© 2019. MANUSCRIPT SUBMITTED TO JOURNAL OF EXPERIMENTAL BIOLOGY (2019) 00, jebxxxxxx. doi:10.1242/jeb.xxxxxx

RESEARCH ARTICLE

How much do model organism phenotypes contribute to the computational identification of human disease genes?

Sarah M. Alghamdi[1] and Paul N. Schofield [2] and Robert Hoehndorf[1]

Summary statement: We investigate the use of model organism phenotypes in the computational identification of disease genes. We identify several data biases and conclude that mouse model phenotypes contribute most to computational disease gene identification whereas other model organisms do not contribute significantly to this task.

ABSTRACT

Computing phenotypic similarity has been shown to be useful in identification of new disease genes and for rare disease diagnostic support. Genotype–phenotype data from orthologous genes in model organisms can compensate for lack of human data to greatly increase genome coverage. Work over the past decade has demonstrated the power of cross-species phenotype comparisons, and several cross-species phenotype ontologies have been developed for this purpose. The relative contribution of different model organisms to computational identification of disease-associated genes is not yet fully explored. We use methods based on phenotype ontologies to semantically relate phenotypes resulting from loss-of-function mutations in different model organisms to disease-associated phenotypes in humans. Semantic machine learning methods are used to measure how much different model organisms contribute to the identification of known human gene–disease associations. We find that mouse genotype-phenotype data is the most important dataset in the identification of human disease genes by semantic similarity and machine learning over phenotype ontologies. Data from other model organisms does not improve identification over that obtained by using the mouse alone, and therefore does not contribute significantly to this task. Our work has implications for the future development of integrated phenotype ontologies, as well as for the use of model organism phenotypes in human genetic variant interpretation.

KEYWORDS: model organism, phenotype, disease gene discovery, ontology, semantic similarity, machine learning

INTRODUCTION

Discovering and building models of human phenotypes in non-human animals has, over the last half century, proved to be of substantial importance in improving our understanding of human disease and its underlying biology (Aitman et al., 2011; Wangler et al., 2017; Brown, 2021; Baldridge et al., 2021), and is providing insights that may be used to develop new therapeutic and diagnostic capabilities. The amount of data available that relates genetics, in particular genetic variation, to phenotypes associated with disease, is increasing rapidly. For example, the Monarch Initiative lists more than 2M phenotypic associations over more than 100 species from dozens of public resources (Shefchek et al., 2020). By comparing the similarities between phenotypic profiles this data can be used to help understand gene function and to identify the genotypic origins of phenotypic variation, which has wide applications in the discovery of the etiology of disease and the identification of candidate disease genes.

The challenge of relating phenotypes accross different species is very significant. The ontologies and controlled vocabularies used to describe phenotypes are species-specific and often structured in markedly different ways (Gkoutos et al., 2017). In order to compare phenotypic profiles between species, several different approaches have been developed to create an overarching phenotype ontology allowing the integration of phenotype-genotype data from multiple species. This can then be used for measuring phenotypic similarity between an instance of one species, for example a human with a genetic disorder, and phenotypes annotated to multiple species and genotypes. This approach mobilises the huge amount of genotype-phenotype data available in public databases such as Mouse Genome Informatics (MGI) (Eppig et al., 2017; Ringwald et al., 2021), Flybase (Larkin et al., 2020) and Online Mendelian Inheritance in Man (OMIM) (Amberger et al., 2018), and maximizes the possibility of finding a phenotype annotation to a potential disease gene where such a relationship has not yet been reported in humans.

The development of a phenotype ontology covering both humans and model organisms has been essential to this task. The main approaches use evolutionary homology (and analogy) between anatomical structures (Mungall et al., 2012) and physiological processes, formalize these in a knowledge base or ontology, and infer relations between phenotypes using automated reasoning (Matentzoglu et al., 2019; Hoehndorf et al., 2011).

Loss-of-function phenotypes are available for several model organisms. These phenotypes have been generated through both hypothesis-driven experiments and large-scale reverse genetics experiments (Brown et al., 2018; Peterson and Murray, 2021). The genotype–phenotype data from model organisms has been used to discover human disease-associated genes using measures of phenotype similarity (Smedley and Robinson, 2015; Meehan et al., 2017; Hoehndorf et al., 2011). For this purpose, cross-species phenotype ontologies have been developed that systematically relate phenotypes of different organisms to each other (Gkoutos et al., 2017). The underlying assumption of phenotype-based methods to discover disease-associated genes is that genes function in

[1]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, 4700 KAUST, 23955 Thuwal, Saudi Arabia
[2]Department of Physiology, Development & Neuroscience, University of Cambridge, Downing Street, CB2 3EG, Cambridge,United Kingdom

Authors for correspondence: sarah.alghamdi.1@kaust.edu.sa (ORCID: 0000-0001-5544-7166); pns12@cam.ac.uk (ORCID: 0000-0002-5111-7263)); robert.hoehndorf@kaust.edu.sa (ORCID: 0000-0001-8149-5890)

evolutionary conserved pathways or modules, and phenotypes associated with a loss or change of function in a gene, are similar to phenotypes observed in a loss or change of function in the human ortholog of that gene (Oti and Brunner, 2007; McGary et al., 2010; Oti et al., 2008; Barabási et al., 2010). These methods are not only used to identify disease-associated genes but also to interpret and prioritize genomic variants associated with disease in tools that combine variant pathogenicity prediction with ranking of candidate genes (Cipriani et al., 2020; Boudellioua et al., 2017).

Phenotype-based methods to identify candidate genes associated with a set of phenotypes in humans are highly successful when the human gene has already been identified as a disease gene and is therefore associated with phenotypes (Köhler et al., 2009), and there are many examples where mouse phenotypes closely resemble human phenotypes and have therefore been used to identify disease-associated genes in humans (Meehan et al., 2017; Brommage et al., 2019; Smedley et al., 2021). Identification of candidate Mendelian disease genes using high-throughput screening suggests that this strategy might be able to identify candidates for inherited diseases of unknown genetic etiology. For example, out of 3,328 genes screened in the mouse, potential models for 360 diseases were reported including novel candidates (Meehan et al., 2017). More recently, IMPC reported knockouts of 1,484 known disease genes, approximately half of which showed phenotypic similarity to human diseases using the Phenodigm platform (Cacheiro et al., 2019). It is estimated that of the 16,847 mouse genes with a human ortholog, 79.9% have a null allele, either derived from hypothesis-driven experiments or large-scale screens such as the IMPC (Peterson and Murray, 2021); there are currently 3,381 genes with mouse–human orthologs for which there are no corresponding mouse loss-of-function phenotypes. MGI reports 1,694 human diseases with one or more mouse models and 7,142 mouse genotypes modeling human diseases (MGI version 6.17; 14 December 2021), but their interpretation is complicated by the inclusion of dominant inheritance and multigenic or humanized models. It has been suggested that the "phenotype gap" might be filled with genotype-phenotype associations from non-mammalian organisms with complementary coverage to the mouse and where loss-of-function mutations in mouse–human orthologs have no phenotype data (Mungall et al., 2016). To date, the contribution of different model organisms to the computational phenotype-driven identification of human disease genes has not been critically evaluated, an assessment that is important for the continued development of strategies and computational approaches to disease gene discovery. It is important to understand and quantify the contribution of more evolutionarily distant model organisms to discovering human disease-associated genes using the methods that have so successfully been applied to the mouse, in particular as, for example, zebrafish phenotypes are used in methods for disease gene discovery and human genetic variant interpretation (Wangler et al., 2017; Smedley et al., 2015; 2016).

We use two different cross-species ontologies and several state of the art methods for phenotype-based identification of disease-associated genes to evaluate the contribution of mouse, zebrafish, fruitfly, and fission yeast loss-of-function phenotypes to discovering human disease genes. We find that only the mouse consistently predicts disease genes whereas the organisms that are more distant do not contribute. As part of our analysis, we find that our evaluation is affected by several biases in how orthologs of disease-associated genes are annotated in model organism databases as well as how phenotype-based methods exploit these annotations; we analyze and correct for some of these biases to support future work in relating phenotype data to human disease.

RESULTS

Contribution of model organisms to disease gene discovery

We collected phenotypes associated with loss-of-function mutations in the mouse, zebrafish, fruitfly, and fission yeast, from model organism databases. The phenotypes are described using different organism-specific phenotype ontologies and we combine the phenotypes using the integrated phenotype ontologies uPheno (Shefchek et al., 2020) and our extension of the PhenomeNET ontology (Hoehndorf et al., 2011) (Pheno-e). Both phenotype ontologies combine the classes that represent phenotypes in different model organisms within a single ontology, thereby allowing us to exploit relations between the phenotypes and compare them. Pheno-e and uPheno also include human phenotypes from the Human Phenotype Ontology (HPO) (Köhler et al., 2021) thereby allowing us to relate mutant model organism phenotypes to human disease-associated phenotypes.

We used the Pheno-e and uPheno ontologies and the phenotypes associated with loss-of-function mutations and human Mendelian diseases to test whether, and how much, different model organisms contribute to the phenotype-based computational discovery of disease-associated genes. For the purpose of evaluating the predictive performance, we used two datasets of gene–disease association: a "human" dataset which includes associations of human genes with Mendelian diseases reported in the Online Mendelian Inheritance in Man (OMIM) (*Online Mendelian Inheritance in Man (OMIM)*, 2020) database, and a "mouse" evaluation set which consists of associations of mouse genes with human disease and represents mouse models of human disease in the MGI database (Ringwald et al., 2021). Then, we measure the semantic similarity between the phenotypes resulting from a gene's loss of function and human diseases (see Supplementary Figure S1). For each disease, we rank all genes by their phenotypic similarity to the disease; we then determine at which rank we identify orthologs of known disease-associated genes.

