Lawrence Berkeley National Laboratory

Title

The Unified Phenotype Ontology : a framework for cross-species integrative phenomics.

Permalink

https://escholarship.org/uc/item/88j9p90s

Journal

Genetics, 229(3)

Authors

Matentzoglu, Nicolas

Bello, Susan

Stefancsik, Ray

et al.

Publication Date

2025-03-17

DOI

10.1093/genetics/iyaf027

Peer reviewed

OXFORD GENETICS

The Unified Phenotype Ontology : a framework for cross-species integrative phenomics

Nicolas Matentzoglu [ID],[1,*,†] Susan M. Bello,[2,†] Ray Stefancsik,[3,†] Sarah M. Alghamdi,[4] Anna V. Anagnostopoulos,[2] James P. Balhoff,[5] Meghan A. Balk,[6] Yvonne M. Bradford,[7] Yasemin Bridges,[8] Tiffany J. Callahan,[9] Harry Caufield,[10] Alayne Cuzick [ID],[11] Leigh C. Carmody,[2] Anita R. Caron,[3] Vinicius de Souza,[3] Stacia R. Engel [ID],[12] Petra Fey,[13] Malcolm Fisher [ID],[14] Sarah Gehrke,[15] Christian Grove,[16] Peter Hansen [ID],[17] Nomi L. Harris,[10] Midori A. Harris,[18] Laura Harris [ID],[3] Arwa Ibrahim,[3] Julius O.B. Jacobsen,[8] Sebastian Köhler,[19] Julie A. McMurry,[15] Violeta Munoz-Fuentes,[20] Monica C. Munoz-Torres [ID],[21] Helen Parkinson,[3] Zoë M. Pendlington,[3] Clare Pilgrim,[18] Sofia M.C. Robb,[22] Peter N. Robinson,[17] James Seager [ID],[11] Erik Segerdell,[14] Damian Smedley [ID],[8] Elliot Sollis [ID],[3] Sabrina Toro,[15] Nicole Vasilevsky,[23] Valerie Wood [ID],[18] Melissa A. Haendel,[15] Christopher J. Mungall,[10] James A. McLaughlin [ID],[3] David Osumi-Sutherland[24]

[1]Semanticly, Ermou 56, Athens, 10563, Attiki, Greece

[2]The Jackson Laboratory, Bar Harbor, ME 04609, USA

[3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridgeshire, CB10 1SD, UK

[4]King Abdullah University of Science and Technology, Computer, Electrical & Mathematical Sciences and Engineering Division, Computational Bioscience Research Center, Thuwal, 23955-6900, Saudi Arabia

[5]Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27517, USA

[6]Natural History Museum, University of Oslo, Oslo 0562, Norway

[7]The Institute of Neuroscience, University of Oregon, 5291 University of Oregon, Eugene, OR 97403-5291, USA

[8]William Harvey Research Institute, Queen Mary University of London, London, E14 NS, UK

[9]Department of Biomedical Informatics, Columbia University Irving Medical Center, Columbia University, New York, NY 10032, USA

[10]Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[11]Department of Biointeractions and Crop Protection, Rothamsted Research, West Common, Harpenden, AL52 JQ, UK

[12]Department of Genetics, Stanford University, Palo Alto, CA 94304, USA

[13]Center for Genetic Medicine, Northwestern University, Chicago, IL 60611, USA

[14]Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

[15]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

[16]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[17]Universitätsmedizin Berlin, Berlin Institute of Health at Charité, Anna-Louisa-Karsch-Straße 2, Berlin 10178, Germany

[18]Department of Biochemistry, University of Cambridge, Cambridge, CB21 TN, UK

[19]Ada Health GmbH, Neue Grünstraße 17, Berlin 10179, Germany

[20]UNEP-WCMC, Cambridge CB3 0DL, UK

[21]Department of Biomedical Informatics, University of Colorado, Anschutz Medical Campus, University of Colorado, Aurora, CO 80045, USA

[22]Stowers Institute for Medical Research, Kansas City, MO 64110, USA

[23]Critical Path Institute, Tucson, AZ 85718, USA

[24]Wellcome Sanger Institute, Hinxton, Saffron Walden CB10 1RQ, UK

*Corresponding author: Email: nicolas.matentzoglu@gmail.com
†Joint first authors.

Phenotypic data are critical for understanding biological mechanisms and consequences of genomic variation, and are pivotal for clinical use cases such as disease diagnostics and treatment development. For over a century, vast quantities of phenotype data have been collected in many different contexts covering a variety of organisms. The emerging field of phenomics focuses on integrating and interpreting these data to inform biological hypotheses. A major impediment in phenomics is the wide range of distinct and disconnected approaches to recording the observable characteristics of an organism. Phenotype data are collected and curated using free text, single terms or combinations of terms, using multiple vocabularies, terminologies, or ontologies. Integrating these heterogeneous and often siloed data enables the application of biological knowledge both within and across species. Existing integration efforts are typically limited to mappings between pairs of terminologies; a generic knowledge representation that captures the full range of cross-species phenomics data is much needed. We have developed the Unified Phenotype Ontology (uPheno) framework, a community effort to provide an integration layer over domain-specific phenotype ontologies, as a single, unified, logical representation. uPheno comprises (1) a system for consistent computational definition of phenotype terms using ontology design patterns, maintained as a community library; (2) a hierarchical vocabulary of species-neutral phenotype terms under which their species-specific counterparts are grouped; and (3) mapping tables between species-specific ontologies. This harmonized representation supports use cases such as cross-species integration of genotype-phenotype associations from different organisms and cross-species informed variant prioritization.

Keywords: phenotype; ontology; integration; semantics

Introduction

Phenotypes are observable or measurable characteristics of an organism resulting from the interaction of its genotype with the environment. Collecting and analyzing information about phenotypes, known as *phenotyping*, is fundamental to biological science and has many applications: clinicians record a patient's phenotypic profile to facilitate a more accurate diagnosis; researchers record phenotypic profiles of model organisms to assess interventions (genetic or drug or otherwise); database curators integrate phenotype data with other data types by extracting phenotypes from data sources that are typically unstructured. As the body of phenotype data has expanded, researchers have looked for ways to use this collective knowledge, but the disparate methods used to record these data have posed an impediment.

Variation in alleles of orthologous genes can result in similar phenotypes across species and taxa. For example, *PAX6* gene mutations can lead to human eye phenotypes similar to the mouse phenotypes caused by *Pax6*-ortholog alleles (Lima Cunha *et al.* 2019). Based on similar *FOXP2* phenotypes across humans, primates, mice, and even birds, important inferences can be made about neural mechanisms that contribute to the evolution of human spoken language (Fisher and Scharff 2009). Evolutionarily conserved functions of myostatin gene orthologs manifest in similar muscle growth phenotypes across several vertebrate species (Rodgers and Garikipati 2008). All these examples suggest that phenotypic similarity frequently correlates with the conserved function of gene products and regulatory networks. Identifying similar phenotypes across species can provide not only supporting evidence for conserved gene function but also the possibility of modeling phenotypes in experimentally accessible model organisms to facilitate useful discoveries in agricultural or medical research.

Phenotype ontologies have been developed to reduce ambiguity and relate similar phenotypes, making it possible for computational methods to group phenotypes easily. However, these ontologies have been developed to meet specific use cases and they are widely used in those communities. For example, the Human Phenotype Ontology (HPO) (Gargano *et al.* 2024) has been designed to provide a standardized vocabulary of phenotypic abnormalities and clinical features encountered in human disease and is a recognized standard for the computational encoding of human phenotyping data. HPO enables computational inference, supports genomic and phenotypic analyses, and has widespread applications in clinical diagnostics and translational research. Similarly, the Mammalian Phenotype Ontology (MP) (Smith and Eppig 2009) has been developed to meet the needs of mammalian model organism data. The specific needs of each community influence the design of each ontology. As a result, despite significant overlap between species-specific ontologies, there are notable differences in their axiomatization, classification, and coverage.

