HHS Public Access

Author manuscript

J Allergy Clin Immunol. Author manuscript; available in PMC 2022 August 03.

Published in final edited form as:

J Allergy Clin Immunol. 2022 January ; 149(1): 369–378. doi:10.1016/j.jaci.2021.04.033.

Curation and expansion of Human Phenotype Ontology for defined groups of inborn errors of immunity

Matthias Haimel, PhD[a,b,c,*], **Julia Pazmandi, MSc**[a,b,c,*], **Raú Jiménez Heredia, MSc**[a,b,c], **Jasmin Dmytrus, MSc**[a,b,c], **Sevgi Köstel Bal, MD, PhD**[a,b,c], **Samaneh Zoghi, PhD**[a,b,c], **Paul van Daele, MD**[d], **Tracy A. Briggs, PhD**[e,f], **Carine Wouters, MD**[g,h], **Brigitte Bader-Meunier, MD**[i,j], **Florence A. Aeschlimann, MD**[i,j], **Roberta Caorsi, MD**[k], **Despina Eleftheriou, MD**[l,m], **Esther Hoppenreijs, MD**[n], **Elisabeth Salzer, MD, PhD**[a,b,c,al], **Shahrzad Bakhtiar, MD**[o], **Beata Derfalvi, MD**[p], **Francesco Saettini, MD**[q], **Maaike A. A. Kusters, MD, PhD**[l,m], **Reem Elfeky, MD**[l,m], **Johannes Trück, MD, DPhil**[r], **Jacques G. Rivière, MD**[s,t], **Mirjam van der Burg, PhD**[u,v], **Marco Gattorno, MD**[k], **Markus G. Seidel, MD**[w], **Siobhan Burns, MD**[x], **Klaus Warnatz, MD**[y,z], **Fabian Hauck, MD, PhD**[aa,ab], **Paul Brogan, MD**[l,m], **Kimberly C. Gilmour, PhD**[m], **Catharina Schuetz, MD**[ac], **Anna Simon, MD, PhD**[ad], **Christoph Bock, PhD**[a,c,ae], **Sophie Hambleton, PhD**[af], **Esther de Vries, MD, PhD**[ag,ah], **Peter N. Robinson, MD**[ai], **Marielle van Gijn, PhD**[aj,‡], **Kaan Boztug, MD**[a,b,c,ak,al,‡]

[a]Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna

[b]St Anna Children's Cancer Research Institute (CCRI), Vienna

[c]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna

[d]Department of Clinical Immunology, Erasmus University Medical Center, Rotterdam

[e]NW Genomic Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester

[f]Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, Manchester

[g]Department of Microbiology and Immunology, Immunobiology, KU Leuven, Leuven

[h]Department of Pediatrics, Division of Pediatric Rheumatology, University Hospitals Leuven, Leuven

[i]Pediatric Immuno-Hematology and Rheumatology Unit, Necker Hospital for Sick Children - AP-HP, Paris

[j]Reference Center for Rheumatic, Autoimmune and Systemic Diseases in Children (RAISE), Paris

Corresponding author: Kaan Boztug, MD, Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD) and St. Anna Children's Cancer Research Institute (CCRI), Zimmermannplatz 10, A-1090 Vienna, Austria. kaan.boztug@rud.lbg.ac.at. Or: Marielle Van Gijn, PhD, Department of Genetics, University Medical Center Groningen, Antonius Deusinglaan 1, 9713AV Groningen, The Netherlands. m.e.van.gijn@umcg.nl.

*These authors contributed equally to this work.

‡These authors contributed equally to this work.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

[k]Center for Autoinflammatory Diseases and Immunodeficiency, IRCCS Istituto Giannina Gaslini, Genova

[l]University College London Great Ormond Street Institute of Child Health, London

[m]Department of Immunology, Great Ormond Street (GOS) Hospital for Children NHS Foundation Trust, London

[n]Department of Paediatric Rheumatology, Radboud University Medical Centre, Nijmegen

[o]Department for Children and Adolescents, Division for Stem Cell Transplantation, Immunology and Intensive Care Unit, Goethe University, Frankfurt

[p]Department of Pediatrics, Division of Immunology, Dalhousie University/IWK Health Centre Halifax, Halifax

[q]Pediatric Hematology Department, Fondazione MBBM, University of Milano Bicocca, via Pergolesi 33, Monza

[r]Division of Immunology, University Children's Hospital Zurich, Zurich

[s]Pediatric Infectious Diseases and Immunodeficiencies Unit, Vall d'Hebron Research Institute, Hospital Universitari Vall d'Hebron, Universitat AutÒnoma de Barcelona, Barcelona

[t]Jeffrey Model Foundation Excellence Center, Barcelona

[u]Department of Immunology, University Medical Center Rotterdam, Rotterdam

[v]Laboratory for Pediatric Immunology, Department of Pediatrics, Leiden University Medical Center, Leiden

[w]Research Unit for Pediatric Hematology and Immunology, Division of Pediatric Hemato-Oncology, Department of Pediatrics and Adolescent Medicine, Medical University Graz, Graz

[x]Department of Immunology, UCL Institute of Immunity & Transplantation, Royal Free Hospital NHS Foundation Trust, London

[y]Division of Immunodeficiency, Department of Rheumatology and Clinical Immunology

[z]Center for Chronic Immunodeficiency, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg

[aa]Department of Pediatrics, Dr. von Hauner Children's Hospital

[ab]Munich Centre for Rare Diseases (M-ZSE^{LMU}), University Hospital, Ludwig-Maximilians-Universität München, Munich

[ac]Department of Pediatrics, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Dresden

[ad]Radboudumc Expertise Centre for Immunodeficiency and Autoinflammation (REIA), Department of Internal Medicine, Radboud University Nijmegen Medical Centre, Nijmegen

[ae]Institute of Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna

[af]Immunity and Inflammation Theme, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne

[ag]Tranzo, Tilburg University, Tilburg

[ah]Laboratory for Medical Microbiology and Immunology, Elisabeth-Tweesteden Hospital, Tilburg

[ai]The Jackson Laboratory for Genomic Medicine, Farmington

[aj]Department of Genetics, University Medical Center Groningen, Groningen

[ak]Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna

[al]St Anna Children's Hospital, Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna

Abstract

Background: Accurate, detailed, and standardized phenotypic descriptions are essential to support diagnostic interpretation of genetic variants and to discover new diseases. The Human Phenotype Ontology (HPO), extensively used in rare disease research, provides a rich collection of vocabulary with standardized phenotypic descriptions in a hierarchical structure. However, to date, the use of HPO has not yet been widely implemented in the field of inborn errors of immunity (IEIs), mainly due to a lack of comprehensive IEI-related terms.

Objectives: We sought to systematically review available terms in HPO for the depiction of IEIs, to expand HPO, yielding more comprehensive sets of terms, and to reannotate IEIs with HPO terms to provide accurate, standardized phenotypic descriptions.