This approach has repeatedly been successfully applied to discover disease-associated genes from model organisms through ontology–based computation of phenotype similarity (Meehan et al., 2017; Washington et al., 2009; Smedley et al., 2021), and further forms the foundation of several computational methods for finding disease-associated genomic variants (Smedley and Robinson, 2015; Smedley et al., 2016; Boudellioua et al., 2017). Multiple different approaches for determining phenotypic similarity have been developed, ranging from hand-crafted semantic similarity measures (Köhler et al., 2009; Smedley et al., 2013; Pesquita et al., 2009) to machine learning approaches (Smaili et al., 2018*a*; Chen et al., 2020). We used four different approaches to compute phenotype similarity between model organism phenotypes and human disease. First, we use Resnik's semantic similarity measure (Resnik, 1999) which relies on the taxonomic relations in the phenotype ontology to determine similarity between two sets of phenotypes. Resnik's similarity compares two phenotype classes whereas we need to compare two sets of phenotype classes (i.e., all the phenotypes associated with the disease and all the phenotypes observed in the model organism). Consequently, we use the "best match average" strategy (Pesquita et al., 2009) (see Materials & Methods) to combine multiple pairwise similarity measurements into a similarity between two sets of phenotypes.

Table 1. Comparison of the performance of predicting gene-disease associations evaluated on diseases associated with genes which have orthologs with at least one phenotype annotation in mouse, fish, fly, and yeast (255 genes). Results are bold if they are significantly different from random (i.e., the confidence intervals does not overlap with the ROCAUC 0.5 of a random classifier).

Overlapping Dataset	Mouse	Fish	Fly	Yeast
Resnik similarity	**0.663** $\pm$ 0.069	0.558 ± 0.072	0.572 ± 0.085	0.475 ± 0.069
OPA2Vec	**0.599** $\pm$ 0.072	**0.585** $\pm$ 0.072	0.526 ± 0.074	0.544 ± 0.069
OWL2Vec*	**0.624** $\pm$ 0.071	0.479 ± 0.073	0.472 ± 0.076	0.483 ± 0.069
DL2Vec	**0.607** $\pm$ 0.071	0.517 ± 0.073	0.486 ± 0.075	**0.594** $\pm$ 0.068

Resnik's similarity uses only the ontology taxonomy whereas phenotype ontologies contain a large amount of additional information in the form of axioms that provide a computational description of the intended meaning of phenotype classes (Hoehndorf et al., 2015; Gkoutos et al., 2017). Therefore, we use the unsupervised machine learning method OPA2Vec (Smaili et al., 2018*a*) which is a deep learning method that learns a "representation" of sets of phenotypes based on ontology axioms as well as natural language information contained in ontologies such as labels and definitions. As a third and fourth approach, we use the deep learning methods OWL2Vec* (Chen et al., 2021) and DL2Vec (Chen et al., 2020) which first converts ontology axioms into a graph, applies a random walk to explore the neighborhood of nodes in that graph, and then generates a feature vector using Word2Vec. The aim of using these methods based on random walks is to exploit more "distant" relations that arise through connecting multiple ontology classes. Figure 1 illustrates the different approaches.

In our first experiment, we focus only on the groups of orthologous genes that have phenotype annotations in the mouse, zebrafish, fruitfly, and fission yeast; the aim is to compare the contributions of different model organism to discovering gene–disease associations on the same set of associations from the "human" dataset. There are 255 human genes with orthologous genes annotated with phenotypes in all organisms we consider, and of these, 88 have a human ortholog associated with a Mendelian disease; several genes are associated with more than one Mendelian disease, and, in total, the 88 genes are associated with 173 Mendelian diseases.

We compare the phenotypic similarity of these genes to human disease phenotypes and, within each organism, we rank the genes by their similarity to each disease. We then evaluate the ranks at which we discover the "correct" gene (i.e., the gene with the human ortholog that is associated with the disease) and quantify the results using the ROCAUC measure (see Materials & Methods). Table 1 summarizes the resulting performance. The results indicate that mouse mutant phenotypes can be used to reliably detect human disease-associated genes by all methods, whereas the other organisms do not consistently show a positive signal, and the quality of the signal is very dependent on the method used.

However, our observations are based on a relatively small set of 88 disease-associated genes that have orthologs with phenotypes in all organisms we study. Therefore, we analyze all genes with phenotypes in the different model organisms separately, incorporating genes that may lack phenotype annotations in other model organisms. We were able to test on 11,672 human genes which have an ortholog in the mouse with phenotype annotations; 3,418 genes in the fish; 6,462 in the fruitfly; and 1,871 in yeast. As in our first experiment, we determine the phenotypic similarity using different semantic similarity measures and evaluate how well-established associations can be recovered.

Table 2 summarizes the ROCAUC values for each organism using the four approaches (Resnik similarity, OPA2Vec, OWL2Vec*, and DL2Vec), and Supplementary Tables S1 and S2 shows the results for the Alliance dataset (Agapite et al., 2022). Similar to the first experiment, mouse phenotypes show the highest performance across all methods we consider and the mouse is the only organism where all four methods to compute phenotype similarity show a predictive performance that is better than random. Resnik similarity shows better-than-random performance for zebrafish and fruitfly phenotypes, but other methods predict disease-associated genes no better than a random classifier (except DL2Vec in fission yeast using human gene–disease associations); in evaluations based on ontology embedding methods the predicted performance is even significantly "worse than random" (i.e., significantly below the ROCAUC 0.5 of a random classifier); this indicates that increased phenotypic dissimilarity between a gene and disease is associated with a higher chance of the gene and disease being associated, a rather counter-intuitive result that requires further exploration. We tested the hypothesis that these results are due to a study bias which results in an increased (phenotypic) distance due to the ontology structure. We break this hypothesis into two parts; first, we hypothesize that genes that have an ortholog that is associated with a Mendelian disease in humans have more, and more specific, phenotype annotations than genes whose ortholog is not associated with a Mendelian disease (or for which no human ortholog is known); this hypothesis tests for a form of study bias within the phenotype annotations. We find that disease-associated genes have a significantly higher total information content compared to non-disease associated genes (mouse: $p = 1.361 \cdot 10^{-43}$, fish: $p = 6.793 \cdot 10^{-20}$, fly: $p = 1.115 \cdot 10^{-12}$, yeast: $p = 0.003$; one-tailed t-test).

If these phenotypes do not match human disease-associated phenotypes well, the distance between these specific (i.e., "deep" within the ontology hierarchy) phenotypes and general (i.e., "shallow" within the ontology hierarchy) human phenotypes is higher than for less specific phenotypes; for example, the distance between the very general human phenotype class *Phenotypic abnormality* (`HP:0000118`) and the general fly phenotype *Phenotypic abnormality of organism* (`FBbtAB:00000001`) is less than the distance between *Phenotypic abnormality of organism* (`FBbtAB:00000001`) and the more specific class *Phenotypic abnormality of eye dorsal compartment* (`FBbtAB:00111608`). To further test whether this holds true across all genes with disease-associated and non-associated homologs in human, we calculate the absolute difference in information content between the phenotypes of the fly model and the most informative human phenotype superclass; the average difference in information content for genes with disease-associated human orthologs is 44 whereas the average difference in information content is 14 for genes with non-associated orthologs ($p \leq 1 \cdot 10^{-60}$, Student's t-test; see Supplementary Materials Section 2). The only method that is not based

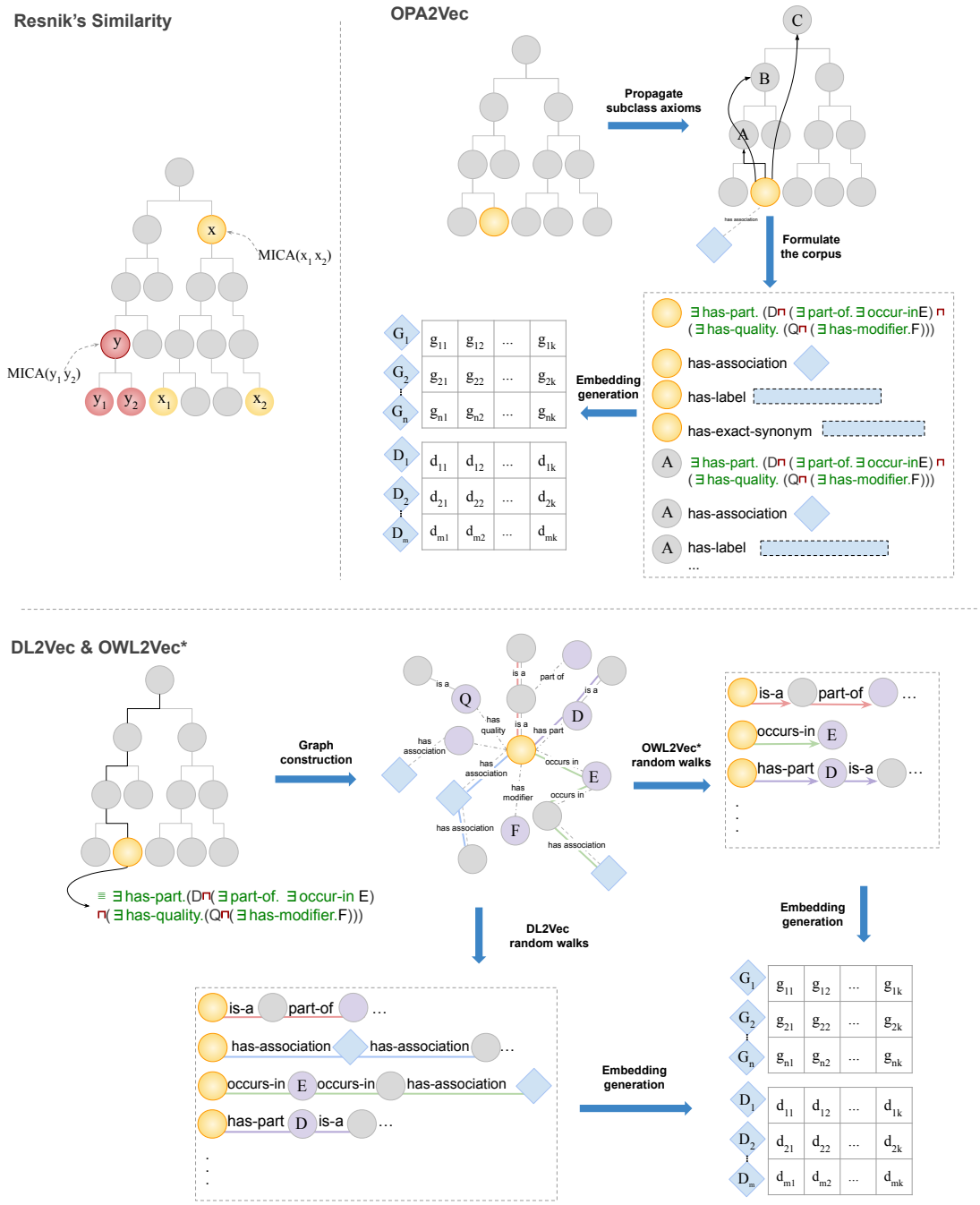

Fig. 1. Illustration of the approaches that we used to calculate phenotypic similarity. Resnik's similarity uses the taxonomy of the ontology. OPA2Vec generates vector representations by using the axioms of the ontologies propagated over the subsumption hierarchy along with the natural language information available in the ontology. DL2Vec and OWL2Vec generate a graph from the ontologies axioms then perform random walks to generate vector representations for genes and diseases, with some differences including that the graphs are directed in OWL2Vec and undirected in DL2Vec.

