageal reflux, and poor suck were provided for feeding difficulty in the atypical Smith-Lemli-Opitz scenario.

Mobile application for workflow exploration

Motivated by participant suggestions, we created a proof-of-concept, open-source, mobile application, PhenoChat (https://github.com/jes8/phenochat), for exploration by interested users (but not for clinical use). Implementation details are provided in the Supplementary Material. Briefly, PhenoChat allows users to build

and send phenotypic descriptions using either workflow interchangeably.

DISCUSSION

Prior literature on diagnostic reasoning and cognitive properties[26–30] suggests that clinicians may employ prototypes (paragon disease presentations) to assess patients and identify relevant genetic diagnoses within WES/WGS results. We designed a novel gene prioritization workflow based upon a prototype-based approach and evaluated it against a workflow that simulated a common phenotype-driven variant/gene prioritization process.[13,15–18] Finally, we demonstrated that gene interpretation could be accelerated using the prototype-based workflow by facilitating prototypical thinking.

Table 2. Prototype selection summary

	Actual scenario diagnosis	Participant-specified prototype	# of participants who selected the prototype[a]
Scenario 1 (n[b] = 5)	Smith-Lemli-Opitz syndrome (MIM 270400)	Smith-Lemli-Opitz syndrome (MIM 270400)	5
Scenario 2 (n[b] = 4)	Atypical Smith-Lemli-Opitz syndrome (MIM 270400)	Smith-Lemli-Opitz syndrome (MIM 270400)	4
Scenario 3 (n[b] = 4)	Tuberous sclerosis 1 (MIM 191100)	Tuberous sclerosis 1 (MIM 191100)	2
		Tuberous sclerosis 2 (MIM 613254)	2
Scenario 4 (n[b] = 5)	Atypical tuberous sclerosis 1 (MIM 191100)	Tuberous sclerosis 1 (MIM 191100)	3
		Tuberous sclerosis 2 (MIM 613254)	3

[a]Counts how many participants selected each prototype as a probable diagnosis. If participants changed prototypes multiple times, they were counted for all prototypes that they had specified.

[b]n = number of participants assigned to scenario using the prototype-based workflow.
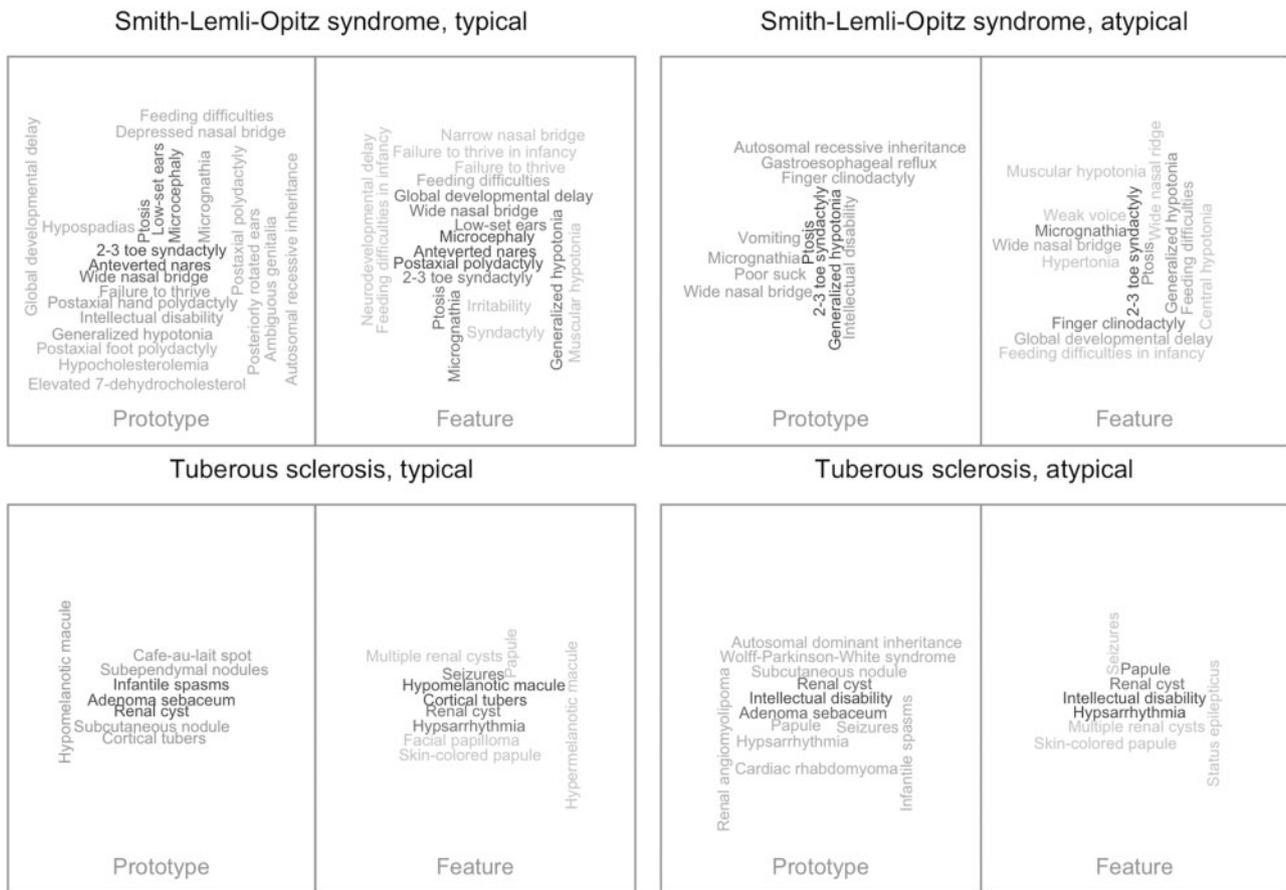
Figure 5. Word-cloud visualization of phenotype selection. For each scenario, phenotype terms collected by the 2 workflows are summarized into word clouds. Darker colors represent terms that were collected more frequently. The underlying data are available in Supplementary Table S2.