Methods: We initiated a collaboration involving expert clinicians, geneticists, researchers working on IEIs, and bioinformaticians. Multiple branches of the HPO tree were restructured and extended on the basis of expert review. Our ontology-guided machine learning coupled with a 2-tier expert review was applied to reannotate defined subgroups of IEIs.

Results: We revised and expanded 4 main branches of the HPO tree. Here, we reannotated 73 diseases from 4 International Union of Immunological Societies–defined IEI disease subgroups with HPO terms. We achieved a 4.7-fold increase in the number of phenotypic terms per disease. Given the new HPO annotations, we demonstrated improved ability to computationally match selected IEI cases to their known diagnosis, and improved phenotype-driven disease classification.

Conclusions: Our targeted expansion and reannotation presents enhanced precision of disease annotation, will enable superior HPO-based IEI characterization, and hence benefit both IEI diagnostic and research activities.

Keywords

HPO; ontology; phenotype; rare diseases; inborn errors of immunity; immunodeficiencies; disease classification; diagnostic support; patient matching; genetic analysis

---

Rare and undiagnosed diseases pose challenges for affected patients, clinicians, and researchers working to improve diagnostic and therapeutic approaches. Because of the rarity, clinicians often see only a few patients with specific rare phenotypes throughout their

careers, leading to considerable diagnostic delay.[1] Genetic research on rare diseases often relies on single pedigrees or a few patients, leaving many patients undiagnosed.[1] Compiling a cohort of patients—so-called patient matching—is often crucial to gain insight into the phenotypic spectrum, natural/clinical history of the disease, and adequate monitoring and treatment strategies. The rare disease community has recognized these challenges and established tools enabling efficient data sharing across institutions and borders, including genetic data exchange through the Matchmaker Exchange platform[2] to solve undiagnosed exomes and genomes.[3] These platforms, however, are highly dependent on accurately phenotyped and categorized patients and standardized disease classifications.

To date, several nomenclatures and reference systems for diseases have been developed.[4,5] In parallel, ontologies were established to provide a more systematic, hierarchical classification of diseases.[6,7] However, these nomenclatures group patients by disease label and do not describe the underlying phenotypic features. Consequently, clinical features, laboratory measurements, and anatomical and functional phenotypes of patients are often described with variable quality and specificity, which hampers patient matching, diagnostic efficiency, genetic variant prioritization in diagnostic pipelines, and global data exchange.

Given these challenges and the need for accurate, standardized phenotyping, the Human Phenotype Ontology (HPO) system was conceptualized and published with initial terminology in 2008.[8,9] To date, HPO provides the most comprehensive deep phenotyping resource for rare diseases for clinicians, researchers, bioinformaticians, and electronic health record systems in the world. HPO is used in many projects including the 100,000 Genomes Project, the NIH Undiagnosed Disease Program and Network, the Undiagnosed Diseases Network International, RD-CONNECT, and SOLVE-RD.[1,10–13] HPO is a community-based tool and is increasingly adapted as the standard to describe phenotypic abnormalities for everyday use.[14] Each term in HPO describes a distinct phenotypic feature (eg, lymphadenopathy, HP:0002716), and the HPO tree structure allows similarity measures between patient phenotypes. HPO contains more than 200,000 phenotypic annotations for hereditary diseases, of which 2,120 are considered rare diseases. Inborn errors of immunity (IEIs) form a subgroup of these rare diseases. Clinical experts in IEI agree that a major barrier to the adoption of HPO terminology is that it has not been used widely for IEIs. This is partly due to the lack of disease-specific HPO terms to describe patients with IEI.[15] Adequate depiction of the complex clinical and immunologic phenotypes of IEI disease entities with HPO terms would allow discrimination between heterogeneous groups of IEIs. Illustrating the lack of terms, in 2017, HPO contained more than 11,000 terms, of which 5,000 terms have been applied to the musculoskeletal system, with only 1,000 terms related to IEIs.[9,15] In addition, the phenotypic annotation of IEIs often includes results of specific immunologic assays, which pose a challenge to accurately reflect in HPO terms.[15] Because of the lack of specific HPO terms depicting results of laboratory assays, often a nonspecific broader term is used for the annotation of IEIs. Therefore, HPOs are currently not specific enough to be used for genetic analysis and diagnostic aid for IEIs. In a study addressing the clinical efficacy of genetic testing in IEI, bioinformatics tools using existing HPO terms missed the disease-causing gene in 37% of the patients with known monogenic disorders.[16] In this study, we set out to improve HPO terminology for IEIs by applying established bioinformatic methodologies coupled with expert review. The aims of this project were

therefore to (1) systematically review existing HPO terms for IEIs, (2) revise ontology structures, (3) add missing terms, as well as (4) reannotate existing IEIs with HPO terms, to collectively enable systematic use of HPO by the IEI community.

METHODS

Spearheaded by the European Reference Network on Rare Primary Immunodeficiency, Autoinflammatory and Autoimmune diseases (ERN-RITA) and the European Society for Immunodeficiencies (ESID), we set up working groups comprising members of the participating immunodeficiency societies to revise and expand HPO terms for IEIs. Three workshops, numerous teleconferences, and joint task forces took place over the span of 2 years, with more than 30 participants including expert clinicians, geneticists, researchers working on IEIs, and bioinformaticians. All participating clinicians and geneticists identified through the ERN-RITA, ESID, and the International Society of Systemic Autoinflammatory Diseases are established experts in their fields from different European countries and North America. Additional scientific support provided the indispensable bioinformatics expertise.

Establishment of working structure

A remote working structure (detailed in this article's Methods section in the Online Repository at www.jacionline.org) was launched to address gaps in the HPO tree and in the annotation of IEI diseases.

Expansion and restructuring of disease-related branches of the HPO tree

Disease-specific HPO restructuring was discussed within 4 working groups. Each group focused on a different HPO branch; the suggested changes were agreed on among all participants. Differences between centers and countries in the use of terms and definitions were highlighted during the face-to-face workshops. The results were summarized electronically in Excel documents or pictures and flipchart drawings by the main coordinators before being submitted to HPO. The full list of restructured tree elements and new submitted HPO terms is detailed in the Document S1 in this article's Online Repository at www.jacionline.org. In addition, missing terms describing pulmonary and gastrointestinal complications of primary antibody deficiency (PAD) were discussed during teleconferences and thereafter submitted to update the HPO ontology. A list of HPO resources can be found in this article's Methods section in the Online Repository.