As the evolutionary distance between the species increases, the differences in their phenotype ontologies also expand. Ontologies for describing phenotypes of the fruit fly *Drosophila melanogaster* (Drosophila Phenotype Ontology, DPO) (Osumi-Sutherland *et al.* 2013), nematode *Caenorhabditis elegans* (*C. elegans* Phenotype Ontology, WBPhenotype) (Schindelman *et al.* 2011) and the fission yeast *S. pombe* (Fission Yeast Phenotype Ontology, FYPO) (Harris *et al.* 2013), developed to address specific needs of the respective communities, differ extensively in terms of both term organization (the taxonomy, i.e. hierarchy of terms) and the scope (for example, anatomical, cellular, or molecular level) and granularity of phenotypes covered. (See Table 1 for a list of eukaryotic single and multicellular species-specific phenotype ontologies developed to describe scientific data in their respective communities).

In addition, phenotype standardization and integration across species is complicated by the fact that communities use different approaches to annotate phenotypes: while some use pre-composed phenotype ontology terms (for example, HP:0007843 "Attenuation of retinal blood vessels"), others, such as Saccharomyces Genome Database (SGD) and Zebrafish Information Network (ZFIN), use a post-composed approach: the different constituents of the phenotype are captured individually during the curation process using several terms from multiple domain-specific ontologies (for example: GO:0061304 "retinal blood vessel morphogenesis"—PATO:0002302 "decreased process quality") (Mungall *et al.* 2010).

While each approach to describing phenotypes provides standardization for the use cases of a specific community, comparing phenotype data from more than 1 species at scale is difficult and/or very time consuming, as it cannot be done computationally. In contrast, the use of species-neutral ontologies, such as the Gene Ontology (GO) (Ashburner *et al.* 2000) and Uberon (Mungall *et al.* 2012) allows for easier interoperability across a range of taxons. GO is regularly used in many types of large-scale molecular biology experiments, including in genomics, transcriptomics, proteomics, or metabolomics. Annotations made in 1 species may be

Table 1. Domain-specific phenotype ontologies currently integrated into uPheno.

Ontology	Taxon	Term count (release version)	Reference
MP	Mammalia	14,206 (v2024-08-08)	Smith and Eppig (2009)
HPO	*Homo sapiens*	18,987 (v2024-08-13)	Gargano *et al.* (2024)
Zebrafish Phenotype Ontology (ZP)	*Danio rerio*	47,443 (v2024-04-18)	https://github.com/obophenotype/zebrafish-phenotype-ontology
WBPhenotype	Nematoda	2,649 (v2024-06-05)	Schindelman *et al.* (2011)
DPO	Drosophilidae	253 (v2024-04-25)	Osumi-Sutherland *et al.* (2013)
Dictyostelium Phenotype Ontology (DDPHENO)	*Dictyostelium discoideum*	1,017 (v2023-08-26)	Fey *et al.* (2019)
Planarian Phenotype Ontology (PLANP)	Planaria	647 (v2020-03-28)	Nowotarski *et al.* (2021)
XPO	Xenopus	20,340 (v2024-04-18)	Fisher *et al.* (2022)
FYPO	*Schizosaccha romyces pombe*	8,056 (v2024-08-01)	Harris *et al.* (2013)
Pathogen-host interaction phenotype ontology (PHIPO)	General	1,104 (v2024-04-04)	https://github.com/PHI-base/phipo
Molecular glyco-phenotype ontology (MGPO)	General	120 (v2024-04-18)	Gourdine *et al.* (2019)
Ascomycete Phenotype Ontology (APO)	Ascomycota	342 (v2024-04-26)	Engel *et al.* (2010)

automatically applied to other species based on orthology, and cross-species data is easily visualized on many platforms. Similarly, Uberon, the cross-species anatomy ontology, can be used to visualize expression data across species with minimal effort (Aleksander *et al.* 2023; Bult and Sternberg 2023). Phenomics is a relatively young discipline and it has yet to establish a similar level of standardization to GO (Brown *et al.* 2018; Rahman and Rahman 2019). Phenomics vocabularies are developed largely as independent, siloed projects by different communities for different purposes, are often species-specific, and even within the same organism can have multiple, incompatible representations. This ultimately hampers the ability to compare phenotype data across species.

A species-neutral phenotype ontology could integrate phenotype data from any species, allowing for more straightforward incorporation of emerging model organisms and other non-model species, and expanding the scope of comparative biological research. Tools that combine human-specific data with data from only 1 or 2 model organisms have already proven significant performance gains in variant analysis (Smedley *et al.* 2015), disease diagnosis (Sun and Hu 2016), and potential disease model identification (Dickinson *et al.* 2016). Such integrated data could help clinicians to select models that best address their research questions, identify cases where phenotypes are or are not associated with variants in orthologous genes, and uncover factors that influence disease penetrance and severity (Cirincione *et al.* 2018). Phenotype integration also provides a robust approach to align molecular-level phenotypes across large evolutionary distances. For example, the reuse of phenotypic data and variant associations from yeast models such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* can enable predictions related to the molecular basis for diseases, particularly when conserved residues are present in human orthologs.

Efforts such as the Monarch Initiative, the Alliance of Genome Resources, Planteome, and PhenomeNET have integrated a selection of phenotype data employing a variety of methodologies such as the Entity–Quality (EQ) methodology (Bult and Sternberg 2023; Putman *et al.* 2024; Rodríguez-García *et al.* 2017; Cooper *et al.* 2024) and lexical and logical matching. The Unified Phenotype Ontology (uPheno) framework described here builds upon and improves these approaches to establish a unified structure for capturing phenotypic information across species, maintained as a community initiative and applied to a variety of cross-species use cases including clinical diagnostics, data discoverability, and data standardization.

Results

uPheno framework

We have developed uPheno, a framework for cross-species integrative phenomics. uPheno has 3 main components: the uPheno ontology, a library of design patterns (templates) for computationally tractable phenotype definitions, and a number of standardized mappings to connect disparate phenotype ontologies. The uPheno ontology currently integrates 12 species-specific phenotype ontologies (Table 1, Supplementary Fig. 1), which are used by a wide range of databases from the domain of model organisms, including all databases participating in the Alliance of Genome Resources (Bult and Sternberg 2023), and leverages previous efforts to integrate species-specific anatomy ontologies, most notably in the Uberon ontology (Haendel *et al.* 2014) and Cell Ontology (CL) (Diehl *et al.* 2016).

Every phenotype term in the uPheno ontology represents a deviation from a reference phenotype (for example, wild-type) defined

using a specific design pattern from our library (see *Data availability*). This enables phenotype classes to be defined in a consistent logical framework rather than defining each phenotype class manually. For example, the phenotype term UPHENO:0001471 "increased size of the heart" can automatically be generated, along with labels and logical axioms, and accurately classified by instantiating an *increasedSizeOfAnatomicalEntity* pattern with a UBERON:0000948 "heart" class from the anatomy ontology Uberon. In addition to the uPheno ontology, which includes logical connections to all species-specific ontologies, standardized mapping tables are provided with direct links between species-specific and species-neutral ontologies.

Library of computational phenotype patterns

The majority of ontologies in biomedical sciences, especially those covering model organisms, are curated manually using tools such as Protege (Musen 2015). The use of design patterns to augment ontology development processes in the Open Biological and Biomedical Ontologies (OBO) (Jackson *et al.* 2021) community became popular with the emergence of easy-to-use templating systems such as DOSDP (Osumi-Sutherland *et al.* 2017). Rather than manually specifying a term such as "abnormally increased glucose levels in the blood" (which includes writing a human-readable definition, a label, and logical axioms), a DOSDP pattern defines a template for terms of the type "abnormally increased X levels in the Y", including the exact structure of the label, definition, and all its surrounding axioms. This ensures that all terms are consistently labeled (which is particularly difficult in large-scale ontologies such as the phenotype ontologies) and consistently axiomatized. With the addition of reasoning to ontology build pipelines, this consistent axiomatization drives consistent classification (i.e. organization in a hierarchical structure). Some ontologies, such as ZP or XPO (Fisher *et al.* 2022), are entirely bootstrapped from phenotype patterns (see *Discussion*), which reduces the overhead of maintaining them.