on distances in our test is Resnik's similarity (which relies on the information content of the most informative shared ancestor), and this is also the only method not showing ROCAUCs below 0.5. Overall, these tests demonstrate that the ROCAUC results significantly lower than 0.5 are due to study bias combined with how the similarity methods utilize the ontology structure to determine similarity (i.e., based on distances traversed between classes).

As the mouse is the only model organism that consistently predicts gene–disease associations, we tested whether combining mouse phenotypes with other organism phenotypes would change the prediction results, i.e., whether combining information from multiple model organisms can improve predictions (i.e., test whether phenotypes of different organisms complement each other). We tested this on varying sets of genes depending on

Table 2. Predicting gene-disease associations

Human evaluations data set				
	Mouse	Fish	Fly	Yeast
Resnik's	**0.786** $\pm$ 0.011	**0.598** $\pm$ 0.018	**0.595** $\pm$ 0.018	0.525 $\pm$ 0.032
OPA2Vec	**0.630** $\pm$ 0.013	0.459 $\pm$ 0.017	0.306 $\pm$ 0.016	0.517 $\pm$ 0.032
OWL2Vec*	**0.711** $\pm$ 0.012	0.449 $\pm$ 0.018	0.418 $\pm$ 0.018	0.500 $\pm$ 0.032
DL2Vec	**0.761** $\pm$ 0.011	0.473 $\pm$ 0.018	0.389 $\pm$ 0.017	**0.557** $\pm$ 0.032
Mouse evaluations data set				
	Mouse	Fish	Fly	Yeast
Resnik's	**0.942** $\pm$ 0.009	**0.592** $\pm$ 0.026	**0.639** $\pm$ 0.039	0.525 $\pm$ 0.061
OPA2Vec	**0.696** $\pm$ 0.017	0.462 $\pm$ 0.025	0.304 $\pm$ 0.035	0.406 $\pm$ 0.060
OWL2Vec*	**0.817** $\pm$ 0.014	0.447 $\pm$ 0.026	0.435 $\pm$ 0.038	0.479 $\pm$ 0.061
DL2Vec	**0.887** $\pm$ 0.012	0.460 $\pm$ 0.026	0.404 $\pm$ 0.037	0.447 $\pm$ 0.061

whether they have phenotypes in two model organisms. Supplementary Table S3 shows the results. We find that combining mouse phenotypes with phenotypes of other model organisms does not significantly change the prediction results.

So far, we performed our analysis only using the Pheno-e ontology. It was unclear whether our results demonstrated an inability of the Pheno-e ontology to compare phenotypes adequately or if they reflect a property of the underlying data and the methods used to analyze it. Consequently, we used the cross-species phenotype ontology uPheno (Shefchek et al., 2020) and repeated the same analysis of predicting gene–disease associations using the four phenotype similarity computation methods; the results and comparison to Pheno-e are shown in Table 3. The results indicate that Pheno-e and uPheno have comparable performance and do not consistently show significant differences in predictive performance across different model organism and analysis methods.

We explored the characteristic scope of input data from different organisms, and how intrinsic bias in coverage of the genome-phenome space cognate with humans might affect the contribution of each organism to computational prediction. For example some species lack organ systems present in humans, and others have quite distant physiology, for example in the immune system. There may also be biases in the selection of experimental systems between different model organisms, dependent partly on previously demonstrated value of those systems and on the historical development of study; furthermore, tractability, proven value for a particular area of investigation, or cost may also explain differences between model organisms. Consequently, we investigated the predictive performance for different types of diseases separately, using the top-level classification of diseases in the DO (Schriml et al., 2021) (see Supplementary Tables S4–S51). Our analysis of the disease classes to which different model organisms contribute most show that, for example, fly and fish contribute to disease of the brain, central nervous system and nervous system diseases. The fruitfly demonstrates substantial predictive performance for mental diseases and behavioral diseases, whereas yeast is predictive mainly for metabolic disorders.

Supervised prediction

One advantage of similarity measures that rely on embeddings is that they can be used as input to "supervised" machine learning approaches and thereby give rise to supervised similarity measures (Smaili et al., 2018*b*). In supervised machine learning, some examples of existing and absent associations between genes and diseases are used to train a model that can determine whether a new gene–disease pairs is associated or not. Using the ontology embeddings as input to supervised machine learning methods has previously resulted in significantly improved prediction of gene–disease associations (Smaili et al., 2018*a*).

We train a machine learning model (an artificial neural network), and use the output of this model to classify pairs of gene- and disease-embedding into two classes, depending on whether the gene is associated with the disease (positives) or not (negatives). We evaluate the performance using a 10-fold cross-validation strategy (see Materials & Methods); Table 4 shows the results. We find that the supervised machine learning approach improves the predictive performance significantly over the unsupervised similarity-based approach, not only when using mouse model phenotypes but also for all other organisms. Furthermore, the supervised model is able to predict gene–disease associations significantly better than a random classifier using all embedding methods and organisms, and further improves significantly over all unsupervised prediction approaches.

However, while the predictive performance is substantially higher than random, it is somewhat surprising that the predictive performance when using phenotypes from distant organisms such as fly or fish matches the performance of using mouse phenotypes, and that even yeast phenotypes apparently are able to identify a large number of gene–disease associations quite accurately when there are so few orthologous genes (estimated to be around 2,000 (O'Brien, 2004)), many without known disease associations in OMIM, the evaluation dataset. Neural networks may be able to exploit non-biological signals in training datasets to achieve relatively high predictive performance without producing biologically meaningful prediction results. For example, genes that are well studied and have a higher number of annotations may be associated with more diseases, or more likely be associated with diseases; ranking genes higher solely based on the number of annotations they received could therefore improve prediction performance even without a specific biological signal. To test this hypothesis, we design a "naïve" classifier that predicts gene–disease associations solely based on the sum of the information content of phenotypes within a gene, i.e., it can be used to test whether genes that are annotated with more and more specific phenotypes are more likely associated with any disease. The naïve classifier ranks all genes based on the sum of the information content of their phenotype annotations, and predicts, for each disease D, the genes in descending order ranked by their information content; this prediction is independent of the disease D, i.e., the same list of genes is predicted in the same order for each disease (see

Table 3. Comparison of the performance of the Pheno-e and the uPheno ontologies to predict gene-disease associations using mouse, fly and yeast on the human evaluation data set. Note that Fish data are not included in this comparison as uPheno uses the zebrafish phenotype ontology and Pheno-e defines fish phenotype classes axiomatically.

	Mouse		Fly		Yeast	
	Pheno-e	uPheno	Pheno-e	uPheno	Pheno-e	uPheno
Resnik's	**0.786** $\pm$ 0.011	**0.746** $\pm$ 0.011	**0.593** $\pm$ 0.018	**0.645** $\pm$ 0.018	0.525 $\pm$ 0.032	0.509 $\pm$ 0.032
OPA2Vec	**0.626** $\pm$ 0.013	**0.590** $\pm$ 0.013	0.375 $\pm$ 0.018	0.326 $\pm$ 0.017	0.522 $\pm$ 0.032	0.519 $\pm$ 0.032
OWL2Vec	**0.711** $\pm$ 0.012	**0.751** $\pm$ 0.011	0.418 $\pm$ 0.018	0.488 $\pm$ 0.018	0.500 $\pm$ 0.032	0.514 $\pm$ 0.032
DL2Vec	**0.755** $\pm$ 0.011	**0.756** $\pm$ 0.011	0.444 $\pm$ 0.018	0.436 $\pm$ 0.018	**0.550** $\pm$ 0.032	**0.538** $\pm$ 0.032

Table 4. Predicting gene-disease associations using supervised methods and our proposed naïve classifier

Human evaluations data set

	Mouse	Fish	Fly	Yeast
MLP - OPA2Vec	0.898 $\pm$ 0.007	0.836 $\pm$ 0.013	0.872 $\pm$ 0.012	0.763 $\pm$ 0.0275
MLP - OWL2Vec	0.875 $\pm$ 0.009	0.826 $\pm$ 0.013	0.874 $\pm$ 0.012	0.775 $\pm$ 0.027
MLP - DL2Vec	0.897 $\pm$ 0.007	0.880 $\pm$ 0.012	0.893 $\pm$ 0.011	0.781 $\pm$ 0.0267
Naïve Classifier	0.722 $\pm$ 0.011	0.689 $\pm$ 0.016	0.753 $\pm$ 0.014	0.510 $\pm$ 0.031

Materials & Methods). The result of the prediction by the machine learning model, together with the naïve classifier results, are shown in Table 4. The results demonstrate that there is substantial bias in the underlying data that can be exploited by the naïve classifier, and is likely exploited by the machine learning models as well.