Standardized reannotation of rare, genetically diagnosed diseases

A 4-step process was developed for a standardized reannotation effort across working groups and to consistently annotate IEIs (spanning more than 300 different diseases in Online Mendelian Inheritance in Man) with HPO terms (see Fig 1). Because IEIs represent a large and heterogeneous group of rare diseases, we here decided to selectively focus on defined subgroups of IEI to test the feasibility and usefulness of such an endeavor. First, publications were collected by experts for each disease within the subgroups (minimum of 2 articles per disease), representing key phenotypic presentation(s) of the specific disease. In the second step, HPO terms were extracted from the provided publications for each disease using machine learning[17] (explained in detail in this article's Methods section in the Online

Repository) and summarized into Excel documents. Third, a 2-tier expert review evaluated the text-mined terms, suggested additional terms if required, and the responsible working group agreed (defined as at least 80% agreement among group experts) on the final HPO annotations for each disease. Fourth, the validated terms were submitted to HPO. Document S2 in this article's Methods section in the Online Repository contains the reannotated diseases, and the list of reannotated terms for each disease is available in Document S3 in this article's Methods section in the Online Repository.

Standardized reannotation of genetically undiagnosed diseases

The methods above were specifically designed for application in (very) rare diseases, where the number of patients and therefore the described phenotypic spectrum and clinical presentation is sparse. In case of diseases and disease groups where an adequate amount of patient and phenotype data were available, in addition to a True/False annotation, the frequency of each phenotypic item was assessed. The frequencies correspond to the following representation in patients: common = Frequent (79%-30%); sometimes = Occasional (29%-5%); rare = 5 Very rare (<4%-1%).

Patient cohort

We randomly selected 30 patients who harbored a genetic diagnosis in one of the reannotated diseases from a large pediatric referral center research database. Clinical summaries of these patients before genetic diagnosis were retrieved by an expert clinician. The clinical summaries were parsed and HPO terms were extracted using machine learning as described in this article's Methods section in the Online Repository.

HPO information content measures, and disease patient similarity measures

Information content of all HPO terms was assessed with the *R* package ontologyIndex v2.5.[18] The phenotypic similarity of diseases and patients before and after reannotation was compared using the R package ontologySimilarity v2.3.[18] The Euclidean distances between the diseases were computed on the basis of similarity measures, clustered with hierarchical clustering and visualized with ggtree using the R packages ggtree[19] and ape v5.2.[20]

A detailed description including the data processing pipeline and tools is available in this article's Methods section in the Online Repository.

RESULTS

Systematic evaluation and expansion of the HPO structure and terms relevant to IEIs

Our approach has resulted in the restructuring of 4 main branches of the HPO tree, namely (1) abnormality of the immune system (HP:0002715), (2) abnormality of metabolism/homeostasis (HP:0001939), (3) abnormality of the integument (HP:0001574), and (4) abnormality of the cardiovascular system (see Fig 2, A, and Document S1 in this article's Online Repository at www.jacionline.org). Together, this revision prompted the replacement/restructuring of 67 terms, and the addition of 57 new terms to the HPO tree, among them "recurrent fever," "unusual infections," "IgG levels in blood" (see Fig 2, B,

and comprehensive list in Documents S1 and S2 in this article's Online Repository at www.jacionline.org).

Directed expansion of PAD terms

Overall, the PAD working group focused on replacing broad and nonspecific terms with terms that describe phenotypes in more detail and accuracy (eg, "partially absent total IgG/IgA/IgM in blood" and "(near) absent total IgG/IgA/IgM in blood" instead of "hypogammaglobulinemia") (Fig 2, B). In addition, we proposed that the full detailed spectrum of specific antibody as well as IgG-subclass deficiencies be described by separate HPO terms. For example, we described individual terms related to "decreased specific antibody response to vaccination in blood" divided according to the response to different types of vaccination (protein, protein-conjugated polysaccharide, and unconjugated polysaccharide).

Standardized reannotation of rare, genetically diagnosed IEIs

We started by a systematic review of 4 disease categories of the International Union of Immunological Societies (IUIS) classification of IEIs, as proof of concept: diseases affecting cellular and humoral immunity (IUIS Table 1), diseases of immune dysregulation (IUIS Table 4), autoinflammatory disorders (IUIS Table 7), and genetically undiagnosed predominantly antibody deficiencies (IUIS Table 3), detailed in Table E1 and Document S3 in this article's Online Repository at www.jacionline.org. As a first step, we assessed the already available HPO annotation for each disease in the v2019-06-03 HPO release (see this article's Methods section in the Online Repository). We found that 15% of diseases considered (11 of 73 diseases in total) did not have any associated HPO terms (Fig 3, A). Overall, we found that on average 13.3 phenotype terms were available per disease (Fig 3, B), later referred to as "existing terms."

The text-mining and evaluation process was separated into 4 steps shown in Fig 3, C. We have first focused on the reannotation of 72 genetically diagnosed IEIs, and genetically undiagnosed PADs. For genetically diagnosed IEIs, text mining was based on 162 expert-curated articles, on average 2.57 articles per disease (Fig 3, D). This resulted in 4,517 extracted phenotype terms, 66.42 terms per disease (Fig 3, E). Of these terms, 3,242—or 71% per disease (47.67 of 66.42)—were accepted as correctly attributed terms by the expert reviewers (Fig 3, F). Expert suggestions added up to 529 additional HPO terms, in addition to the existing and text-mined terms.

After reannotation, a mean of 63.1 terms were available for each disease, resulting in a 4.7-fold gain in the number of available annotations (Fig 3, G). The mean information content as measured by the overall frequency of terms in each disease's annotations has increased from 6.17 to 8.3 (Fig 3, H) after reannotation.

The new annotation of diseases consisted mainly of text-mined terms (70.6%) (Fig 3, I), followed by already existing terms (9.3%) and additional suggestions by experts (9.3%; adding a further 5.2 additional terms per disease) (see Document S3 in this article's Online Repository).

Standardized reannotation of genetically undiagnosed PADs

PADs form a heterogeneous group, and most PADs do not (as yet) have a genetic diagnosis. We collected articles describing the heterogeneous PADs related to common variable immunodeficiency disorders, agammaglobulinemia, selective IgM deficiency, selective IgA deficiency, IgG-subclass deficiency, specific antibody deficiency, and unclassified antibody deficiency subgroups. In total, 541 terms were text mined from these articles, many of these in more than 1 PAD subgroup, and 245 of these terms (45.2%) were annotated as correctly associated to the respective PAD subgroup by the expert reviewers (Fig 3, J). Of these 245 terms, the experts annotated 16.3% as commonly found in PADs, 48.97% as sometimes associated (albeit less commonly), and 34.7% as rarely associated with PAD (Fig 3, K).