We have developed 262 phenotype term templates that cover cases such as "abnormally increased X levels in the Y" (where X is a chemical entity and Y is an anatomical location), "abnormal X morphology" (where X is an anatomical entity), or "abnormal rate of X" (where X is a biological process). Details on the engineering methodology can be found in the *Methods* section. All patterns are available as part of a library of phenotype patterns on GitHub (see *Data availability*).

Phenotypes that affect anatomical entities (UBERON:0001062) or biological processes (GO:0008150) feature prominently in the shared uPheno pattern library constituting approximately 75% of the pattern templates (Fig. 1). Patterns involving anatomical entities make up over 65% of patterns and cover both the morphology and physiology of these entities. Examples involving abnormal anatomical entities include the pattern *abnormalLengthOfAnatomicalEntity* which can be applied to phenotypes characterized by the abnormal length of any anatomical entity, such as HP:0200011 "Abnormal length of corpus callosum", MP:0011999 "abnormal tail length", and ZP:0022039 "head length, abnormal".

Besides phenotypes described by anatomical entity abnormalities, researchers often report the alterations in biological processes associated with specific genetic mutations. The second most frequent phenotype pattern group in the uPheno library relates to biological processes (10.7%). For example, the pattern *abnormallyDecreasedRateOfContinuousBiologicalProcess* can be applied to such diverse process phenotypes as MP:0020234 "decreased basal metabolism", ZP:0101378 "glycolytic process decreased rate, abnormal", FBcv:0000791 "decreased speed of aging", ZP:0001531 "blood circulation decreased rate, abnormal", ZP:0102933
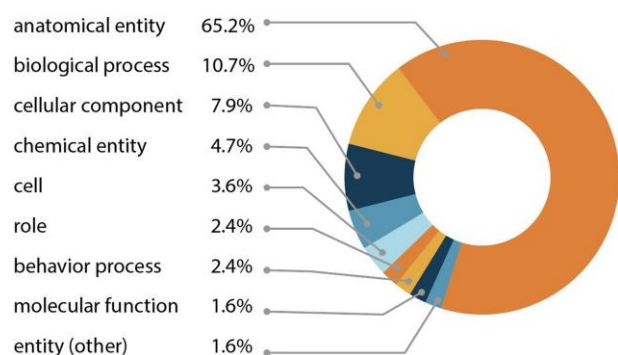
Fig. 1. Distribution of entity types in the uPheno pattern library. All phenotype definitions reference at least one affected entity. The percentage of patterns using an entity type relative to all pattern templates is indicated. The main entity categories in uPheno phenotype pattern templates include: anatomical entity (UBERON:0001062), biological process (GO:0008150), cellular component (GO:0005575), chemical entity (CHEBI:24431), cell (CL:0000000), role (CHEBI:50906), behavior process (NBO:0000313), molecular function (GO:0003674), other entities (BFO:0000001).

"digestion decreased rate, abnormal", and FYPO:0000419 "decreased rate of cytokinesis".

Over 11% of the pattern templates involve cellular (CL:0000000) or cellular component (GO:0005575) phenotypes. Cell component phenotypes can be observed in both single- and multicellular organisms, thus making the relevant uPheno pattern templates applicable to diverse taxa. For example, the *abnormally DecreasedNumberOfCellularComponent* pattern template can be used in cases where a decrease in the number of mitochondria is observed, such as HP:0040013 "Decreased mitochondrial number", MP:0011629 "decreased mitochondrial number", DDPHENO:0000271 "decreased number of mitochondria", and FYPO:0003820 "mitochondria present in decreased numbers during vegetative growth".

Other templates allow the standardization of phenotypic description and annotation related to chemical entities (CHEBI:24431), chemical roles (CHEBI:50906), behavioral processes (NBO:0000313), molecular function (GO:0003674), and developmental processes (GO:0032502).

uPheno ontology

The uPheno ontology is a computational logic-based ontology built using the W3C Web Ontology Language (OWL). uPheno combines existing phenotype ontologies into a single ontology and introduces common grouping classes such as UPHENO:0082544 "mitochondrion phenotype" (Fig. 2a). For example, HP:0001640 "Cardiomegaly", MP:0000274 "enlarged heart" and ZP:0000532 "heart increased size, abnormal" all classify under a common species-neutral grouping UPHENO:0001471 "increased size of the heart", which is in turn classified under UPHENO:0075162 "size of heart phenotype" (Fig. 2b). The grouping classes are primarily built using the uPheno pattern library and rely on external species-neutral ontologies for the component parts. For example, anatomical phenotype terms are created using anatomy terms from Uberon (Mungall *et al.* 2012; Haendel *et al.* 2014), cell type phenotype terms from CL (Diehl *et al.* 2016), and physiological and subcellular phenotypes from GO (Ashburner *et al.* 2000).

The overall structure of the uPheno ontology relies heavily on the structure of the ontologies used to build the classes. By using a wellestablished "entity–quality" (EQ) modeling framework (see *Methods*), we can define a phenotype in terms of its constituent

parts (for example, an anatomical and a chemical entity) and use the hierarchical structure of the respective source ontologies for these parts to classify our phenotype terms using an automated logic-based reasoner such as Elk (Kazakov *et al.* 2014). For example, the phenotype UPHENO:0047922 "increased thickness of the aortic valve leaflet" is classified as a "heart morphology phenotype" (UPHENO:0076810) because "thickness" (PATO:0000915) is considered a subclass of "morphology" (PATO:0000051) in the PATO ontology and the "aortic valve" (UBERON:0002137) is considered a part of the "heart" (UBERON:0000948). For details about this approach refer to the *Methods* section. There are currently 35782 terms in uPheno, of which only 7 are manually classified; all 7 are grouping classes such as UPHENO:3000006 "taste/olfaction phenotype" which are difficult to define using a simple EQ logical definition, see Methods. The remaining terms are defined using the automated reasoning method described above. The advantage of logic-based reasoning compared with machine-learning approaches is that the reference ontologies such as PATO and UBERON have been curated by human experts over many years, which limits the potential for errors in classification.

The uPheno classes enable expressive querying and effective classification of phenotypes across species. For example, a user might want to find all genes where perturbations alter heart morphology regardless of species. To achieve this, they can simply retrieve all subclasses of "heart morphology phenotype" (UPHENO:0076810), for example, using the Ontology Lookup Service (OLS) (McLaughlin *et al.* 2025; https://www.ebi.ac.uk/ols4/ontologies/upheno) (Fig. 2c). This straightforward retrieval of similar phenotypes across taxa can be used for a variety of applications, such as finding relevant literature across species, identifying candidate genes for phenotypes with an unknown genetic basis, or comparing the phenotypic spectrum produced by mutations in orthologous genes across species. The use of the EQ logical framework moreover enables querying for phenotypes using logical expressions. For example, a bioinformatician interested in heart morphology phenotypes across species could use an OWL class expression ('has part' some (morphology and ('characteristic of part of' some ('heart')) and (qualifier some abnormal)).

uPheno currently integrates 12 species-specific phenotype ontologies to varying degrees; see Table 1. The deepest level of integration is for ontologies that cover vertebrates, such as HPO, MP, ZP, and XPO. The integration of other ontologies is more variable; for example, WBPhenotype, DDPHENO and DPO are well integrated, while FYPO and APO are at earlier stages of integration (Fig. 3).

For curation scenarios where no species-specific vocabulary exists, uPheno provides a standardized species-neutral vocabulary that can be used to capture phenotype data (see the Online Mendelian Inheritance in Animals (OMIA) example in *Discussion*).

Cross-species mappings

Cross-species mappings can be used to make datasets interoperable across species, for example, by linking HP phenotypes from a human study to similar MP phenotypes in a mouse study. To facilitate these types of integrations, we publish a number of cross-species mappings derived from the cross-species ontologies (e.g. Uberon, GO) used in the logical definitions of the terms. For example, MP:0003855 "abnormal forelimb zeugopod morphology" maps to HP:0002973 "Abnormal forearm morphology" as they both are defined using the same anatomical term UBERON:0002386 "forelimb zeugopod".