DISCUSSION

We have evaluated the contribution of different model organism phenotypes to the computational identification of human gene–disease associations through the use of a variety of semantic similarity and machine learning methods. We find that the main contribution towards discovering human disease-associated genes using these methods comes from mouse phenotypes, whereas other model organism data do not contribute significantly to this task. The premise that pooling genotype–phenotype data from multiple organisms to enhance the phenotype-driven prediction of human disease genes, or interpret human genetic variants, is in principle sound, and has driven the development of multiple cross-species phenotype ontologies. The assumption has been that, as long as the knowledge contained in the ontology is "true", then this should help bridge the "phenotype gap", i.e., the human genes that have no phenotype associations in human but do have in model organisms. However, a critical evaluation of the main types of methods in use, machine learning and semantic similarity, indicates that the contribution of the non-mammalian model organism phenotypes to this task is computationally insignificant in comparison to mouse data. We identify two problems with the inconsistency of the results obtained by different methods; the first is bias generated by the use of the structure of the cross-species ontologies available, and the second we have identified as issues such as annotation density; however, there may be further biases that affect the results.

We tested the impact of a number of different parameters on our finding. First, our results hold true across two cross-species phenotype ontologies, Pheno-e and uPheno (Matentzoglu et al., 2019). Both ontologies have similar content and goals but are based on different ontology design patterns (Gkoutos et al., 2017; Alghamdi et al., 2019). We compared the two ontologies in our analyses to test whether the underlying ontology design patterns have a significant impact but we did not consistently identify significant differences between both ontologies, indicating that our results hold true independent of the choice of phenotype ontology. Further, we used different analysis methods, focusing both on traditional semantic similarity measures (Pesquita et al., 2009) that are largely defined based on explicit assumption of how similarity should be computed, as well as methods based on unsupervised and supervised machine learning with ontologies (Kulmanov et al., 2020). An increased number of, and more specific, annotations will bias estimates of semantic similarity. The effect of these biases has been demonstrated when comparing between model organism and human disease phenotypes (Kulmanov and Hoehndorf, 2017), and also when predicting gene–disease associations where this bias can be corrected when detected (Cornish et al., 2018). The machine learning methods we employed are largely based on "paths" in graph-based representations of the ontologies, whereas the semantic similarity we used is based on information content of classes without considering "paths" explicitly; in particular, "distance" is not a relevant consideration in our chosen semantic similarity measure whereas distance is relevant in the machine learning methods we consider. We find that the notion of distance introduces a bias in prediction results, similar to biases found in some semantic similarity measures (Kulmanov and Hoehndorf, 2017; Cornish et al., 2018); using these methods should consequently be considered carefully, in particular as their blackbox nature makes it challenging to identify the reason for a prediction.

We identified and tested the impact of different biases within phenotype-based methods for finding candidate genes. We found a general study bias where disease-associated genes (or genes whose human ortholog is disease-associated) have generally more, and more specific, annotations than non-associated genes, and this affects not only semantic similarity measures but also machine learning methods; even more concerning, supervised machine learning methods can exploit biases in the data to make accurate predictions based on non-biological properties of the data (such as number and type of phenotype annotations). Again, use of black-box models such as neural networks presents the danger of hiding the biases and how they are utilized in decision making.

We demonstrate here that assessment of the contribution of different model organisms to disease gene identification depends critically on the methods used, and we present evidence that supervised machine learning methods systematically overestimate the contribution of some model organisms, mainly by exploiting biases in phenotype data. Similar biases affect the evaluation of gene–disease and variant pathogenicity prediction methods

(Grimm et al., 2015), and are challenging to detect and correct for. In the future, evaluation datasets and methods need to be developed that are less likely to overfit to biases in training and testing data, generalize well across different organisms, and are robust to noise in phenotype annotations.

The use of model organisms for understanding genotype–phenotype relations in humans is well established as a valuable strategy. Few organisms present a complete model of the human (Sundberg et al., 2013), but aspects of, for example, a disease phenotype might be studied more conveniently and with good fidelity in certain species or strains, yielding valuable insights from several model organisms (Hmeljak and Justice, 2019). While the mouse is anatomically and physiologically closest to humans and therefore phenotypes are more likely to be easily related, we know that in particular systems or metabolic pathways valuable information can come from much more distant organisms, for example insights into cell cycle control or ageing from yeast (Pardo and Boriek, 2020). The computational use of model organism phenotypes to identify the underlying genetics of disease is a recent development, based on the premise that, as we do not have functional information for all genes in humans, combining this knowledge from model organisms can massively increase the knowledge that can be brought to bear (Mungall et al., 2010). Based on our data, there are 13,789 human genes that have an ortholog in the model organisms we investigate which have been assigned one or more phenotypes (11,672 human genes have an ortholog in the mouse with phenotype annotations; 3,418 genes in the fish; 6,462 in the fruitfly; and 1,871 in yeast), and therefore over 63% of human genes have orthologs in model organisms with phenotype annotations (Willyard, 2018). Figure 2 shows the pairwise overlap of genes with phenotypes in mouse, fish, fly and yeast.

A question not so far addressed is which of these annotations add to the power of computational approaches to discover disease-associated genes. By using two cross-species phenotype ontologies we are able to show that the data from the mouse explains the majority of the human disease gene associations and that very little, if any, data from other models organisms contributes to this task. A previous focused study comparing mouse and fish phenotypes to predict disease genes in different disease categories for the Phenodigm algorithm (Oellrich et al., 2014) also showed the mouse to be overall more useful but suggested that the zebrafish made contributions in specific disease areas. The authors suggested that this may be due to increased coverage of cardiovascular diseases in the data from mutant fish. Our findings are consistent with this, and also suggest that, as a consequence, the resulting performance is due to biases in the number and specificity of annotations and not due to the intrinsic relatedness between phenotypes.

We further analyzed the broader disease classes to which model organisms contribute (Supplementary Tables S4–S51). Our results are consistent with the common selection of model organisms for different disease groups. For example, *Zebrafish* are widely used in models of cardiovascular disease (Dahme et al., 2009; Prykhozhij and Berman, 2018; Narumanchi et al., 2021), and yeast as a model of metabolic disorders (Cervelli and Galli, 2021). Interestingly, we did not identify model organisms besides the mouse for immune system or eye disease including the retina. Our analyses reflect the intrinsic strengths of each model organism and the bias in the choice of organism to use when investigating particular types of disorders, either as a consequence of the utility of the models developed, or the importance of the diseases modeled.

A deeper consideration of the differences and commonalities between phenotype–genotype relations in model organism species indicates why an analysis at a phenotypic level may be particularly sensitive to evolutionary distance. Intuitively, different organisms have different ontogeny and anatomy; for example, fish have fins which are homologous to mammalian limbs but differ in organization, and flies have wings and legs which have no homologs in mammals. Mutations in genes associated with phenotypes in fish and flies are often not associated with comparable phenotypes in humans simply because the structures are lacking or are profoundly different. For example, the human ortholog of the dishevelled gene (Dsh) in *Drosophila*, which affects segment polarity, causes autosomal dominant Robinow syndrome (OMIM:616331) (Patton and Afzal, 2002); yet, the phenotypes are not readily relatable because most of the structures affected in the human have no homologs in the fly. Nevertheless, decades of experimental investigation show that the underlying molecular processes in which these genes are involved are highly conserved but have evolved to be used in different morphogenetic or physiological processes (Wangler et al., 2017). Recognition of this problem has lead to the concept of orthologous phenotypes or phenologs relating different phenotypes in different organisms resulting from mutation in the orthologous gene, but these are very difficult to identify in a phenotype-led approach (McGary et al., 2010). For similar reasons, it is challenging to predict phenotypic pleiotropy in genetically distant organisms due to the inevitable differences in genome, genetic interactions, anatomy, and physiology (Wagner and Zhang, 2011; Chesmore et al., 2017) and the fact that phenotypes are emergent properties of an organism (Varela et al., 1974). This means that model organisms can be extremely valuable in understanding and investigating the molecular mechanisms underlying normal physiology as well as patho-physiology, but identifying the participants in such processes from whole organism, or often even cellular, phenotypes can be extremely difficult and species-dependent (Kulmanov and Hoehndorf, 2020). Conversely, we have examples where careful phenotypic characterization of mouse models has identified the genetic origin of human diseases and in some cases expanded the phenotypic characterization of quite familiar disorders (Brommage et al., 2019; Thiele et al., 2012), emphasizing the power of phenotypic homology in closely related organisms.

Our study illustrates some of the limitations in the utility of non-mammalian phenotype–genotype data in computational discovery of genes responsible for human disease. It highlights more broadly intrinsic problems in "black box" methods for machine learning, the weaknesses of both phenotype ontologies and methods for estimating semantic similarity, data overfitting, and the intrinsic biases in data collection. We hope that understanding these problems will help in the development of new approaches, possibly new ontological tools, and inform the collection and annotation of new data.

MATERIALS AND METHODS

Ontologies

Several foundational ontologies are used for the axiomatisation of species-specific phenotype ontologies, and we reused them for the construction of the Pheno-e ontology. We used the Gene Ontology (GO) (Ashburner et al., 2000) downloaded from http://purl.obolibrary.org/obo/go.owl ; the Cell Ontology (CL) (Diehl et al., 2016) downloaded from http://purl.obolibrary.org/obo/cl-basic.