Patient-disease matching

We set out to showcase the efficacy of our reannotation effort by highlighting the potential diagnostic impact of optimized disease annotation. To do this, we have selected 30 clinical cases from a large immunology referral center research database (see Online Repository Document S3). HPO terms were matched to patient phenotypes by experts from the clinical synopsis, and the phenotypic similarity to all HPO-annotated diseases was calculated on the basis of these selected patient HPO terms (Fig 4, A), as illustrated by a concrete clinical example of a patient with tumor necrosis factor receptor–associated periodic syndrome (Fig 4, B). Overall, we show a significant improvement by 47% in the specificity of patient phenotype matching to correct diagnosis (from 0.49 to 0.72; $P = 1.8 \times 10^{-07}$; Fig 4, C), and a significantly better ranking of the correct clinical diagnosis across all possible diseases after reannotation: in most cases, the correct diagnosis was in the top 10 of matched diseases (Fig 4, D) after reannotation, and the rank of the correct diagnosis for individual patients was highly significantly improved, from a mean of 285 to 19 (14.9-fold improvement; $P = 9.1 \times 10^{-07}$; Fig 4, E*).*

Phenotype-driven disease classification

We tested the efficacy of our approach in selecting biologically and clinically meaningful phenotypes by assessing the HPO-based phenotypic similarity of diseases before and after reannotation. In particular, we assessed whether the similarity was greater within or between IUIS clinically defined groups. We found that the phenotype-driven disease classification after reannotation has resulted in a clustering more in concordance with the IUIS-based clinical classification (see Fig 5, A and B).

DISCUSSION

Unified data standards, consistent classification, and robustly verified clinical data are vital pillars supporting diagnostic pipelines and data-driven research. Although databases and vocabularies that aim to provide accurate phenotypic descriptions exist,[5–9] there are still major gaps in the depiction of IEIs in these data sets. Here, we used a cross-community collaboration to review, expand, and improve the depiction of IEIs in HPO, and reannotate IEIs with HPO terms. We reviewed 4 separate branches of the HPO tree and submitted 57 new and expanded HPO terms, most of which are now included in the official HPO data set. We introduced a semi-automated reannotation pipeline, which combines ontology-guided

machine learning and a 2-tier expert review to reannotate 4 main categories of IEIs. The basis of the ontology-guided machine learning was the expert-curated list of articles (162 in total), which was submitted to the PanelApp[21] to serve as a public resource. The text-mined phenotypes were subjected to expert review to confer face validity or refute the putative new HPO terms. IEIs and their current HPO terms covered by the working groups were scrutinized in-depth, resulting in high-quality annotations. Overall, we have achieved a 4.7-fold gain in the number of HPO terms annotating each disease. These annotations included unspecific (frequently annotated) as well as specific (less frequently annotated) HPO terms holding less and more information content, respectively. Combined, the mean information content increased from 6.17 to 8.3.

Each reannotated disease showed an increase in information content and a quantitative gain in the number of available HPO terms. Through patient-disease matching and disease-similarity examples, we illustrated that these gains and increases translated to significant qualitative improvement in patient-disease matching in an independent cohort of patients with IEI (Fig 4), and phenotype-driven classification of IEIs that more closely resembles clinical consensus (Fig 5). Although neither of these measures are systematic assessments of global patient-disease matching and disease-similarity comparisons, they highlight that there is considerable benefit by the revision of specific subclasses of diseases. Once a near-complete HPO phenotype reannotation of almost all IEIs is available, it will be intriguing to assess how well patients with genetic diagnoses match reannotated Online Mendelian Inheritance in Man (OMIM) diseases in a clinical setting, how patient matching to genetic diagnosis is transformed, and whether these changes ultimately lead to an earlier diagnosis. Finally, once a detailed and accurate phenotypic description is available for all IEIs, identification phenotype-driven patient subgroups will be common practice, and a more objective entirely phenotype-driven classification and ontology of IEIs can become a reality.

Accurate phenotypic description of patients holds promise for diagnostic utility and for the discovery of novel diseases. Phenotype-driven genetic diagnostic tools now exist, but their full clinical potential is hampered by the lack of complete phenotypic descriptions for most types of IEIs. Phenotips[22] is a free and open source software for collecting and analyzing phenotypic information of patients with genetic disorders that is widely used in the rare disease community. Tools such as Exomiser use HPO terms to annotate and to prioritize potentially casual variants.[23] New integrative “omics” approaches and the analysis of large-scale data with artificial intelligence will allow us to go from a one-size-fits-all to a more personalized medicine, including in IEIs. We see the potential to integrate the richer phenotyping of previously undiagnosed groups of patients with IEI with available sequencing data to accelerate disease gene discovery and at the same time increase the diagnostic rate in new patients.[24]

Novel disease-gene or phenotype associations depend on sufficient numbers of cases as well as a control cohort of comparable quality. Cross-institute and cross-country collaborations for cohorts of undiagnosed, but well-phenotyped patients could shed light on novel disease-causing genes of the immune system. Trusted and accepted data and information sharing platforms are already being developed[13,22] to provide robust and sufficiently granular HPO terms as a standardized way of phenotyping patients. Electronic health records[25]

could facilitate the transfer of HPO terms by integrating with available sharing platforms. Capturing HPO annotations of novel rare diseases or cases is an ongoing challenge for a complete disease representation. Thus, it is important that alongside of updating the official IUIS classification, HPO descriptions of disorders are curated once every several years. We suggest a community effort for such regular reviews of HPO regarding IEIs, such as a team of experts, part of big international groups of clinicians such as ESID or ERN-RITA, the Clinical Immunology Society (CIS), or other similar organizations. Publication standards that require the submission of HPO annotations upfront would greatly improve this process.

Once phenotyped patients are available, robust and global approaches are accessible[2] to find phenotypically similar cases. These comparisons are performed by advanced machine learning algorithms. However, machine learning can also be a very powerful tool to automate the identification of relevant phenotype information in publications or clinical notes. We applied an ontology-guided machine learning tool to support the annotation of diseases and explored the full spectrum of terms—from very relevant to not relevant at all. The same process can be applied to unstructured clinical notes to accelerate in-depth annotation of patients. For patients with electronic health records,[25] abnormal clinical values can automatically be translated into HPO codes[26] for a more precise diagnostic application and integrated with sharing platforms as mentioned before. The foundation of these comparisons is an ontology with a comprehensive set of terms, which is widely used.

Because there is currently no criterion standard on how to perform an expert-based review of ontologies, guidance on annotating diseases with HPO phenotypes can vary between diseases, disease classes, and centers. IEIs are rare diseases, and often there are only a few patients described (sometimes only 1 kindred in case of ultra-rare diseases). Therefore, the depth of currently available published phenotypes is at times limited. The low number of patients and insufficient depth of available phenotypes bring up a question as to which diseases to include in phenotyping exercises of this nature. On the one hand, focusing on IEIs that are commonly accepted, with multiple patients diagnosed and well described by multiple researchers, can increase the depth of phenotyping. However, this approach excludes at least 10% of IEIs (the ultra-rare diseases). On the other hand, an all-inclusive approach including every disease systematically means that we rely on sparsely phenotyped patients and perhaps insufficient data for ultrarare disorders. A warning of accuracy by indicating the frequency of each phenotype for diseases could soon be possible, with the addition of phenotype frequency to the HPO data set, an expansion that is currently work in progress. This implies the need for a responsive system, capable of assimilating new phenotypic information as the pool of confidently diagnosed patients increases.