These mappings are published in a simple spreadsheet that complies with the exchange format Simple Standard for Sharing Ontological Mappings (SSSOM) (Matentzoglu and Balhoff *et al.*
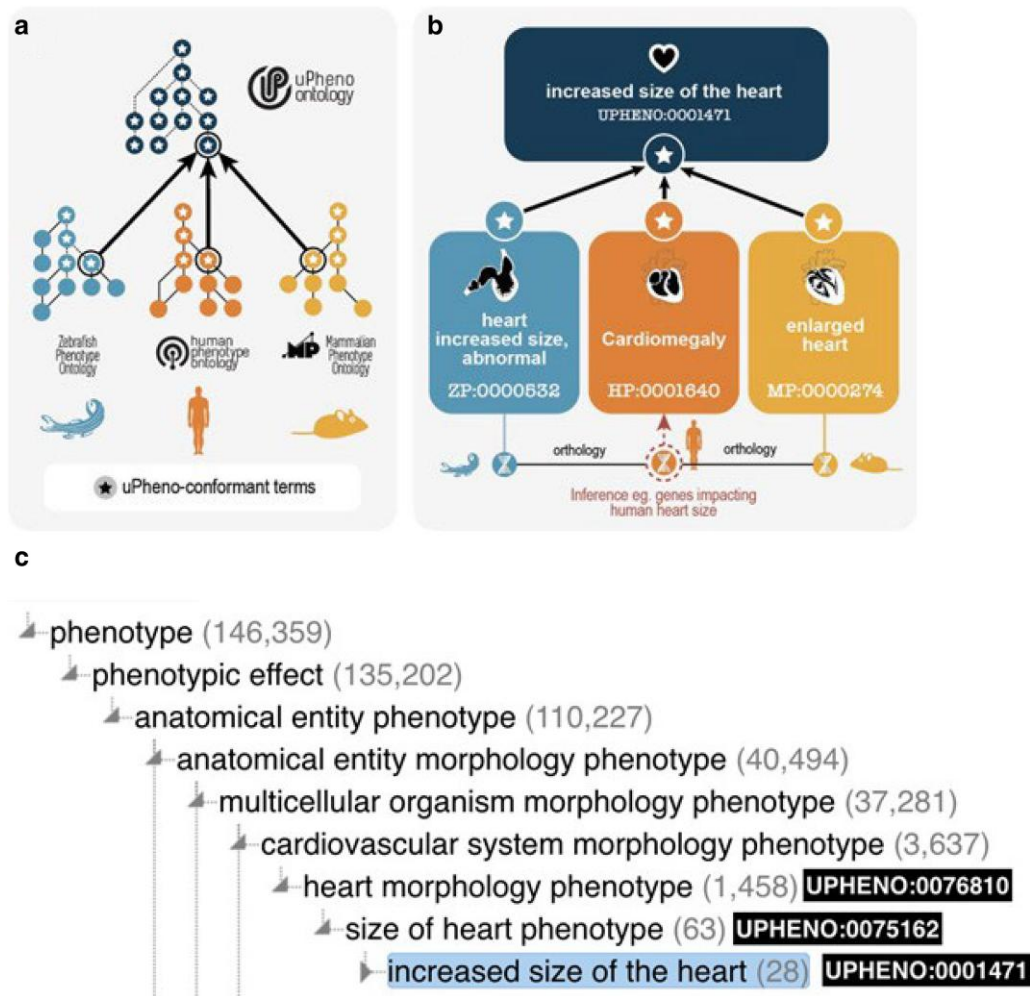
Fig. 2. Structure of the uPheno ontology. uPheno is a framework for consistent and logical definition of phenotype categories using ontology design patterns that provides a hierarchical vocabulary of species-neutral phenotype terms under which their species-specific counterparts are grouped. The ontology design templates are based on shared features of existing phenotypic descriptions from various model organisms and represent community consensus. The phenotype pattern template-adherent terms are adopted by species-specific ontologies, thereby contributing to the community-built uPheno framework. uPheno accelerates cross-species inference and computationally amenable comparative phenotype analysis. For example, the interoperable representation of heart phenotypes characterized by increased size, compared with wild-type in distinct species, such as zebrafish and humans, allows the cross-species identification of genes whose alleles can cause similar phenotypes. uPheno contextual hierarchy for increased size of the heart as displayed in the OLS.

2022), connecting phenotype terms from one species-specific phenotype ontology such as ZP to another, such as XPO. uPheno semantic similarity tables provide associations between species-specific phenotype terms and scores that reflect their semantic similarity. All cross-species mappings, semantic similarity tables and manually curated mappings can be obtained from the URLs provided in the *Data availability* section.

Methods

uPheno integrates existing species-specific representations of phenotype data developed by a broad community of model organism, clinical, and research database curators, using a variety of methodologies. Many of these representations already exist as pre-composed (also known as pre-coordinated) ontologies, where specific terms such as "decreased circulating lysine level" are created and assigned unique, permanent identifiers ("MP:0030719"). Other representations instead rely on curating the different aspects of phenotype separately (post-composition, also known as

post-coordination). For example, ZFIN (Bradford *et al.* 2022) follows a sophisticated post-composed curation style, selecting the attribute and the entity terms separately. To facilitate the integration of this phenotypic data, the ZP ontology, supported by the uPheno effort, converts the post-composed curated content in the ZFIN database into a pre-coordinated ontology.

EQ framework

The computational phenotype model underlying the uPheno framework is an extension of the entity–attribute (or EQ) model which is used to describe phenotypes in terms of affected entities and their characteristics (Washington *et al.* 2009). The affected entities included in phenotypic characterizations are called the *bearers* of the observable attributes (also known as observable qualities or characteristics). For example, in the phenotype "enlarged heart" the entity, heart, bears the characteristic or quality of increased size.

The attribute categories (qualities) in uPheno logical axioms that characterize phenotypes are chosen from the Phenotype
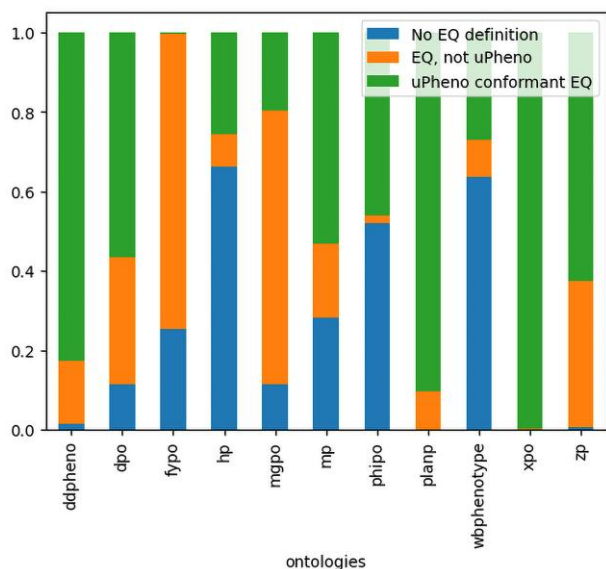
Fig. 3. Current degree of alignment of phenotype ontologies with uPheno. visualization used to quantify the degree of alignment of species-specific phenotype ontologies with uPheno patterns: Proportion of terms that follow a defined uPheno pattern (uPheno-conformant EQ); follow an EQ-style definition (EQ, not uPheno); and terms that do not have a logical definition (no EQ definition). Note that this visualization only quantifies automatically (pattern-based) term alignment and does not include terms aligned using manually defined mappings such as MP to HPO mappings from theMGI database.

and Trait Ontology (PATO) (Gkoutos *et al.* 2005). The entities in uPheno can be broadly categorized as physical objects or processes. The physical objects include anatomical entities and their constituents, such as cells, subcellular structures or components, proteins, and other chemical entities. Examples of process-type entities include GO biological process and GO molecular function classes as well as categories of behavior or roles for chemical compounds. The entity components in uPheno include classes from OBO Foundry (Jackson *et al.* 2021) ontologies, such as Uberon, the CL, GO (cellular component classes), the Chemical Entities of Biological Interest (ChEBI), and the Neuro Behavior Ontology (NBO) (Ashburner *et al.* 2000; Haendel *et al.* 2014; Diehl *et al.* 2016; Gkoutos *et al.* 2012; Hastings *et al.* 2016).