Within the scope of this study, the observed time spent on gene list interpretation was significantly shorter for the prototype-based workflow compared to the feature-based workflow. This suggested that participants likely engaged in prototypical thinking. The main difference in workflow designs was that the prototype-based workflow explicitly kept track of prototypes. Through tracking, the prototype-based workflow likely reminded participants of their reasoning process and encouraged prototypical comparison of genetic diagnoses. This notion was also supported by a secondary finding, in which time spent on gene interpretation was shorter for both workflows when analyzing typical scenarios compared to atypical

scenarios. This difference agreed with reports in prototype theory research regarding faster recall and recognition of typical members of a category compared to atypical members.[33–36] In sum, it was likely that participants employed some level of prototypical thinking in both workflows, while the reasoning process was more efficiently facilitated by the prototype-based workflow.

However, the above evaluation was limited in scope to focus on only 2 genetic diseases that were both well characterized in the literature. More research is needed to generalize the observed workflow performances over different rare genetic diseases that have not yet reached the same level of characterization or expert awareness. The

workflows should also be assessed in terms of (a) performance with diseases that present with heterogeneous, overlapping, or novel phenotypes and (b) incorporation of information beyond gene level. In addition to the scope of the evaluation, the study recruitment was also restricted to medical/biochemical geneticists in order to ensure that participants represent a focused group of users, who would exhibit similar areas of attention/interest when approaching WES/WGS data as well as shared desiderata towards WES/WGS analysis software.[37] Further research should also consider inclusion of other healthcare professionals involved in clinical WES/WGS interpretation (eg, genetic counselors and bioinformaticians) to gain further insights into diverse groups of users and their interaction with different workflows.

While the 2 workflows resulted in equivalent phenotypic information amounts, differences in the content of phenotypic information suggested possible involvement of distinct cognitive processes during phenotype assessment. Phenotype terms collected from feature-based workflows did not deviate greatly from simulated patient descriptions, whereas those collected from prototype-based workflows did. The deviating terms were relevant concepts but not exact synonyms: for example, cafe-au-lait spot was provided in relation to hypomelanotic macule, and renal angiomyolipoma was provided in relation to renal cysts. This observation could be explained by a cognitive tendency towards focusing on deeper structural details when comparing 2 examples as opposed to considering a single example.[30] However, a quantitative experiment is required to conclusively determine involvement of the aforementioned cognitive tendency during phenotype assessment within different workflows.