Our ongoing approach aims to address these gaps for IEIs and to provide an ontology that is practical, useful, and as complete as possible. However, the existence of a well-built ontology and the awareness of clinicians and researchers itself does not guarantee a shift in the community to fully adapt a standardized phenotyping approach. Our approach raised awareness regarding the concept and importance of HPO among the IEI community. Moreover, the process made the participating clinicians aware of the available terms and highlighted where these were lacking. Moving forward, it is very important that official entities adopt HPO terms as the unified means of patient phenotyping. We hypothesize that

as soon as the widely used registries such as the Undiagnosed Disease Network[11] or the IUIS[27] use HPO to refer to phenotypic annotation, this will propel the IEI field toward adopting HPO as the main nomenclature for phenotyping patients with IEI. One promising move in this direction is the recent expansion of the ESID registry working definitions for the clinical diagnosis of IEIs,[28] which derives HPO terms from OrphaNet using the ORDO Ontological Module (HOOM) platform,[29] prompted by our HPO initiative.

Conclusions

Our work reviewed and expanded the phenotypic depiction of multiple subclasses of IEIs, and to our knowledge, this initiative is the first endeavor of its kind with the aim of standardizing IEI phenotypes. Our semiautomated annotation-based approach is scalable to include all IEIs as illustrated herein. We propose our reannotation approach as a blueprint for systematic HPO (re) annotation for additional immunologic and nonimmunologic diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the European Research Council (ERC Consolidator Grant no. 820074 “iDysChart” to K.B.), by the Austrian Science Fund (FWF) project P 29951-B30 (to K.B.), and by funding from the European Union’s Horizon 2020 research and innovation program under the EJP RD COFUND-EJP N° 825575 (to C.B.). Additional financial support for the workshops was granted by the Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, the European Reference Network on Rare Primary Immunodeficiency, Autoinflammatory and Autoimmune Diseases (ERN-RITA), and the European Society for Immunodeficiencies (ESID).

Abbreviations used

ESID	European Society for Immunodeficiencies
HPO	Human Phenotype Ontology
IEI	Inborn error of immunity
IUIS	International Union of Immunological Societies
PAD	Primary antibody deficiency

REFERENCES

1. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. Genet Med 2012;14:51–9. [PubMed: 22237431]
2. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat 2015;36:915–21. [PubMed: 26295439]
3. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 2015;36:928–30. [PubMed: 26220891]

4. Hernandez-Ibarburu G, Perez-Rey D, Alonso-Oset E, Alonso-Calvo R, de Schepper K, Meloni L, et al. ICD-10-CM extension with ICD-9 diagnosis codes to support integrated access to clinical legacy data. Int J Med Inform 2019;129:189–97. [PubMed: 31445254]
5. Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). Hum Mutat 2011;32:564–7. [PubMed: 21472891]
6. Pavan S, Rommel K, Mateo Marquina ME, Höhn S, Lanneau V, Rath A. Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. PLoS One 2017;12:e0170365. [PubMed: 28099516]
7. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res 2019;47:D955–62. [PubMed: 30407550]
8. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008;83:610–5. [PubMed: 18950739]
9. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res 2017;45:D865–76. [PubMed: 27899602]
10. Ramoni RB, Mulvihill JJ, Adams DR, Allard P, Ashley EA, Bernstein JA, et al. The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease. Am J Hum Genet 2017;100:185–92. [PubMed: 28157539]
11. Taruscio D, Groft SC, Cederroth H, Melegh B, Lasko P, Kosaki K, et al. Undiagnosed Diseases Network International (UDNI): white paper for global actions to meet patient needs. Mol Genet Metab 2015;116:223–5. [PubMed: 26596705]
12. Gall T, Valkanas E, Bello C, Markello T, Adams C, Bone WP, et al. Defining disease, diagnosis, and translational medicine within a homeostatic perturbation paradigm: the National Institutes of Health Undiagnosed Diseases Program Experience. Front Med (Lausanne) 2017;4:62. [PubMed: 28603714]
13. Thompson R, Johnston L, Taruscio D, Monaco L, Béroud C, Gut IG, et al. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med 2014;29:S780–7. [PubMed: 25029978]
14. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 2019;47:D1018–27. [PubMed: 30476213]
15. Chinn IK, Chan AY, Chen K, Chou J, Dorsey MJ, Hajjar J, et al. Diagnostic interpretation of genetic studies in patients with primary immunodeficiency diseases: a working group report of the Primary Immunodeficiency Diseases Committee of the American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol 2020;145:46–69. [PubMed: 31568798]
16. Rae W, Ward D, Mattocks C, Pengelly RJ, Eren E, Patel SV, et al. Clinical efficacy of a next-generation sequencing gene panel for primary immunodeficiency diagnostics. Clin Genet 2018;93:647–55. [PubMed: 29077208]
17. Arbabi A, Adams DR, Fidler S, Brudno M. Identifying clinical terms in medical text using ontology-guided machine learning. JMIR Med Inform 2019;7:e12596. [PubMed: 31094361]
18. Greene D, Richardson S, Turro E. ontologyX: a suite of R packages for working with ontological data. Bioinformatics 2017;33:1104–6. [PubMed: 28062448]
19. Yu G, Lam TT-Y, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using Ggtree. Mol Biol Evol 2018;35:3041–3. [PubMed: 30351396]
20. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 2019;35:526–8. [PubMed: 30016406]
21. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet 2019;51:1560–5. [PubMed: 31676867]
22. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. Pheno-Tips: patient phenotyping software for clinical and research use. Hum Mutat 2013;34:1057–65. [PubMed: 23636887]

23. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc 2015;10:2004–15. [PubMed: 26562621]

24. Westbury SK, Turro E, Greene D, Lentaigne C, Kelly AM, Bariana TK, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. Genome Med 2015;7:36. [PubMed: 25949529]

25. Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The use of FHIR in digital health– a review of the scientific literature. Stud Health Technol Inform 2019;267:52–8. [PubMed: 31483254]

26. Shefchek KA, Harris NL, Gargano M, Matentzoglu N, Unni D, Brush M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res 2020;48:D704–15. [PubMed: 31701156]

27. Bousfiha A, Jeddane L, Picard C, Ailal F, Bobby Gaspar H, Al-Herz W, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. J Clin Immunol 2018;38:129–43. [PubMed: 29226301]

28. Seidel MG, Kindle G, Gathmann B, Quinti I, Buckland M, van Montfrans J, et al. The European Society for Immunodeficiencies (ESID) Registry Working Definitions for the Clinical Diagnosis of Inborn Errors of Immunity. J Allergy Clin Immunol Pract 2019;7:1763–70. [PubMed: 30776527]

29. Gasteiger LM, Robinson PN, Pazmandi J, Boztug K, Seppänen MRJ, Seidel MG. Supplementation of the ESID registry working definitions for the clinical diagnosis of inborn errors of immunity with encoded human phenotype ontology (HPO) terms. J Allergy Clin Immunol Pract 2020;8:1778. [PubMed: 32389282]

Key messages

- HPO is a robust resource for supporting IEI diagnostics and genetics with adequate ontology breadth and disease annotation depth.
- Following systematic reannotation of IEIs, the HPO-based phenotype-driven classification improved and now closely resembles clinical consensus.
- Significant increase in matching patients to the correct diagnoses is achieved by systematic reannotation of IEIs.