The most basic EQ model involves 2 primitive classes that are part of an asserted class expression, where the "E" entity is the affected entity (e.g. an anatomical entity, such as "limb") or a biological process (e.g. "limb development"). The "Q" component is a quality (attribute) class from PATO. The equivalent class axioms in uPheno follow or extend the basic EQ model to represent phenotypic characteristics. uPheno is intended to represent phenotypic states that deviate from a reference, therefore, they always include a PATO:0000460 "abnormal" component. This is expressed as an equivalent class axiom. For example, UPHENO:0076810 "heart morphology phenotype" is expressed in OWL Manchester syntax (http://www.w3.org/TR/owl2-manchester-syntax/) as follows:

has part

some (

morpholo

gy and

characteristic of part of

some heart and has modifier

some abnormal)

The relationships RO:0000052 "characteristic of" and RO:0002314 "characteristic of part of" from the OBO Relations Ontology (RO) (Mungall *et al.* 2023) are used to connect the entity that is observed to be phenotypically affected and the characteristic it exhibits (e.g. "size" or "amount"). The entity part of the EQ statement can be composite, i.e. comprising more than 1 entity. For example, if an abnormality of a biological process occurs in a particular anatomical location, then the entity will be defined as a biological process (GO:0008150) which "occurs in" (BFO:0000066) an "anatomical location" (UBERON:0001062). More complex patterns defining entities can be found, such as chemical entities that play a certain role (for example a CHEBI:25212 "metabolite" in an UBERON:0001062 "anatomical location").

This formal logic representation of phenotypes in OWL enables the use of logical inference through automated reasoners; see *uPheno ontology* section above for an example on how the use of EQ statements in conjunction with reference ontologies such as ChEBI and Uberon enables entirely automated classification of phenotypes. More information about how OWL ontologies and reasoning can be leveraged in the biological and biomedical sciences can be found in Hoehndorf *et al.* (2015).

The DOSDP framework (and ODK)

Dead Simple OWL Design Patterns (DOSDPs) (Osumi-Sutherland *et al.* 2017) allow efficient and scalable definition of ontology term templates which can support the construction and maintenance of large numbers of ontology classes. The EQ modeling framework is especially well suited for template-based ontology development. DOSDP term templates, which are specified in YAML, support the specification of both logical axioms and annotation axioms (e.g. for synonyms, labels and definitions) with variable slots. Separate tables (stored as tsv or csv files) specify fillers for these variables. The templates and tables can be parsed and converted into OWL axioms by dosdp-tools, which is part of the Ontology Development Kit (ODK) (Matentzoglu and Goutte-Gattat *et al.* 2022). The resulting axioms can be built into a class hierarchy and/or incorporated into an existing OWL ontology by ROBOT (Jackson *et al.* 2019) and other ODK components. The release system of the uPheno ontology is implemented as an ODK workflow, which makes it easily executable in a platform-agnostic manner through Docker.

uPheno templates

uPheno phenotype pattern templates are designed to help align the modeling of similar or related phenotypic categories across multiple taxonomic domains. uPheno utilizes the DOSDP templating system and the EQ framework to define phenotype templates. The uPheno templates are the result of collective curation by a community of ontology editors called the Phenotype Ontology Reconciliation Effort (https://obophenotype.github.io/upheno/reference/reconciliationeffort/). This community effort is pivotal not only to the definition of phenotype templates described in this section, but also in their implementation in the species-specific phenotype ontologies. The curation process operates as follows: when a need for a pattern arises, a member of the community requests a template. Next, another member of the

```
pattern_name: abnormalAnatomicalEntity
pattern_iri: http://purl.obolibrary.org/obo/upheno/patterns/abnormalAnatomicalEntity.yaml
description: "Any unspecified abnormality of an anatomical entity."

contributors:
 -  https://orcid.org/0000-0002-9900-7880
 -  https://orcid.org/0000-0001-9076-6015

classes:
 quality: PATO:0000001
 abnormal: PATO:0000460
 anatomical entity: UBERON:0001062

relations:
 'characteristic_of_part_of':
 RO:0002314 has_modifier:
 RO:0002573
 has_part: BFO:0000051

vars:
 anatomical_entity: "'anatomical entity'"

name:
 text: "abnormal %s"
 vars:
   - anatomical_entity

def:
 text: "Abnormality of %s."
 vars:
   - anatomical_entity

equivalentTo:
 text: "'has_part' some ('quality' and ('characteristic_of_part_of' some %s) and ('has_modifier' some 'abnormal'))"
 vars:
   - anatomical_entity
```

Fig. 4. DOSDP pattern for the representation of abnormal anatomical entity phenotypes. Species-specific phenotype ontologies implement this pattern in phenotype terms such as "Abnormality of the cardiovascular system" (HP:0001626) and "gall bladder quality, abnormal" (ZP:0006529).

Table 2. Relationships used to logically define terms in uPheno.

Relationship	Meaning
characteristic of	Relates the phenotypic quality to the "bearer", i.e. the entity that is affected by the phenotype.
characteristic of part of	Relates the phenotypic quality to the "bearer" or one of its parts.
part of	A general mereological relation that denotes parthood between 2 entities, such as an anatomical entity and one of its parts.
has part	The inverse relationship of part of. In the context of our EQs, this is used to relate the phenotype itself to its primary quality.
towards	In the case of a relational quality, i..e. a quality that involves 2 entities (such as "fused with"), the towards relation is used to describe the second involved entity (the first is related using "characteristic of").
has modifier	Phenotypes can be either normal or abnormal. The relation is used to connect the primary phenotypic quality to a respective modifier.
occurs in	Used to connect biological processes to the anatomical entities in which they are taking place.

community develops a draft template in DOSDP format and makes a pull request on GitHub. The broader community can review the pattern and provide feedback suggesting changes to wording, definitions, naming templates, and other aspects. Once the template is approved, it is presented at a specific monthly call that is organized by the reconciliation effort for the purpose of advancing uPheno patternization across the phenotype ontology editors community. All present members who approve of the pattern now add their signature to the template (in the form of an ORCID, see below), indicating their approval and their intention to implement the pattern in the phenotype ontology they represent (for example, HPO, MP, DDPHENO, etc.).

As an example, a slightly simplified version of the *abnormal AnatomicalEntity* pattern can be seen in Fig. 4. The full pattern can be downloaded at the URL indicated by the "pattern_iri" field.

The 2 most important elements of any pattern are the template for the logical definition (equivalentTo) and the list of contributors. The equivalentTo field specifies the logical axiom pattern to be used to define the phenotype term. The contributor field is used to record the members of the reconciliation effort who have reviewed a particular pattern (see above).

The main relationships used by phenotype patterns can be seen in Table 2. Other relationships are used in specific cases. For example, GOREL:0001006 "acts on population of" is used for the definition of cell proliferation patterns in order to align with the logical definition of cell proliferation process terms in GO.

The uPheno patterns described here are collected and available in the "uPheno pattern library", which makes it easy for phenotype ontology editors to identify a suitable pattern for a given phenotype (see *Data availability*).

Cross-species mappings and semantic similarity

Locating genetic variants with similar phenotypes across species can suggest new disease candidates, or provide new insights into gene function. For example, *PRG4* mutations have been causally implicated in "Camptodactyly of finger" (HP:0100490) phenotypes in human genetic studies, and there are mouse knockout (KO) models of the orthologous *Prg4* gene whose phenotype annotations include "camptodactyly" (MP:0003807), thereby providing independent evidence of causality and increasing confidence that the gene-to-phenotype association is correct (Smith and Eppig 2009; Marcelino *et al.* 1999; Rhee *et al.* 2005). To leverage these similar phenotypes for data analysis (especially when potentially orthologous variants relating to analogous phenotypes are unknown), we publish those mappings in a standardized format called the SSSOM (Matentzoglu and Balhoff *et al.* 2022). SSSOM allows the precise specification of the mapping relation, such as semapv:crossSpeciesExactMatch, to signify a relation between 2 identical phenotypic characteristics of homologous anatomical structures. Metadata can also be attached to distinguish matches which have been determined through manual curation (semapv:ManualMappingCuration), lexical matching (semapv:LexicalMatching) and logical matching using automated reasoning (semapv:LogicalMatching). Mapping tables are particularly useful for basic lookup tasks and providing cross-links between resources. For example, the International Mouse Phenotyping Consortium (IMPC) (Groza *et al.* 2023) and Mouse Genome Informatics (MGI) (Baldarelli *et al.* 2024) websites use mapping tables to allow searching using an HPO phenotype name to discover data connected to an analogous MP phenotype term.