Upon observing no difference in user satisfaction, optional comments provided in user satisfaction questionnaires were examined. Specific comments suggested that the study findings should be translated by implementing the best of both worlds. Feature-based workflow participants pointed out that (1) having to enter each phenotype did not enhance productivity and thus opted to enter only those deemed highly discriminatory; and (2) it was occasionally difficult to code phenotypes *impromptu*. Meanwhile, prototype-based workflow participants highlighted that (1) a typical feature could not be found in phenotype suggestions (likely due to limitations of disease-phenotype annotations); and (2) some thought it was redundant to refine the phenotype list. The above comments suggested that perceived deficiencies of one workflow could be remedied by the other, and flexibility to use either workflow for phenotype specification (as demonstrated by PhenoChat) or incorporating the prototype-based workflow into existing feature-based workflows would provide synergistic utility for prospective implementations.

Heuristics biases, such as overconfidence, anchoring effect, and self-confirmatory bias, are potential risks associated with the prototype-based workflow. However, more research is required to identify when these biases may arise and when the benefits of a prototype-based workflow may outweigh the risks, as cases in which heuristics provided diagnostic advantages over computer-derived approaches have been demonstrated.[22] Meanwhile, implementations of the prototype-based workflow should consider incorporating measures aimed at reducing heuristics biases. For example, a decision schema based on variant interpretation guidelines[38] can be incorporated into the workflow steps to enforce closer examination of user-identified variants.

CONCLUSION

In summary, we explored the utility of augmenting clinical reasoning and cognitive characteristics of experts within computer-assisted gene prioritization. We found that clinicians interpreted genes faster using a prototype-based gene prioritization workflow. Clinician feedback suggested that the prototype-based workflow may provide optimal utility if implemented in synergy with common feature-based variant/gene prioritization workflows. However, further investigation is warranted to confirm the above findings across diverse rare genetic diseases. WES/WGS informatics methods that complement human–computer interactions offer promise for overcoming the informatics bottleneck in clinical genome analysis.

CONTRIBUTORS

JJYL, CDMvK, and WWW designed the study and contributed to the interpretation of results. JJYL generated the simulated clinical scenarios, conducted the user study, performed the data analysis, and drafted the manuscript. CDMvK reviewed the simulated clinical scenarios and assisted with participant recruitment. CDMvK and WWW supervised the user study execution and data analysis. All authors edited and approved the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

REFERENCES

1. Tarailo-Graovac M, Shyr C, Ross CJ, *et al*. Exome sequencing and the management of neurometabolic disorders. *N Engl J Med* 2016; 374 (23): 2246–55.
2. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, *et al*. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet* 2017; 25 (2): 176–82.
3. Stavropoulos DJ, Merico D, Jobling R, *et al*. Whole genome sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *NPJ Genom Med* 2016; 1.

4. Chong JX, Buckingham KJ, Jhangiani SN, *et al*. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *Am J Hum Genet* 2015; 97 (2): 199–215.

5. Stark Z, Schofield D, Alam K, *et al*. Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. *Genet Med* 2017; 19 (8): 867–74.

6. Vissers LELM, van Nimwegen KJM, Schieving JH, *et al*. A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet Med* 2017; 19 (9): 1055–63.

7. Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. *Nat Rev Genet* 2017; 18 (10): 599–612.

8. Online Mendelian Inheritance in Man, OMIM®. Baltimore, MD: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. https://omim.org/ Accessed June 28, 2017.

9. Landrum MJ, Lee JM, Benson M, *et al*. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018; 46 (D1): D1062–7.

10. Köhler S, Vasilevsky NA, Engelstad M, *et al*. The human phenotype ontology in 2017. *Nucleic Acids Res* 2017; 45 (D1): D865–76.

11. Wishart DS, Feunang YD, Marcu A, *et al*. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 2018; 46 (D1): D608–17.

12. Girdea M, Dumitriu S, Fiume M, *et al*. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat* 2013; 34 (8): 1057–65.

13. Robinson PN, Köhler S, Oellrich A, *et al*. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res* 2014; 24 (2): 340–8.

14. James RA, Campbell IM, Chen ES, *et al*. A visual and curatorial approach to clinical variant prioritization and disease gene discovery in genome-wide diagnostics. *Genome Med* 2016; 8: 13.