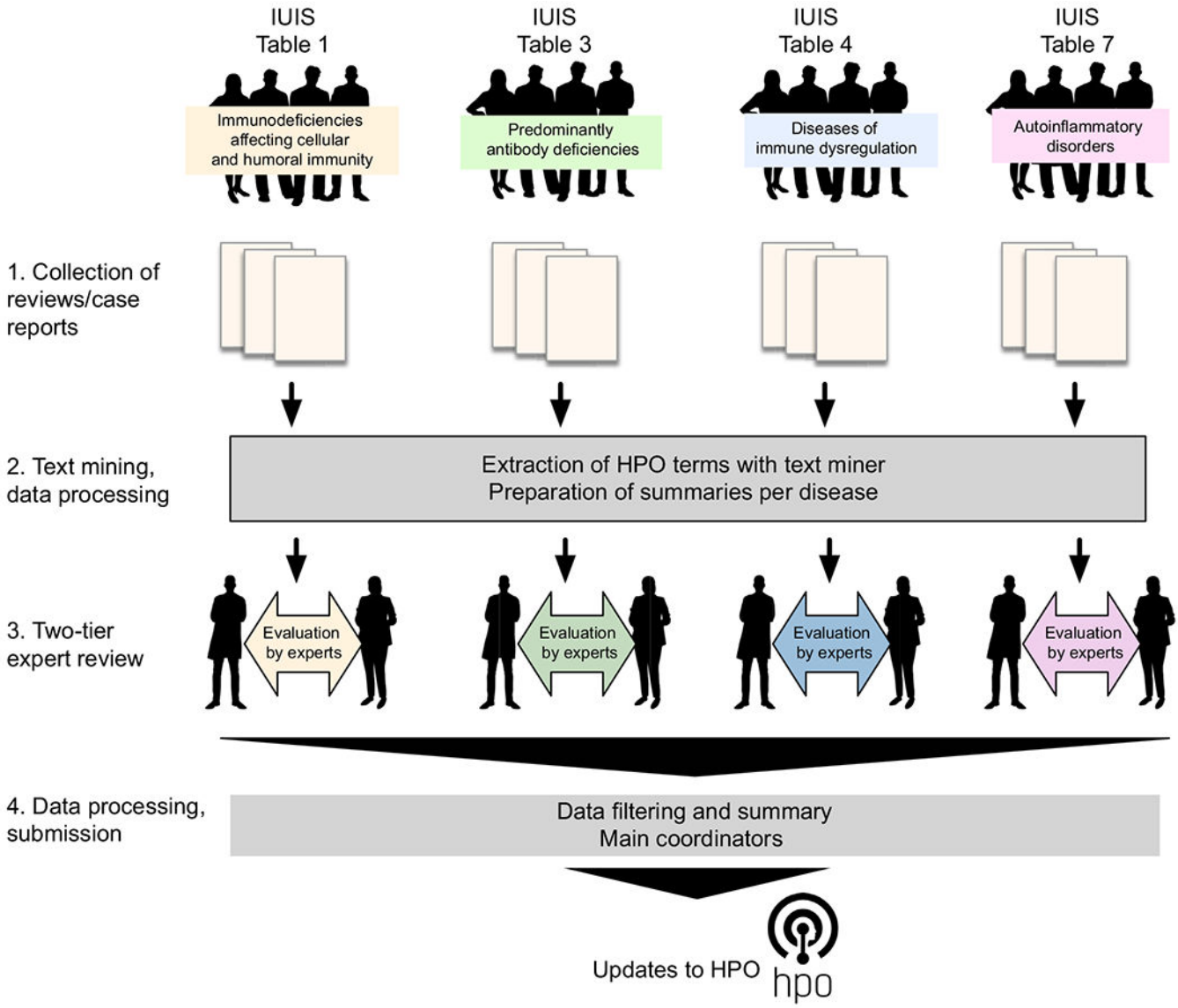

FIG 1.

Pipeline for standardized reannotation of IEI diseases. First, scientific publications were collected by experts for each disease within the subgroups. Second, HPO terms were extracted from the provided publications for each disease using machine learning and summarized into Excel documents. Third, a 2-tier expert review evaluated the text-mined terms, suggested additional terms if required, and the responsible working group agreed on the final HPO annotations for each disease. Fourth, data were collated, and the agreed terms were submitted to HPO.

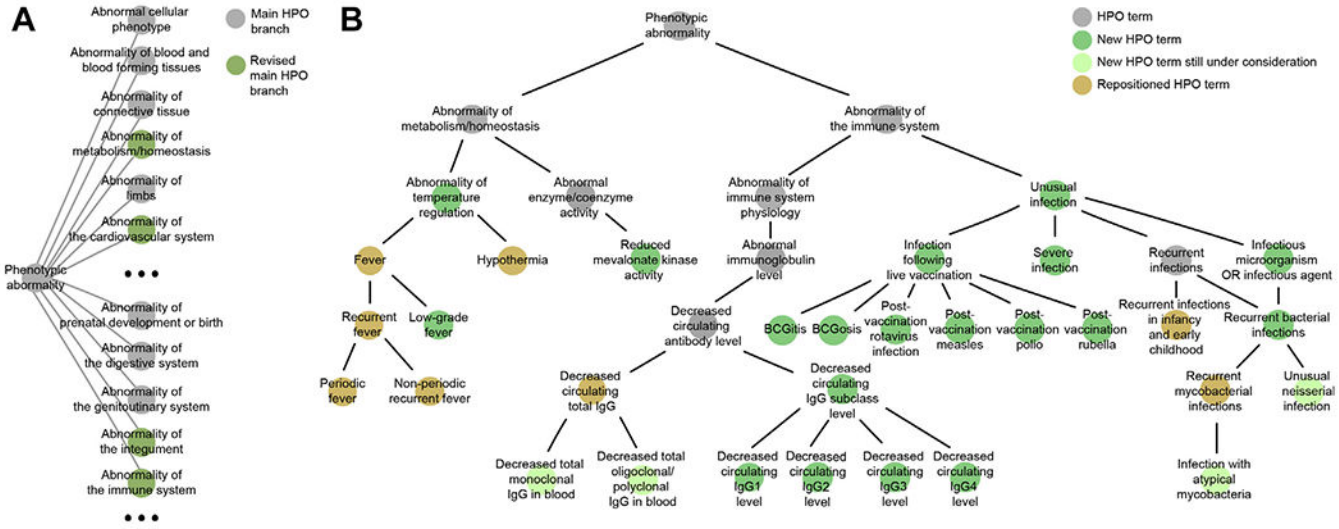

FIG 2.