The uPheno framework also enables the computational identification of semantically similar phenotypes. Similar phenotypes are more likely to be related to similar mechanisms, which makes this information very valuable for applications such as variant prioritization. For example, Huntington's disease and Parkinson's disease are both neurodegenerative disorders where involuntary motor symptoms, chorea and tremors, respectively, are prominent phenotypic features. *HTT* gene mutations have been causally implicated in Huntington's disease. We could therefore predict *HTT* might be a candidate gene for Parkinson disease based on phenotypic similarity. This is exploited to prioritize pathogenic variants in tools such as Exomiser (Smedley *et al.* 2015), which is used widely in clinical practice. Estimating "phenotypic similarity" in Exomiser uses 2 main measures: Jaccard, which is simply a measure of how similar 2 phenotypes are with respect to their position in the ontology (sibling terms are more similar than distantly related terms); and the PhenoDigm score (Smedley *et al.* 2013), which is based on Jaccard similarity but normalizes against Information Content, which itself is a measure of "how informative/interesting" a phenotype is (phenotypes that are more specific have stronger known gene associations and are positioned lower in the ontology hierarchy and are considered "more informative").

Discussion

uPheno applications

uPheno is applicable to a variety of phenotypic analysis projects and tools. The IMPC (Groza *et al.* 2023), a global resource for whole gene KO mouse lines, plans to use uPheno cross-species mappings to make mouse phenotypes discoverable using HPO phenotype terms (Groza *et al.* 2023). Similarly, MGI has recently prototyped the use of cross-species mappings for discovering gene-to-phenotype associations (Baldarelli *et al.* 2024) and intends to incorporate uPheno

mappings into this tool. The Monarch Initiative Knowledge Graph (Putman *et al.* 2024) uses uPheno alongside the Ontology of Biological Attributes (OBA) (Stefancsik *et al.* 2023), which enables analyzing biomedical data across species, with a specific focus on phenotypes, diseases, and their genetic underpinnings.

uPheno has also been applied to the phenomics-informed study of disease. In a study to determine whether model organism phenotype data contributes to the computational discovery of human gene-disease associations and to what extent, Alghamdi et. al. used uPheno and Pheno-e (Hoehndorf *et al.* 2011) (an extension of the PhenomeNET ontology) to semantically relate phenotypes resulting from loss-of-function mutations in mouse, zebrafish, fruit fly, and fission yeast model organisms to disease-associated human phenotypes (Alghamdi *et al.* 2022). An informatics pipeline developed by Cary *et al.* presented an Alzheimer's disease risk assessment score across biological domains. The approach utilized phenotypes of model organism orthologs to human genes extracted from the uPheno ontology (Cary *et al.* 2024). InpherNet is a machine-learning approach that can aid monogenic disease diagnosis where patient-based annotation is incomplete or lacking. It leverages the uPheno ontology to obtain organismal and cellular-level gene phenotype data (Yoo *et al.* 2021).

In disease diagnostics, variant prioritization tools such as Exomiser (Smedley *et al.* 2015), EmbedPVP (Althagafi *et al.* 2024), and EvORanker (Canavati *et al.* 2024) leverage uPheno's cross-species phenotypic similarity mappings to improve variant prioritization by comparing human phenotypes to those of model organisms like mice and zebrafish. For example, in a cohort of pediatric patients presenting with a range of clinical phenotypes including global developmental delay, seizures, and generalized hypotonia, (Ji *et al.* 2019) used Exomiser to achieve an overall molecular diagnostic rate of 36%. The Phenotypic Inference Evaluation Framework (PhEval) (Bridges *et al.* 2024) has recently been developed to benchmark diagnostic yield in Exomiser when informed by similar cross-species phenotypes mapped using uPheno.

uPheno can also be used to bootstrap the generation of species-specific phenotype ontologies. Instead of building an ontology manually, uPheno pattern templates and spreadsheets of relevant entities can be used to automate the creation of ontology terms. Xenbase (Fisher *et al.* 2023), the Xenopus model organism knowledgebase, has developed the Xenopus phenotype ontology (XPO) (Fisher *et al.* 2022) using uPheno templates in combination with high-level terms from the Xenopus Anatomy Ontology (Segerdell *et al.* 2008), the PATO and the GO. The PLANP ontology for the Planarian Flatworm has been generated using uPheno patterns for use in phenotype annotation. These terms are being used to help researchers identify genes with comparable phenotypes when perturbed using RNA interference.

Limitations

While uPheno allows the species-neutral description of a phenotype such as "abnormally enlarged heart", it does not address what reference the phenotype is a comparison to (e.g. a control group or wild type), nor does it capture an effect size (e.g. whether the phenotype is slightly outside of the clinically normal range or significantly changed). In practice, all phenotype ontologies are used in contexts where different comparators and effect sizes are assumed. For example, for quantitative traits in the GWAS Catalog, the annotation with a phenotype/trait term indicates that the effect allele is associated with an increase/decrease in the trait compared with the mean of the entire sample; for binary traits, the trait annotation indicates that the effect allele is found at higher/lower frequency in cases compared with controls. Since

the overall goal of uPheno is to make phenotype information comparable, it would be impractical to create different classification axes for every case (e.g. having an "abnormally increased heart size", a "significantly increased heart size", a "abnormally increased heart size compared with wild-type"). Thus, the presence of a phenotype term as part of, for example, a gene-to-phenotype association, cannot automatically be associated with a specific comparator or effect size. Instead, this information needs to be supplied in the metadata of the phenotype annotation, for example, the experimental conditions.

A second important limitation is that, while the uPheno framework has significantly improved the alignment of species-specific phenotype ontologies with each other and with uPheno, it does not automatically lead to their complete alignment. The implementation of uPheno patterns with uPheno-conformant reference ontologies such as Uberon or Uberon-aligned ontologies by the species-specific phenotype ontologies is costly in developer time, so coverage for ontologies such as HPO and MP is unlikely to be complete in the near future. Complex phenotypes are a specific concern, as they are frequently described in different ways across species-specific phenotype ontologies. For example, "anencephaly" (HP:0002323, MP:0001890) share the features of the absence of most or all of the brain (encephalon) tissue, and both are deemed to be defects in the developmental process of neural tube closure in the respective HPO and MP definitions. It is possible to incorporate these phenotypes into uPheno patterns. However, there is a choice of creating logical axioms that focus on morphological features, compared with modeling this phenotype from a developmental process perspective. The ongoing efforts and collaboration of model organism ontology editors can remedy this, and similar problems, by not only defining shared uPheno templates but also reviewing their decisions which specific phenotype should be defined using which pattern.

Nevertheless, significant coverage has already been reached, and the process of alignment is ongoing.

Future work

In the future, we would like to make uPheno more accessible to researchers by integrating the ontology into tools to enable use cases such as finding related phenotypes across species without the need for specialized ontology training. We also plan to improve the upper-level structure of the phenotype hierarchy to improve the findability of phenotype terms for users browsing the ontology.

uPheno has primarily focused on integrating phenomics data from the model organism community. We are expanding our efforts to include non-model animal species and address use cases relevant to the veterinary field. This undertaking has been driven by the team at the Online Mendelian Inheritance in Animals (OMIA, https://omia.org/home/) (Nicholas 2021) with whom we are collaborating. OMIA is a freely available, curated knowledge base that offers up-to-date information on inherited disorders, traits, and associated genes and variants in animals. uPheno was chosen to represent phenotypes and clinical and pathological signs data in OMIA to enhance their data's computational analysis and data interoperability.