15. Thuriot F, Buote C, Gravel E, *et al*. Clinical validity of phenotype-driven analysis software PhenoVar as a diagnostic aid for clinical geneticists in the interpretation of whole-exome sequencing data. *Genet Med* 2018; 20: 942–9.

16. Javed A, Agrawal S, Ng PC. Phen-Gen: combining phenotype and genotype to analyze rare disorders. *Nat Methods* 2014; 11 (9): 935–7.

17. 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.

18. Singleton MV, Guthery SL, Voelkerding KV, *et al*. Phevor combines multiple biomedical ontologies for accurate identification of disease-causing alleles in single individuals and small nuclear families. *Am J Hum Genet* 2014; 94 (4): 599–610.

19. Philippakis AA, Azzariti DR, Beltran S, *et al*. The Matchmaker Exchange: a platform for rare disease gene discovery. *Hum Mutat* 2015; 36 (10): 915–21.

20. Berg JS. Exploring the importance of case-level clinical information for variant interpretation. *Genet Med* 2017; 19 (1): 3–5.

21. Weymann D, Laskin J, Roscoe R, *et al*. The cost and cost trajectory of whole-genome analysis guiding treatment of patients with advanced cancers. *Mol Genet Genomic Med* 2017; 5 (3): 251–60.

22. Stark Z, Dashnow H, Lunke S, *et al*. A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. *Eur J Hum Genet* 2017; 25 (11): 1268–72.

23. Baldridge D, Heeley J, Vineyard M, *et al*. The Exome Clinic and the role of medical genetics expertise in the interpretation of exome sequencing results. *Genet Med* 2017; 19 (9): 1040–8.

24. Bland A, Harrington EA, Dunn K, *et al*. Clinically impactful differences in variant interpretation between clinicians and testing laboratories: a single-center experience. *Genet Med* 2018; 20 (3): 369–73.

25. Wilson JR. Fundamentals of ergonomics in theory and practice. *Appl Ergon* 2000; 31 (6): 557–67.

26. Cianci P, Selicorni A. Gestalt diagnosis" for children with suspected genetic syndromes. *Ital J Pediatr* 2015; 41 (Suppl 2): A16.

27. Gurrieri F, Tiziano FD, Zampino G, *et al*. Recognizable facial features in patients with alternating hemiplegia of childhood. *Am J Med Genet A* 2016; 170 (10): 2698–705.

28. Concannon N, Hegarty A-M, Stallings RL, *et al*. Coffin-Lowry phenotype in a patient with a complex chromosome rearrangement. *J Med Genet* 2002; 39: e41.

29. Rosch E, Mervis CB, Gray WD, *et al*. Basic objects in natural categories. *Cogn Psychol* 1976; 8 (3): 382–439.

30. Gentner D, Loewenstein J, Thompson L. Learning and transfer: a general role for analogical encoding. *J Educ Psychol* 2003; 95 (2): 393–408.

31. Köhler S, Schulz MH, Krawitz P, *et al*. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet* 2009; 85 (4): 457–64.

32. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int J Hum Comput Interact* 1995; 7 (1): 57–78.

33. Rosch E, Simpson C, Miller RS. Structural bases of typicality effects. *J Exp Psychol Hum Percept Perform* 1976; 2 (4): 491–502.

34. Smith EE, Shoben EJ, Rips LJ. Structure and process in semantic memory: a featural model for semantic decisions. *Psychol Rev* 1974; 81 (3): 214–41.

35. Hampton JA. Concepts as prototypes In: Ross BH, ed. *Psychology of Learning and Motivation*. San Diego, CA: Elsevier Academic Press; 2006: 79–113.

36. Aerts D, Broekaert J, Gabora L, *et al*. Generalizing prototype theory: a formal quantum framework. *Front Psychol* 2016; 7: 418.

37. Shyr C, Kushniruk A, van Karnebeek CDM, *et al*. Dynamic software design for clinical exome and genome analyses: insights from bioinformaticians, clinical geneticists, and genetic counselors. *J Am Med Inform Assoc* 2016; 23 (2): 257–68.

38. Richards S, Aziz N, Bale S, *et al*. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17 (5): 405–24.