Revision and expansion of the HPO tree. **A,** Schematic representation of the restructuring of the HPO tree. Main branches of the HPO tree where restructuring was performed are marked with light green. **B,** “Abnormality of temperature,” “Abnormality of immunoglobulin level,” and “Unusual infections” as examples of revised branches of the HPO tree. New additions and suggestions are marked with green, and repositioned terms are marked with yellow.

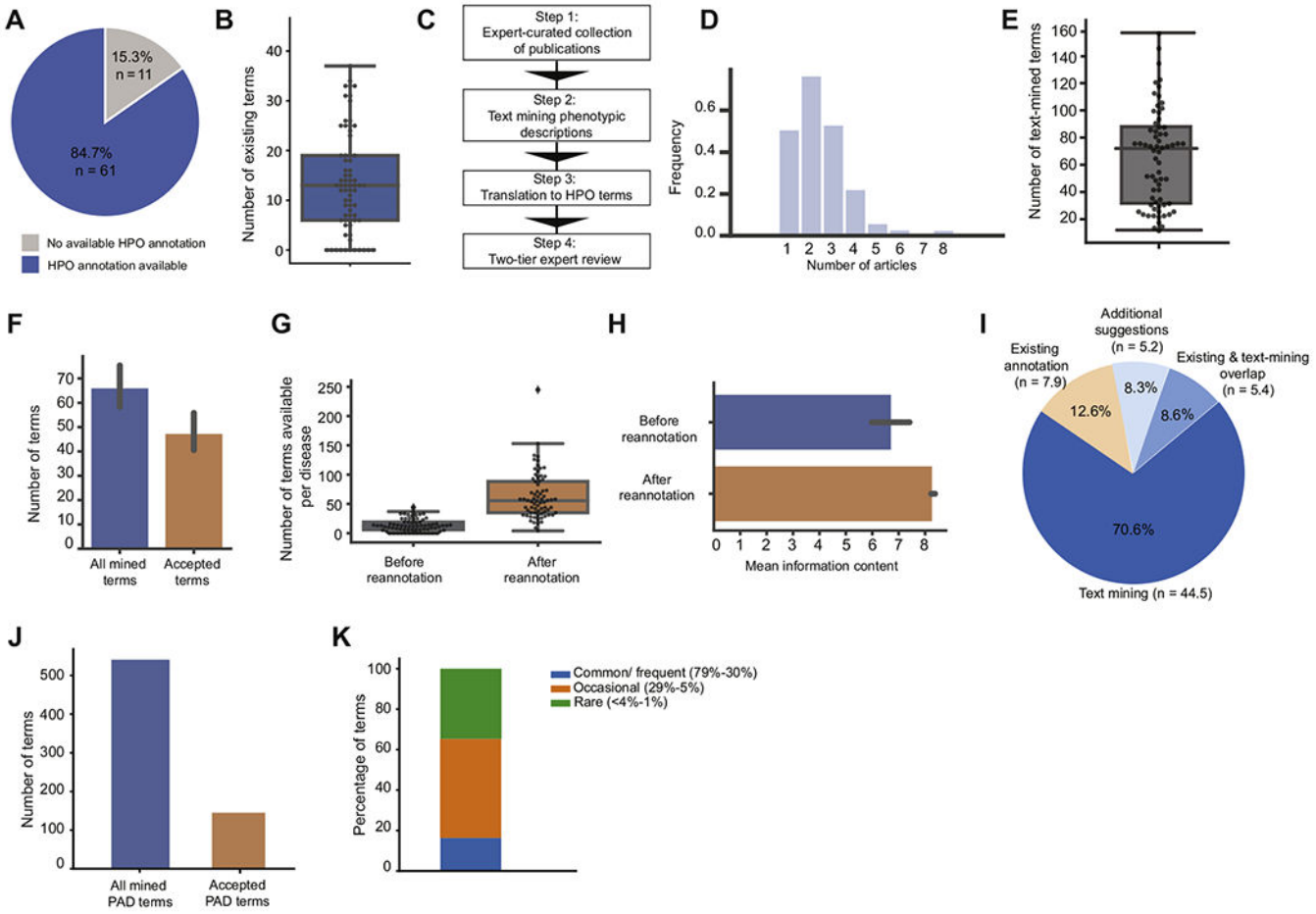

FIG 3.

Results of disease reannotation. **A,** HPO annotation availability in the subset of 72 diseases. **B,** Distribution of number of available HPO terms per disease. **C,** Pipeline for the reannotation process. **D,** Distribution of the number of articles used per disease for the reannotation pipeline. **E,** Number of mined terms per disease. Each dot represents a disease. **F,** All mined vs all accepted terms. **G,** Number of available terms per disease before and after reannotation. Each dot represents a disease. **H,** Mean information content available per disease before and after reannotation. **I,** The aggregate mean annotation per disease after reannotation. **J,** All text-mined terms from PAD publications. **K,** Frequency distribution of different PAD terms according to the experts.