Data availability

The uPheno ontology, pattern library and associated files can be found here: https://github.com/obophenotype/upheno/blob/master/docs/reference/data-availability.md, or at the URLs provided in the text. The uPheno ontology can be browsed using OLS (https://www.ebi.ac.uk/ols4/ontologies/upheno).

Supplemental material available at GENETICS online.

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Alghamdi SM, Schofield PN, Hoehndorf R. 2022. Contribution of model organism phenotypes to the computational identification of human disease genes. Dis Model Mech. 15(7):dmm049441. doi: 10.1242/dmm.049441

Althagafi A, Zhapa-Camacho F, Hoehndorf R. 2024. Prioritizing genomic variants through neuro-symbolic, knowledge-enhanced learning. Bioinformatics. 40(5):btae301. doi:10.1093/bioinformatics/btae301

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25(1):25–29. doi:10.1038/75556

Baldarelli RM, Smith CL, Ringwald M, Richardson JE, Bult CJ. 2024. Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. Genetics. 227(1):iyae031. doi:10.1093/genetics/iyae031

Bradford YM, Van Slyke CE, Ruzicka L, Singer A, Eagle A, Fashena D, Howe DG, Frazer K, Martin R, Paddock H, et al. 2022. Zebrafish information network, the knowledgebase for Danio rerio research. Genetics. 220(4):iyac016. doi:10.1093/genetics/iyac016

Bridges Y, de Souza V, Cortes KG, Haendel M, Harris NL, Korn DR, Marinakis NM, Matentzoglu N, McLaughlin JA, Mungall CJ, et al. 2024. Towards a standard benchmark for variant and gene prioritisation algorithms: PhEval. Phenotypic inference Evaluation

framework [Preprint]. bioRxiv 2024.06.13.598672. https://doi.org/10.1101/2024.06.13.598672.

Brown SDM, Holmes CC, Mallon A-M, Meehan TF, Smedley D, Wells S. 2018. High-throughput mouse phenomics for characterizing mammalian gene function. Nat Rev Genet. 19(6):357–370. doi:10.1038/s41576-018-0005-2

Bult CJ, Sternberg PW. 2023. The alliance of genome resources: transforming comparative genomics. Mamm Genome. 34(4):531–544. doi:10.1007/s00335-023-10015-2

Canavati C, Sherill-Rofe D, Kamal L, Bloch I, Zahdeh F, Sharon E, Terespolsky B, Allan IA, Rabie G, Kawas M, *et al.* 2024. Using multi-scale genomics to associate poorly annotated genes with rare diseases. Genome Med. 16(1):4. doi:10.1186/s13073-023-01276-2

Cary GA, Wiley JC, Gockley J, Keegan S, Amirtha Ganesh SS, Heath L, Butler RR, Mangravite LM, Logsdon BA, Longo FM, *et al.* 2024. Genetic and multi-omic risk assessment of Alzheimer's disease implicates core associated biological domains. Alzheimers Dement. 10(2):e12461. doi:10.1002/trc2.12461

Cirincione AG, Clark KL, Kann MG. 2018. Pathway networks generated from human disease phenome. BMC Med Genomics. 11(S3):75. doi:10.1186/s12920-018-0386-2

Cooper L, Elser J, Laporte M-A, Arnaud E, Jaiswal P. 2024. Planteome 2024 Update: Reference Ontologies and Knowledgebase for Plant Biology. Nucleic Acids Res. 52(D1):D1548–D1555. doi:10.1093/nar/gkad1028

Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H, *et al.* 2016. High-throughput discovery of novel developmental phenotypes. Nature. 537(7621):508–514. doi:10.1038/nature19356

Diehl AD, Meehan TF, Bradford YM, Brush MH, Dahdul WM, Dougall DS, He Y, Osumi-Sutherland D, Ruttenberg A, Sarntivijai S, *et al.* 2016. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. J Biomed Semantics. 7(1):44. doi:10.1186/s13326-016-0088-7

Engel SR, Balakrishnan R, Binkley G, Christie KR, Costanzo MC, Dwight SS, Fisk DG, Hirschman JE, Hitz BC, Hong EL, *et al.* 2010. Saccharomyces Genome Database provides mutant phenotype data. Nucleic Acids Res. 38(Database):D433–6 . doi:10.1093/nar/gkp917

Fey P, Dodson RJ, Basu S, Hartline EC, Chisholm RL. 2019. dictyBase and the Dicty Stock Center (version 2.0) - a progress report. Int J Dev Biol. 63(8-9-10):563–572. doi:10.1387/ijdb.190226pf

Fisher M, James-Zorn C, Ponferrada V, Bell AJ, Sundararaj N, Segerdell E, Chaturvedi P, Bayyari N, Chu S, Pells T, *et al.* 2023. Xenbase: key features and resources of the Xenopus model organism knowledgebase. Genetics. 224(1):iyad018. doi:10.1093/genetics/iyad018

Fisher ME, Segerdell E, Matentzoglu N, Nenni MJ, Fortriede JD, Chu S, Pells TJ, Osumi-Sutherland D, Chaturvedi P, James-Zorn C, *et al.* 2022. The Xenopus phenotype ontology: bridging model organism phenotype data to human health and development. BMC Bioinformatics. 23(1):99. doi:10.1186/s12859-022-04636-8

Fisher SE, Scharff C. 2009. FOXP2 as a molecular window into speech and language. Trends Genet. 25(4):166–177. doi:10.1016/j.tig.2009.03.002

Gargano MA, Matentzoglu N, Coleman B, Addo-Lartey EB, Anagnostopoulos AV, Anderton J, Avillach P, Bagley AM, Bakštein E, Balhoff JP, *et al.* 2024. The Human Phenotype Ontology in 2024: phenotypes around the world. Nucleic Acids Res. 52(D1):D1333–D1346. doi:10.1093/nar/gkad1005

Gene Ontology Consortium; Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, Feuermann M, Gaudet P, Harris NL, *et al.* 2023. The Gene Ontology knowledgebase in 2023. Genetics. 224(1):iyad031. doi: 10.1093/genetics/iyad031

Gkoutos GV, Green ECJ, Mallon A-M, Hancock JM, Davidson D. 2005. Using ontologies to describe mouse phenotypes. Genome Biol. 6(1):R8. doi:10.1186/gb-2004-6-1-r8

Gkoutos GV, Schofield PN, Hoehndorf R. 2012. The neurobehavior ontology: an ontology for annotation and integration of behavior and behavioral phenotypes. Int Rev Neurobiol. 103:69–87. doi:10.1016/B978-0-12-388408-4.00004-6

Gourdine J-PF, Brush MH, Vasilevsky NA, Shefchek K, Köhler S, Matentzoglu N, Munoz-Torres MC, McMurry JA, Zhang XA, Robinson PN, *et al.* 2019. Representing glycophenotypes: semantic unification of glycobiology resources for disease discovery. Database. 2019:baz114. doi:10.1093/database/baz114

Groza T, Gomez FL, Mashhadi HH, Muñoz-Fuentes V, Gunes O, Wilson R, Cacheiro P, Frost A, Keskivali-Bond P, Vardal B, *et al.* 2023. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. Nucleic Acids Res. 51(D1):D1038–D1045. doi:10.1093/nar/gkac972

Haendel MA, Balhoff JP, Bastian FB, Blackburn DC, Blake JA, Bradford Y, Comte A, Dahdul WM, Dececchi TA, Druzinsky RE, *et al.* 2014. Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. J Biomed Semantics. 5(1):21. doi:10.1186/2041-1480-5-21

Harris MA, Lock A, Bähler J, Oliver SG, Wood V. 2013. FYPO: the fission yeast phenotype ontology. Bioinformatics. 29(13):1671–1678. doi:10.1093/bioinformatics/btt266

Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. 44(D1):D1214–9. doi:10.1093/nar/gkv1031

Hoehndorf R, Schofield PN, Gkoutos GV. 2011. PhenomeNET: a whole-phenome approach to disease gene discovery. Nucleic Acids Res. 39(18):e119. doi:10.1093/nar/gkr538