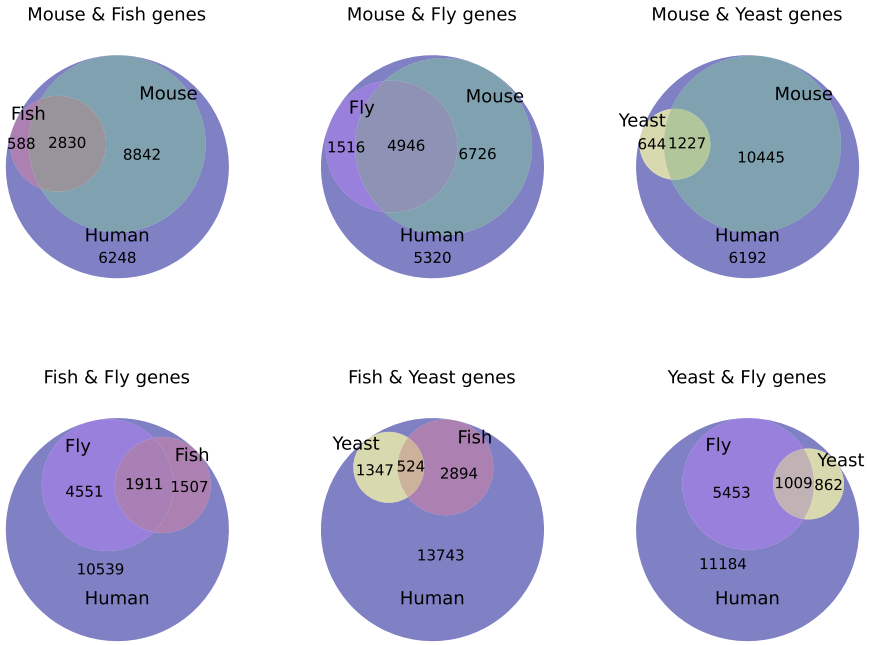

Fig. 2. This figure present human genes with model organism orthologs. The pairwise intersection of model organisms with phenotypes is illustrated in sub-graphs. Each of these sub-graphs represents 18,508 human genes in total.

owl; Phenotype and Trait Ontology (PATO) (Gkoutos et al., 2005) downloaded from http://purl.obolibrary.org/obo/pato.owl; Uber Anatomy Ontology (UBERON) (Mungall et al., 2012) downloaded from http://purl.obolibrary.org/obo/uberon.owl; Zebrafish Anatomy and Development Ontology (ZFA) (Van Slyke et al., 2014) downloaded from http://purl.obolibrary.org/obo/zfa.owl; Neuro Behavior Ontology (NBO) (Gkoutos et al., 2012) downloaded from http://purl.obolibrary.org/obo/nbo.owl; Biological Spatial Ontology (BSPO) (Dahdul et al., 2014) downloaded from http://purl.obolibrary.org/obo/bspo.owl; *Drosophila* Gross Anatomy Ontology (FB-BT) (Costa et al., 2013) downloaded from http://purl.obolibrary.org/obo/fbbt.owl.

The phenotype ontologies used were: Mammalian Phenotype Ontology (MP) (Smith and Eppig, 2012) downloaded from http://purl.obolibrary.org/obo/mp.owl; Human Phenotype Ontology (HP) (Köhler et al., 2018) downloaded from http://purl.obolibrary.org/obo/hp.owl; *Drosophila* Phenotype Ontology (DPO) (Osumi-Sutherland et al., 2013) downloaded form http://purl.obolibrary.org/obo/dpo.owl and Fission Yeast Phenotype Ontology (FYPO) (Harris et al., 2013) downloaded from http://purl.obolibrary.org/obo/fypo.owl. The latest version of the ontologies is used for every update of Pheno-e; the results reported here use ontologies downloaded in February 2021.

Data Sets and Phenotype Annotations

For constructing the model organism phenotype classes we used the following:

- From the ontologies MP (Smith and Eppig, 2012), HP (Köhler et al., 2018), DPO (Osumi-Sutherland et al., 2013) and FYPO (Harris et al., 2013), we reconstructed the phenotype classes for mice, human, fly and yeast respectively.
- From FlyBase (Thurmond et al., 2018) we used `allele_phenotypic_data_fb_2021_01.tsv`, which provides the alleles phenotypes association using controlled vocabulary for *Drosophila melanogaster*. We used this file to create the abnormal anatomy classes (FBabAB).
- From ZFIN, we used `phenoGeneCleanData_fish.txt` which contains zebrafish gene–phenotype associations to create classes representing zebrafish phenotypes.

For generating representations of genes and diseases and for predicting gene–disease associations we used the following files downloaded on 07-Feb-2021:

- Human disease–phenotype annotations were obtained from the HPO database (Köhler et al., 2018) `phenotype_annotation.tab`. comprising manual and semi-automated annotations representing disease identifiers from three databases OMIM (Amberger and Hamosh, 2017), Orphanet (Weinreich et al., 2008) and DECIPHER (Firth et al., 2009).
- Mouse gene–phenotype annotations were obtained from the Mouse Genome Informatics (MGI) database (Ringwald et al., 2021) MGI_GenePheno.rpt which use MP (Smith and Eppig, 2012).

- We generated the fly gene–phenotype annotations using files from FlyBase (Thurmond et al., 2018), `allele_phenotypic_data_fb_2021_01.tsv` which represents the allele–phenotype annotations and `fbal_to_fbgn_fb_2021_01.tsv` which contains allele–gene mappings.
- We obtained yeast gene–phenotype annotations from Pombase (Harris et al., 2013) which represent the phenotypes annotations for fission yeast *Schizosaccharomyces pombe* `phenotype_annotations.pombase.phaf.gz`.

For evaluation purposes, we used two gene-disease association datasets from MGI (Ringwald et al., 2021). The first dataset is a "human" dataset that includes associations of human genes with Mendelian diseases using identifiers from OMIM database (*Online Mendelian Inheritance in Man (OMIM)*, 2020). MGI acquire this dataset for human gene–disease associations from OMIM and additional data from other sources including NCBI's Gene Review (Eppig et al., 2017). This dataset includes 2,848 human gene, 3644 OMIM disease and 11,778 human gene–disease associations. The second dataset is a "mouse" evaluation set that includes associations of mouse genes with human disease and represents mouse models of human disease in the MGI database. This dataset contains 2,459 mouse genes, 2,157 disease and 8,101 mouse gene–disease associations. This dataset is acquired by curating data on mouse models from the scientific literature and high throughput experiments (Eppig et al., 2017). Both datasets are included in the file `MGI_DO.rpt` available from MGI. Gene–phenotype and gene–disease annotations are derived by manual curation (Bello et al., 2016). The version we used was downloaded in February 2021. Additionally, we utilised data on gene–disease associations from the Alliance of Genome Resources (*Alliance of Genome Resources Portal: unified model organism research platform*, 2020), an effort to integrate data resources among the major model organism databases. We mapped Disease Ontology (DO) (Schriml et al., 2022) identifiers to OMIM using the DO and MGI database cross references. As a result, we identified 4,022, 2,366, 2,681, and 406 OMIM diseases that were associated with mouse, fish, fly, and yeast genes, respectively.

To find orthologous genes between different organisms we used several files. Human–mouse orthology was obtained from `HMD_HumanPhenotype.rpt` from MGI. Human–zebrafish and mouse–zebrafish orthologs were obtained from `human_orthos.txt` and `mouse_orthos.txt` from ZFIN. We obtained human–fly orthology from `dmel_human_orthologs_disease_fb_2020_06.tsv` from FlyBase. Human–yeast orthologs were obtained from `pombeorthologs`. We obtained mouse–fly and mouse–yeast orthologs from OMA (Train et al., 2017).

Pheno-e and integration of model organism phenotypes

The PhenomeNET Ontology was developed by utilizing existing phenotype ontology class descriptions and reformulating them according to a set of ontology design patterns so that different phenotype ontologies can be integrated (Hoehndorf et al., 2011). The uPheno ontology similarly establishes bridging axioms to connect phenotypes from different species-specific ontologies (Matentzoglu et al., 2019). The current version of the PhenomeNET ontology does not contain classes for yeast and fly phenotypes while these two species are covered in uPheno. We therefore expanded PhenomeNET to include phenotypes from fly and yeast. We obtained the phenotype class descriptions from the DPO and FYPO ontologies, and reformulated them using the PhenomeNET design patterns. The new classes we created use the pattern

$$?Phenotype \equiv \exists has\_part.(?E \sqcap \exists has\_quality.?Q) \quad (1)$$

In this pattern, $?E$ characterizes the entity underlying the phenotype (either from an anatomy ontology or the GO) and $?Q$ is a quality from the PATO ontology. We use relations from the OBO Relation Ontology (Smith et al., 2005) ; the relations we use in constructing PhenomeNET and Pheno-e include *part-of*, *results-from*, *during*, *has-quality*, *has-central-participant*, *occurs-in*, and *towards*.

FlyBase has two types of abnormal phenotype classes; it associates alleles with classes from the DPO as well as with classes from the fly anatomy ontology, indicating that an anatomical or developmental structure was found to be abnormal in a mutant fly. In order to integrate those anatomical abnormalities in the PhenomeNET ontology and therefore use them in cross-species phenotype analysis, we added the abnormal anatomical structures as new classes in PhenomeNET and associated the alleles with these classes:

$$?FBbtAB \equiv \exists has\_part.(\exists part\_of.?FBbt \sqcap \exists has\_quality.(Quality \sqcap \exists has\_modifier.Abnormal)) \quad (2)$$

Quality and *Abnormal* are classes from the PATO ontology. For example, `FBal0148512` is an allele associated with wing abnormalities (Végh and Basler, 2003), and we associate the allele with the newly defined class *Phenotypic abnormality of wing*, defined accordingly to the pattern in 2 where $?FBbt$ is the class *Wing* from the fly anatomy ontology.

Similarly to PhenomeNET, we define homologous and analogous anatomical structures as equivalent (for the purpose of the ontology). For example, we defined the nervous system in fly (`FBbt:00005093`) to be equivalent to the nervous system in the zebrafish (`ZFA:0000396`), and the nervous system in the Uberon multi-species anatomy ontology (`UBERON:0001016`). Through these equivalence class assertions, we can deductively infer an equivalence between *nervous system phenotype* (`MP:0003631`), *abnormal neuroanatomy* (`FBcv:0000435`), and *Abnormality of the nervous system* (`HP:0000707`), thereby enabling the direct comparison of mouse, fly, and human phenotypes.