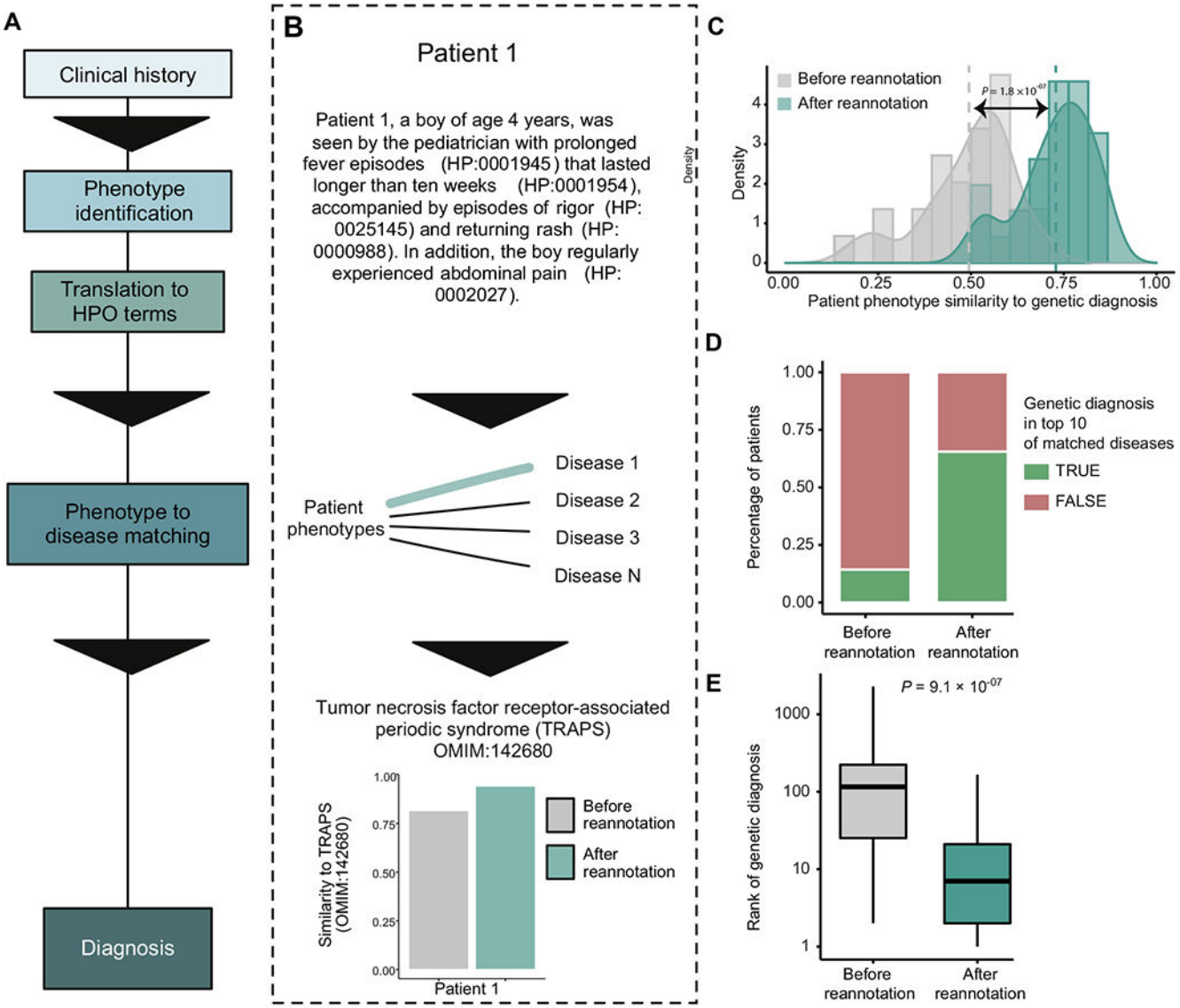

FIG 4.

Patient-disease matching. **A,** Schematic overview of the different steps of patient-to-disease matching. First, the phenotypes were identified in a patient's clinical history. Second, these phenotypes were translated to HPO terms. Finally, patient phenotype to disease matching was measured by Lin similarity. **B,** Matching patient 1 to a diagnosis. **C,** Similarity of patients in patient cohort to genetic diagnosis before and after reannotation. **D,** The rank of correct clinical diagnosis more often is in the top 10 of matched diseases after reannotation. **E,** Improvement of ranks of clinical diagnosis before and after reannotation. Significance was assessed by Student *t* test.

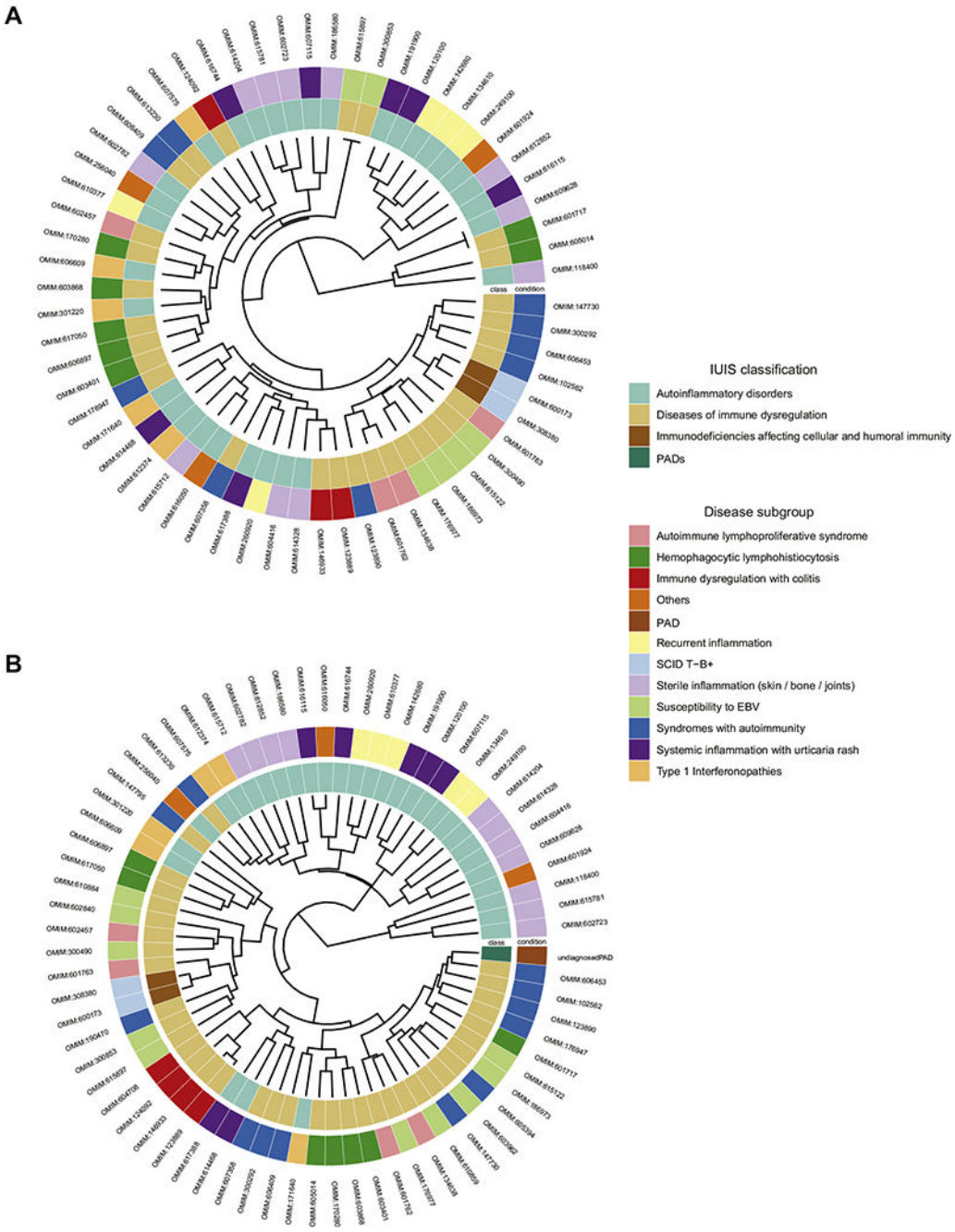

FIG 5.

Phenotypic similarity of diseases before and after reannotation. Diseases are annotated with the IUIS disease group (inner circle), subgroup (outer circle), and OMIM identifier. **A,** Clustering of diseases based on phenotypic similarity before reannotation. **B,** Clustering of diseases based on phenotypic similarity after reannotation. *OMIM,* Online Mendelian Inheritance in Men; *SCID,* severe combined immunodeficiency.