Hoehndorf R, Schofield PN, Gkoutos GV. 2015. The role of ontologies in biological and biomedical research: a functional perspective. Brief Bioinform. 16(6):1069–1080. doi:10.1093/bib/bbv011

Jackson R, Matentzoglu N, Overton JA, Vita R, Balhoff JP, Buttigieg PL, Carbon S, Courtot M, Diehl AD, Dooley DM, *et al.* 2021. OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. Database. 2021:baab069. doi:10.1093/database/baab069

Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. 2019. ROBOT: A Tool for Automating Ontology Workflows. BMC Bioinformatics. 20(1):407. doi:10.1186/s12859-019-3002-3

Ji J, Shen L, Bootwalla M, Quindipan C, Tatarinova T, Maglinte DT, Buckley J, Raca G, Saitta SC, Biegel JA, *et al.* 2019. A semiautomated whole-exome sequencing workflow leads to increased diagnostic yield and identification of novel candidate variants. Cold Spring Harb Mol Case Stud. 5(2):a003756 . doi:10.1101/mcs.a003756

Kazakov Y, Krötzsch M, Simančík F. 2014. The incredible ELK: From polynomial procedures to efficient reasoning with $\mathcal{EL}$ ontologies. J Automat Reason. 53(1):1–61. doi:10.1007/s10817-013-9296-3

Lima Cunha D, Arno G, Corton M, Moosajee M. 2019. The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye. Genes. 10(12):1050. doi:10.3390/genes10121050

Marcelino J, Carpten JD, Suwairi WM, Gutierrez OM, Schwartz S, Robbins C, Sood R, Makalowska I, Baxevanis A, Johnstone B, *et al.* 1999. CACP, encoding a secreted proteoglycan, is mutated in camptodactyly-arthropathy-coxa vara-pericarditis syndrome. Nat Genet. 23(3):319–322. doi:10.1038/15496

Matentzoglu N, Balhoff JP, Bello SM, Bizon C, Brush M, Callahan TJ, Chute CG, Duncan WD, Evelo CT, Gabriel D, *et al.* 2022a. A

Simple Standard for Sharing Ontological Mappings (SSSOM). Database. 2022:baac035. doi:10.1093/database/baac035

Matentzoglu N, Goutte-Gattat D, Tan SZK, Balhoff JP, Carbon S, Caron AR, Duncan WD, Flack JE, Haendel M, Harris NL, *et al.* 2022b. Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies. Database. 2022:baac087. doi:10.1093/database/baac087.

McLaughlin J, Lagrimas J, Iqbal H, Parkinson H, Harmse H. 2025. OLS4: A new Ontology Lookup Service for a growing interdisciplinary knowledge ecosystem [Preprint], arXiv, arXiv:2501.13034 [cs.IR]. https://doi.org/10.48550/arXiv.2501.13034

Mungall C, Matentzoglu N, Balhoff J, Osumi-Sutherland D, Duncan B, pgaudet M, Tan S, Hoyt CT, Pilgrim C, Overton JA, *et al.* 2023. Oborel/obo-Relations: 2023-08-18 Release. Zenodo. doi:10.5281/zendo.8263469

Mungall CJ, Gkoutos GV, Smith CL, Haendel MA, Lewis SE, Ashburner M. 2010. Integrating phenotype ontologies across multiple species. Genome Biol. 11(1):R2. doi:10.1186/gb-2010-11-1-r2

Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, Haendel MA. 2012. Uberon, an integrative multi-species anatomy ontology. Genome Biol. 13(1):R5. doi:10.1186/gb-2012-13-1-r5

Musen MA; Protégé Team. 2015. The Protégé Project: A Look Back and a Look Forward. AI Matters. 1(4):4–12. doi:10.1145/2757001.2757003

Nicholas FW. 2021. Online Mendelian Inheritance in Animals (OMIA): a record of advances in animal genetics, freely available on the Internet for 25 years. Anim Genet. 52(1):3–9. doi:10.1111/age.13010

Nowotarski SH, Davies EL, Robb SMC, Ross EJ, Matentzoglu N, Doddihal V, Mir M, McClain M, Sánchez Alvarado A. 2021. Planarian Anatomy Ontology: a resource to connect data within and across experimental platforms. Development. 148(15): dev196097. doi:10.1242/dev.196097

Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. 2017. Dead simple OWL design patterns. J Biomed Semantics. 8(1):18. doi:10.1186/s13326-017-0126-0

Osumi-Sutherland D, Marygold SJ, Millburn GH, McQuilton PA, Ponting L, Stefancsik R, Falls K, Brown NH, Gkoutos GV. 2013. The Drosophila phenotype ontology. J Biomed Semantics. 4(1):30. doi:10.1186/2041-1480-4-30

Putman TE, Schaper K, Matentzoglu N, Rubinetti VP, Alquaddoomi FS, Cox C, Caufield JH, Elsarboukh G, Gehrke S, Hegde H, *et al.* 2024. The Monarch Initiative in 2024: an analytic platform integrating phenotypes, genes and diseases across species. Nucleic Acids Res. 52(D1):D938–D949. doi:10.1093/nar/gkad1082

Rahman J, Rahman S. 2019. The utility of phenomics in diagnosis of inherited metabolic disorders. Clin Med. 19(1):30–36. doi:10.7861/clinmedicine.19-1-30

Rhee DK, Marcelino J, Baker M, Gong Y, Smits P, Lefebvre V, Jay GD, Stewart M, Wang H, Warman ML, *et al.* 2005. The secreted glycoprotein lubricin protects cartilage surfaces and inhibits synovial cell overgrowth. J Clin Invest. 115(3):622–631. doi:10.1172/JCI200522263

Rodgers BD, Garikipati DK. 2008. Clinical, agricultural, and evolutionary biology of myostatin: a comparative review. Endocr Rev. 29(5):513–534. doi:10.1210/er.2008-0003

Rodríguez-García MÁ, Gkoutos GV, Schofield PN, Hoehndorf R. 2017. Integrating phenotype ontologies with PhenomeNET. J Biomed Semantics. 8(1):58. doi:10.1186/s13326-017-0167-4

Schindelman G, Fernandes JS, Bastiani CA, Yook K, Sternberg PW. 2011. Worm Phenotype Ontology: integrating phenotype data within and beyond the C. elegans community. BMC Bioinformatics. 12(1):32. doi:10.1186/1471-2105-12-32

Segerdell E, Bowes JB, Pollet N, Vize PD. 2008. An ontology for Xenopus anatomy and development. BMC Dev Biol. 8(1):92. doi:10.1186/1471-213X-8-92

Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, Siragusa E, Zemojtel T, Buske OJ, Washington NL, *et al.* 2015. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat Protoc. 10(12):2004–2015 . doi:10.1038/nprot.2015.124

Smedley D, Oellrich A, Köhler S, Ruef B, Westerfield M, Robinson P, Lewis S, Mungall C. 2013. PhenoDigm: analyzing curated annotations to associate animal models with human diseases. Database. 2013:bat025. doi:10.1093/database/bat025

Smith CL, Eppig JT. 2009. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip Rev Syst Biol Med. 1(3):390–399. doi:10.1002/wsbm.44

Stefancsik R, Balhoff JP, Balk MA, Ball RL, Bello SM, Caron AR, Chesler EJ, de Souza V, Gehrke S, Haendel M, *et al.* 2023. The Ontology of Biological Attributes (OBA)-computational traits for the life sciences. Mamm Genome. 34(3):364–378. doi:10.1007/s00335-023-09992-1

Sun YV, Hu Y-J. 2016. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. Adv Genet. 93:147–190. doi:10.1016/bs.adgen.2015.11.004

Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. 2009. Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol. 7(11): e1000247. doi:10.1371/journal.pbio.1000247

Yoo B, Birgmeier J, Bernstein JA, Bejerano G. 2021. InpherNet accelerates monogenic disease diagnosis using patients' candidate genes' neighbors. Genet Med. 23(10):1984–1992. doi:10.1038/s41436-021-01238-2