The extended PhenomeNET ontology (Pheno-e) contains 16,083 human phenotype (HP) classes, 13,698 mammalian phenotype (MP) classes, 35,954 Zebrafish phenotype classes (PHENO classes, defined in Pheno-e), 3,111 fly phenotype classes (FBcv

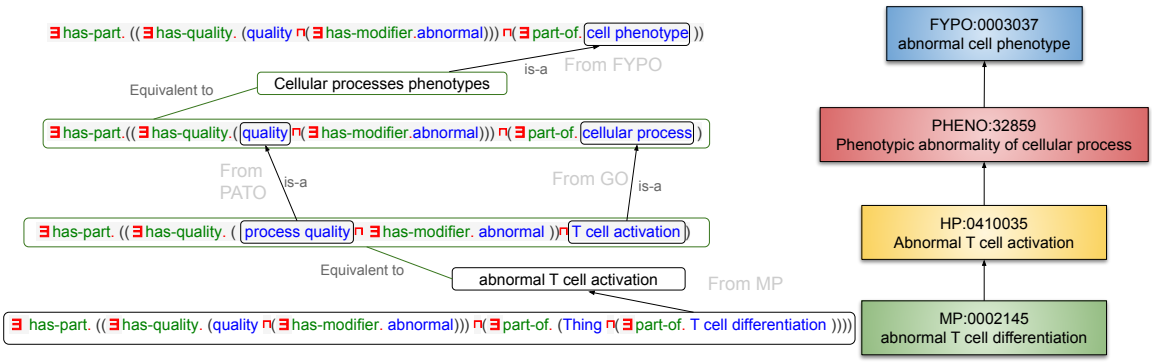

Fig. 3. Example of inferred hierarchy relating classes from different organisms phenotypes.

Table 5. Pheno-e summary of direct and indirect inferred sub-classes and super-classes axioms between different organisms phenotype classes

Subclass Matrix

	Human	Mammalian	Zebrafish	Drosophila	Yeast
Human Phenotypes	5484	2086	1189	79	54
Mammalian phenotypes	2669	5806	2132	98	110
Zebrafish phenotypes	4016	6155	17617	398	937
Drosophila phenotypes	120	173	209	4143	28
Fission yeast phenotpes	62	128	332	25	2457

Superclass Matrix

	Human	Mammalian	Zebrafish	Drosophila	Yeast
Human Phenotypes	15795	15389	15573	15573	7804
Mammalian phenotypes	13166	14096	13160	13160	7797
Zebrafish phenotypes	35954	30793	35951	35951	17916
Drosophila phenotypes	3017	2530	3016	19153	153
Fission yeast phenotpes	5807	3987	5807	5807	7322

classes and abnormal anatomy FBbtAB classes), and 7,636 classes of yeast phenotype (FYPO) classes.

We use phenotype datasets consisting of 8,031 OMIM diseases annotated with HP classes, 14,210 mouse genes annotated with MP classes, 6,182 zebrafish genes annotated with PHENO classes, 13,512 fly genes annotated with FBbtAB and FBcv classes , and 4,443 yeast genes annotated with FYPO classes.

Using automated reasoning over the Pheno-e ontology, we are able to infer relations between classes from different organisms; in particular, we are able to automatically infer whether two classes are equivalent or whether one class is a subclass of another class. We show the number of inferred relations in Pheno-e between the different species in Table 5, and for uPheno in Table 6. The tables show that it is possible to relate a large number of model organism phenotypes to human phenotypes through the Pheno-e and uPheno ontologies.

Figure 3 illustrates an example of inferred relations between phenotype classes of different organisms and resources. In this example, the class *Abnormal T cell activation* (`HP:0410035`) has as a (zebrafish) superclass *Phenotypic abnormality of cellular process* (`PHENO:32859`) . This inference was made because of the background available from PATO and GO as the class *process quality* (`PATO:0001236`) is a subclass of *quality* (`PATO:0000001`), and *T cell activation* (`GO:0042110`) is a subclass of *cellular process* (`GO:0009987`).

Phenotype similarity

We apply a set of different methods to compare the similarity of phenotypes associated with a loss of function model organism mutant and human disease phenotypes.

Resnik semantic similarity

We calculated Resnik similarity (Resnik, 1995) between genes and diseases annotated with phenotype classes; the use of integrated phenotype ontologies enables the direct comparison of phenotypes.

Resnik's similarity is a similarity measure based on information content, defined as

$$IC(class) = -log(p(class)) \quad (3)$$

where the probability of a class is defined as the frequency of annotation with the class. The similarity between two ontology classes is defined as the information content of the most informative common ancestor (MICA) of two classes:

$$sim_{Resnik}(g_i, d_j) = IC(MICA(g_i, d_j)) \quad (4)$$

Table 6. uPheno summary of direct and indirect inferred shared ancestor generic class count between different organisms phenotype classes

Shared Ancestor Generic class count

	Human	Mammalian	Zebrafish	Drosophila	Yeast
Human Phenotypes	6414	2903	1426	91	67
Mammalian phenotypes	2903	7786	2007	87	75
Zebrafish phenotypes	1426	2007	6446	139	142
Drosophila phenotypes	91	87	139	177	53
Fission yeast phenotpes	67	75	142	53	177

As we compare groups of classes, we use the best match average method (BMA) (Pesquita et al., 2009) to calculate the similarity between genes and diseases:

$$sim_{BMA}(gene, disease) = \frac{\sum_{i=1}^{g_n} \max_{1\leq j\leq d_n}(sim_{Resnik}(g_i, d_j))}{2 * g_n} + \frac{\sum_{i=1}^{d_n} \max_{1\leq j\leq g_n}(sim_{Resnik}(d_i, g_j))}{2 * d_n} \quad (5)$$

OPA2Vec

As second method, we "embed" phenotypes in a real-valued vector space using OPA2Vec (Smaili et al., 2018*a*); an embedding is a structure-preserving map from one algebraic structure (ontology axioms) into another (vector space), i.e., an embedding preserves (some) properties of the first structure within the second. OPA2Vec is mainly based on preserving syntactic relations in asserted and inferred ontology axioms.

We generated embeddings for diseases and genes using OPA2Vec based on the phenotypes associated with the diseases and genes and the axioms in the phenotype ontology. For the training, we use the skip gram, set `mincount` to 0, embedding size to 100, and window size to 5. We then compute similarity between genes and diseases based on the cosine similarity of their embeddings.

DL2Vec and OWL2Vec*

Another method we used to generate feature embeddings is the DL2Vec (Chen et al., 2020) method. DL2Vec converts description logic axioms into an undirected graph representations and uses a random walk to explore the graph; the walks are then treated as sentences and encoded using a language model. The graph is generated from the ontology axioms, and each phenotype class becomes one node in this graph; we add the gene and disease identifiers to this graph and connect them to the phenotype classes with which they are annotated. To generate the walks we choose the walk length to be 30, with 50 number of walks, and we used a skip gram method with window size set to 10, `mincount` to 1, and embedding size 100.

OWL2Vec*(Chen et al., 2021) is an embedding method similar to DL2Vec and based on a similar graph representation. OWL2Vec* graphs are directed and do not include equivalence or disjoint class axioms. We use random walker with walk depth 7 and 30 iterations with projection on structure document. and we used a skip gram method with window size set to 5, `mincount` to 1, and negatives to 5 and embedding size 100. For both DL2Vec and OWL2Vec* embeddings, we compare the phenotypic similarity between genes and diseases using cosine similarity.

Prediction of gene–disease associations

In addition to predicting gene–disease associations based on phenotypic similarity, we also use supervised prediction of these associations. For this purpose, we use a multilayer perceptron (MLP) with a single hidden layer. The input of the MLP is the concatenated embeddings of a disease and a gene. We use a hidden layer half the size of the input and a binary output using a sigmoid function, indicating whether the gene and disease are associated through a gene–disease relation or not; we further use the value of the sigmoid to rank genes for a disease. We randomly generated five negatives to each positive. For the training, we use the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 0.001 and maximum number of iterations of 300. To evaluate, we used 10-fold cross validation, stratified by diseases.

Naïve classifier

We hypothesize that some of our results are due to imbalanced or biased data. To test this hypothesis, we define a "naïve" classifier that predicts gene-disease associations on the basis of the information content of phenotypes of a gene alone; this "classifier" ranks all genes based on the sum of the information content of their phenotype annotations, sorts genes in descending order, and predicts the same ranked list of genes for each disease (i.e., the classifier is independent of the disease). The aim of this "naïve" classifier is to test whether genes annotated with more and more specific phenotypes are generally more likely to be associated with a disease, i.e., it tests for a kind of annotation bias.

Evaluating predictive performance

Our evaluation is based on estimating how well the different approaches rank disease-associated genes given a set of disease-associated phenotypes, for phenotypes from different organisms. Higher phenotypic similarity between a gene and a disease indicates higher likelihood that the gene (or its human ortholog) is associated with that disease. We evaluated two data sets from the MGI file `MGI_DO.rpt`, one for human gene–disease associations from OMIM and another MGI-curated dataset of mouse models of human disease.

For the evaluation, for each disease D_i in our evaluation set, we rank all genes $G_1, ..., G_n$ based on their phenotypic similarity to D_i. For each disease D_i, we determine the rank (or ranks) at which the associated gene (or genes) appear in this ranked list. We use this information to determine the false positive and true positive rate at each rank; we average the true and false positive rates across all diseases and use this to determine the receiver operating characteristic (ROC) curve and the area under the ROC curve (ROCAUC).

When using supervised methods to predict gene–disease associations, we use the same evaluation in a 10-fold cross validation setting, and we rank genes based on the output of the sigmoid unit of our machine learning model.

Implementation

We used several tools and libraries, such as the OWLAPI for generating the ontology groovy and python scripts for data processing.

We also used several python libraries like sklearn, numpy, pandas, PyTorch (Paszke et al., 2017) for the supervised learning. For calculating Resnik semantic similarity we used the Semantic Measures Library (SML) (Harispe et al., 2013).

Acknowledgements

We acknowledge use of resources from the KAUST Supercomputing Laboratory. PNS acknowledges the support of The Alan Turing Institute.

Competing interests

The authors declare that they have no competing interests.

Contribution

PNS and RH conceived of the experiments; SMA, PNS and RH designed and interpreted the experiments; SMA performed and implemented all computational and statistical experiments; PNS and RH acquired the funding.

Funding

This work was supported by funding from King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No. URF/1/3790-01-01, URF/1/4355-01-01, and FCC/1/1976-34-01.

Data availability

All data and software required to reproduce our results are freely available at https://github.com/bio-ontology-research-group/mo-phenotype-analysis

REFERENCES

Agapite, J., Albou, L.-P., Aleksander, S. A., Alexander, M., Anagnostopoulos, A. V., Antonazzo, G., Argasinska, J., Arnaboldi, V., Attrill, H., Becerra, A., Bello, S. M., Blake, J. A., Blodgett, O., Bradford, Y. M., Bult, C. J., Cain, S., Calvi, B. R., Carbon, S., Chan, J., Chen, W. J., Cherry, J. M., Cho, J., Christie, K. R., Crosby, M. A., Davis, P., da Veiga Beltrame, E., Pons, J. L. D., D'Eustachio, P., Diamantakis, S., Dolan, M. E., dos Santos, G., Douglass, E., Dunn, B., Eagle, A., Ebert, D., Engel, S. R., Fashena, D., Foley, S., Frazer, K., Gao, S., Gibson, A. C., Gondwe, F., Goodman, J., Gramates, L. S., Grove, C. A., Hale, P., Harris, T., Hayman, G. T., Hill, D. P., Howe, D. G., Howe, K. L., Hu, Y., Jha, S., Kadin, J. A., Kaufman, T. C., Kalita, P., Karra, K., Kishore, R., Kwitek, A. E., Laulederkind, S. J. F., Lee, R., Longden, I., Luypaert, M., MacPherson, K. A., Martin, R., Marygold, S. J., Matthews, B., McAndrews, M. S., Millburn, G., Miyasato, S., Motenko, H., Moxon, S., Muller, H.-M., Mungall, C. J., Muruganujan, A., Mushayahama, T., Nalabolu, H. S., Nash, R. S., Ng, P., Nuin, P., Paddock, H., Paulini, M., Perrimon, N., Pich, C., Quinton-Tulloch, M., Raciti, D., Ramachandran, S., Richardson, J. E., Gelbart, S. R., Ruzicka, L., Schaper, K., Schindelman, G., Shimoyama, M., Simison, M., Shaw, D. R., Shrivatsav, A., Singer, A., Skrzypek, M., Smith, C. M., Smith, C. L., Smith, J. R., Stein, L., Sternberg, P. W., Tabone, C. J., Thomas, P. D., Thorat, K., Thota, J., Toro, S., Tomczuk, M., Trovisco, V., Tutaj, M. A., Tutaj, M., Urbano, J.-M., Auken, K. V., Slyke, C. E. V., Wang, Q., Wang, S.-J., Weng, S., Westerfield, M., Williams, G., Wilming, L. G., Wong, E. D., Wright, A., Yook, K., Zarowiecki, M., Zhou, P. and Zytkovicz, M. (2022), 'Harmonizing model organism data in the alliance of genome resources', *Genetics* **220**(4).
URL: *https://doi.org/10.1093/genetics/iyac022*

Aitman, T. J., Boone, C., Churchill, G. A., Hengartner, M. O., Mackay, T. F. C. and Stemple, D. L. (2011), 'The future of model organisms in human disease research', *Nature Reviews Genetics* **12**(8), 575–582.
URL: *https://doi.org/10.1038/nrg3047*

Alghamdi, S. M., Sundberg, B. A., Sundberg, J. P., Schofield, P. N. and Hoehndorf, R. (2019), 'Quantitative evaluation of ontology design patterns for combining pathology and anatomy ontologies', *Scientific reports* **9**(1), 1–12.

Alliance of Genome Resources Portal: unified model organism research platform (2020), *Nucleic acids research* **48**(D1), D650–D658.

Amberger, J. S., Bocchini, C. A., Scott, A. F. and Hamosh, A. (2018), 'OMIM.org: leveraging knowledge across phenotype–gene relationships', *Nucleic Acids Research* **47**(D1), D1038–D1043.
URL: *https://doi.org/10.1093/nar/gky1151*

Amberger, J. S. and Hamosh, A. (2017), 'Searching online mendelian inheritance in man (omim): a knowledgebase of human genes and genetic phenotypes', *Current protocols in bioinformatics* **58**(1), 1–2.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000), 'Gene ontology: tool for the unification of biology', *Nature genetics* **25**(1), 25.

Baldridge, D., , Wangler, M. F., Bowman, A. N., Yamamoto, S., Schedl, T., Pak, S. C., Postlethwait, J. H., Shin, J., Solnica-Krezel, L., Bellen, H. J. and Westerfield, M. (2021), 'Model organisms contribute to diagnosis and discovery in the undiagnosed diseases network: current state and a future vision', *Orphanet Journal of Rare Diseases* **16**(1).
URL: *https://doi.org/10.1186/s13023-021-01839-9*

Barabási, A.-L., Gulbahce, N. and Loscalzo, J. (2010), 'Network medicine: a network-based approach to human disease', *Nature Reviews Genetics* **12**(1), 56–68.
URL: *https://doi.org/10.1038/nrg2918*

Bello, S. M., and Eppig, J. T. (2016), 'Inferring gene-to-phenotype and gene-to-disease relationships at mouse genome informatics: challenges and solutions', *Journal of Biomedical Semantics* **7**(1).
URL: *https://doi.org/10.1186/s13326-016-0054-4*

Boudellioua, I., Razali, R. B. M., Kulmanov, M., Hashish, Y., Bajic, V. B., Goncalves-Serra, E., Schoenmakers, N., Gkoutos, G. V., Schofield, P. N. and Hoehndorf, R. (2017), 'Semantic prioritization of novel causative genomic variants', *PLOS Computational Biology* **13**(4), e1005500.
URL: *https://doi.org/10.1371/journal.pcbi.1005500*

Brommage, R., Powell, D. R. and Vogel, P. (2019), 'Predicting human disease mutations and identifying drug targets from mouse gene knockout phenotyping campaigns', *Disease Models & Mechanisms* **12**(5).
URL: *https://doi.org/10.1242/dmm.038224*

Brown, S. D. M. (2021), 'Advances in mouse genetics for the study of human disease', *Human Molecular Genetics* **30**(R2), R274–R284.
URL: *https://doi.org/10.1093/hmg/ddab153*

Brown, S. D. M., Holmes, C. C., Mallon, A.-M., Meehan, T. F., Smedley, D. and Wells, S. (2018), 'High-throughput mouse phenomics for characterizing mammalian gene function', *Nature Reviews Genetics* **19**(6), 357–370.
URL: *https://doi.org/10.1038/s41576-018-0005-2*

Cacheiro, P., , Haendel, M. A. and Smedley, D. (2019), 'New models for human disease from the international mouse phenotyping consortium', *Mammalian Genome* **30**(5-6), 143–150.
URL: *https://doi.org/10.1007/s00335-019-09804-5*

Cervelli, T. and Galli, A. (2021), 'Yeast as a tool to understand the significance of human disease-associated gene variants', *Genes* **12**(9), 1303.
URL: *https://doi.org/10.3390/genes12091303*

Chen, J., Althagafi, A. T. and Hoehndorf, R. (2020), 'Predicting candidate genes from phenotypes, functions, and anatomical site of expression'.

Chen, J., Hu, P., Jimenez-Ruiz, E., Holter, O. M., Antonyrajah, D. and Horrocks, I. (2021), 'Owl2vec*: Embedding of owl ontologies', *Machine Learning* pp. 1–33.

Chesmore, K., Bartlett, J. and Williams, S. M. (2017), 'The ubiquity of pleiotropy in human disease', *Human Genetics* **137**(1), 39–44.
URL: *https://doi.org/10.1007/s00439-017-1854-z*

Cipriani, V., Pontikos, N., Arno, G., Sergouniotis, P. I., Lenassi, E., Thawong, P., Danis, D., Michaelides, M., Webster, A. R., Moore, A. T., Robinson, P. N., Jacobsen, J. O. and Smedley, D. (2020), 'An improved phenotype-driven tool for rare mendelian variant prioritization: Benchmarking exomiser on real patient whole-exome data', *Genes* **11**(4), 460.
URL: *https://doi.org/10.3390/genes11040460*

Cornish, A. J., David, A. and Sternberg, M. J. (2018), 'Phenorank: reducing study bias in gene prioritization through simulation', *Bioinformatics* **34**(12), 2087–2095.

Costa, M., Reeve, S., Grumbling, G. and Osumi-Sutherland, D. (2013), 'The drosophila anatomy ontology', *Journal of biomedical semantics* **4**(1), 1–11.

Dahdul, W. M., Cui, H., Mabee, P. M., Mungall, C. J., Osumi-Sutherland, D., Walls, R. L. and Haendel, M. A. (2014), 'Nose to tail, roots to shoots: spatial descriptors for phenotypic diversity in the biological spatial ontology', *Journal of Biomedical Semantics* **5**(1), 34.

Dahme, T., Katus, H. A. and Rottbauer, W. (2009), 'Fishing for the genetic basis of cardiovascular disease', *Disease Models & Mechanisms* **2**(1-2), 18–22.
URL: *https://doi.org/10.1242/dmm.000687*

Diehl, A. D., Meehan, T. F., Bradford, Y. M., Brush, M. H., Dahdul, W. M., Dougall, D. S., He, Y., Osumi-Sutherland, D., Ruttenberg, A., Sarntivijai, S. et al. (2016), 'The cell ontology 2016: enhanced content, modularization, and ontology interoperability', *Journal of biomedical semantics* **7**(1), 44.

Eppig, J. T., Smith, C. L., Blake, J. A., Ringwald, M., Kadin, J. A., Richardson, J. E. and Bult, C. J. (2017), Mouse genome informatics (mgi): resources for mining mouse genetic, genomic, and biological data in support of primary and translational research, *in* 'Systems Genetics', Springer, pp. 47–73.

Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R. M. and Carter, N. P. (2009), 'Decipher: database of chromosomal imbalance and phenotype in humans using ensembl resources', *The American Journal of Human Genetics* **84**(4), 524–533.

Gkoutos, G. V., Green, E. C., Mallon, A.-M., Hancock, J. M. and Davidson, D. (2005), 'Using ontologies to describe mouse phenotypes', *Genome biology* **6**(1), R8.

Gkoutos, G. V., Schofield, P. N. and Hoehndorf, R. (2012), The neurobehavior ontology: an ontology for annotation and integration of behavior and behavioral phenotypes, *in* 'International review of neurobiology', Vol. 103, Elsevier, pp. 69–87.

Gkoutos, G. V., Schofield, P. N. and Hoehndorf, R. (2017), 'The anatomy of phenotype ontologies: principles, properties and applications', *Briefings in Bioinformatics* **19**(5), 1008–1021.
URL: *https://doi.org/10.1093/bib/bbx035*

Grimm, D. G., Azencott, C., Aicheler, F., Gieraths, U., MacArthur, D. G., Samocha, K. E., Cooper, D. N., Stenson, P. D., Daly, M. J., Smoller, J. W., Duncan, L. E. and Borgwardt, K. M. (2015), 'The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity', *Human Mutation* **36**(5), 513–523.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/humu.22768*

Harispe, S., Ranwez, S., Janaqi, S. and Montmain, J. (2013), 'The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies', *Bioinformatics* **30**(5), 740–742.

Harris, M. A., Lock, A., Bähler, J., Oliver, S. G. and Wood, V. (2013), 'Fypo: the fission yeast phenotype ontology', *Bioinformatics* **29**(13), 1671–1678.

Hmeljak, J. and Justice, M. J. (2019), 'From gene to treatment: supporting rare disease translational research through model systems', *Disease Models & Mechanisms* **12**(2).
URL: *https://doi.org/10.1242/dmm.039271*

Hoehndorf, R., Schofield, P. N. and Gkoutos, G. V. (2011), 'Phenomenet: a whole-phenome approach to disease gene discovery', *Nucleic acids research* **39**(18), e119–e119.

Hoehndorf, R., Schofield, P. N. and Gkoutos, G. V. (2015), 'The role of ontologies in biological and biomedical research: a functional perspective', *Briefings in bioinformatics* **16**(6), 1069–1080.

Kingma, D. P. and Ba, J. (2014), 'Adam: A method for stochastic optimization', *arXiv preprint arXiv:1412.6980* .

Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N. L., Matentzoglu, N., McMurry, J. A. et al. (2018), 'Expansion of the human phenotype ontology (hpo) knowledge base and resources', *Nucleic acids research* **47**(D1), D1018–D1027.

Köhler, S., Gargano, M., Matentzoglu, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., Danis, D., Balagura, G., Baynam, G., Brower, A. M. et al. (2021), 'The human phenotype ontology in 2021', *Nucleic acids research* **49**(D1), D1207–D1217.

Köhler, S., Schulz, M. H., Krawitz, P., Bauer, S., Dölken, S., Ott, C. E., Mundlos, C., Horn, D., Mundlos, S. and Robinson, P. N. (2009), 'Clinical diagnostics in human genetics with semantic similarity searches in ontologies', *The American Journal of Human Genetics* **85**(4), 457–464.
URL: *https://doi.org/10.1016/j.ajhg.2009.09.003*

Kulmanov, M. and Hoehndorf, R. (2017), 'Evaluating the effect of annotation size on measures of semantic similarity', *Journal of Biomedical Semantics* **8**(1).
URL: *https://doi.org/10.1186/s13326-017-0119-z*

Kulmanov, M. and Hoehndorf, R. (2020), 'DeepPheno: Predicting single gene loss-of-function phenotypes using an ontology-aware hierarchical classifier', *PLOS Computational Biology* **16**(11), e1008453.
URL: *https://doi.org/10.1371/journal.pcbi.1008453*

Kulmanov, M., Smaili, F. Z., Gao, X. and Hoehndorf, R. (2020), 'Semantic similarity and machine learning with ontologies', *Briefings in Bioinformatics* **22**(4).
URL: *https://doi.org/10.1093/bib/bbaa199*

Larkin, A., Marygold, S. J., Antonazzo, G., Attrill, H., dos Santos, G., Garapati, P. V., Goodman, J. L., Gramates, L. S., Millburn, G., Strelets, V. B., Tabone, C. J., Thurmond, J., Perrimon, N., Gelbart, S. R., Agapite, J., Broll, K., Crosby, M., dos Santos, G., Falls, K., Gramates, L. S., Jenkins, V., Longden, I., Matthews, B., Sutherland, C., Tabone, C. J., Zhou, P., Zytkovicz, M., Brown, N., Antonazzo, G., Attrill, H., Garapati, P., Larkin, A., Marygold, S., McLachlan, A., Millburn, G., Pilgrim, C., Ozturk-Colak, A., Trovisco, V., Kaufman, T., Calvi, B., Goodman, J., Strelets, V., Thurmond, J., Cripps, R. and and, T. L. (2020), 'FlyBase: updates to the drosophila melanogaster knowledge base', *Nucleic Acids Research* **49**(D1), D899–D907.
URL: *https://doi.org/10.1093/nar/gkaa1026*

Matentzoglu, N., Osumi-Sutherland, D., Balhoff, J. P., Bello, S., Bradford, Y., Cardmody, L., Grove, C., Harris, M. A., Harris, N., Köhler, S. et al. (2019), 'upheno 2: Framework for standardised representation of phenotypes across species', *F1000Research* **8**.
URL: *https://f1000research.com/posters/8-403*

McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B. and Marcotte, E. M. (2010), 'Systematic discovery of nonobvious human disease models through orthologous phenotypes', *Proceedings of the National Academy of Sciences* **107**(14), 6544–6549.
URL: *https://doi.org/10.1073/pnas.0910200107*

Meehan, T. F., , Conte, N., West, D. B., Jacobsen, J. O., Mason, J., Warren, J., Chen, C.-K., Tudose, I., Relac, M., Matthews, P., Karp, N., Santos, L., Fiegel, T., Ring, N., Westerberg, H., Greenaway, S., Sneddon, D., Morgan, H., Codner, G. F., Stewart, M. E., Brown, J., Horner, N., Haendel, M., Washington, N., Mungall, C. J., Reynolds, C. L., Gallegos, J., Gailus-Durner, V., Sorg, T., Pavlovic, G., Bower, L. R., Moore, M., Morse, I., Gao, X., Tocchini-Valentini, G. P., Obata, Y., Cho, S. Y., Seong, J. K., Seavitt, J., Beaudet, A. L., Dickinson, M. E., Herault, Y., Wurst, W., de Angelis, M. H., Lloyd, K. C. K., Flenniken, A. M., Nutter, L. M. J., Newbigging, S., McKerlie, C., Justice, M. J., Murray, S. A., Svenson, K. L., Braun, R. E., White, J. K., Bradley, A., Flicek, P., Wells, S., Skarnes, W. C., Adams, D. J., Parkinson, H., Mallon, A.-M., Brown, S. D. M. and Smedley, D. (2017), 'Disease model discovery from 3, 328 gene knockouts by the international mouse phenotyping consortium', *Nature Genetics* **49**(8), 1231–1238.
URL: *https://doi.org/10.1038/ng.3901*

Mungall, C. J., Gkoutos, G. V., Smith, C. L., Haendel, M. A., Lewis, S. E. and Ashburner, M. (2010), 'Integrating phenotype ontologies across multiple species', *Genome Biology* **11**(1), R2.
URL: *https://doi.org/10.1186/gb-2010-11-1-r2*

Mungall, C. J., McMurry, J. A., Köhler, S., Balhoff, J. P., Borromeo, C., Brush, M., Carbon, S., Conlin, T., Dunn, N., Engelstad, M., Foster, E., Gourdine, J., Jacobsen, J. O., Keith, D., Laraway, B., Lewis, S. E., NguyenXuan, J., Shefchek, K., Vasilevsky, N., Yuan, Z., Washington, N., Hochheiser, H., Groza, T., Smedley, D., Robinson, P. N. and Haendel, M. A. (2016), 'The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species', *Nucleic Acids Research* **45**(D1), D712–D722.
URL: *https://doi.org/10.1093/nar/gkw1128*

Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E. and Haendel, M. A. (2012), 'Uberon, an integrative multi-species anatomy ontology', *Genome biology* **13**(1), R5.

Narumanchi, S., Wang, H., Perttunen, S., Tikkanen, I., Lakkisto, P. and Paavola, J. (2021), 'Zebrafish heart failure models', *Frontiers in Cell and Developmental Biology* **9**.
URL: *https://doi.org/10.3389/fcell.2021.662583*

O'Brien, K. P. (2004), 'Inparanoid: a comprehensive database of eukaryotic orthologs', *Nucleic Acids Research* **33**(Database issue), D476–D480.
URL: *https://doi.org/10.1093/nar/gki107*

Oellrich, A., Koehler, S., Washington, N., Mungall, C., Lewis, S., Haendel, M., Robinson, P. N. and and, D. S. (2014), 'The influence of disease categories on gene candidate predictions from model organism phenotypes', *Journal of Biomedical Semantics* **5**(Suppl 1), S4.
URL: *https://doi.org/10.1186/2041-1480-5-s1-s4*

Online Mendelian Inheritance in Man (OMIM) (2020), *McKusick-Nathans Institute of Genetic Medicine* .
URL: *http://omim.org/*

Osumi-Sutherland, D., Marygold, S. J., Millburn, G. H., McQuilton, P. A., Ponting, L., Stefancsik, R., Falls, K., Brown, N. H. and Gkoutos, G. V. (2013), 'The drosophila phenotype ontology', *Journal of biomedical semantics* **4**(1), 1–10.

Oti, M. and Brunner, H. G. (2007), 'The modular nature of genetic diseases', *Clin Genet* **71**(1), 1–11.
URL: *https://pubmed.ncbi.nlm.nih.gov/17204041/*

Oti, M., van Reeuwijk, J., Huynen, M. A. and Brunner, H. G. (2008), 'Conserved co-expression for candidate disease gene prioritization', *BMC Bioinformatics* **9**(1).
URL: *https://doi.org/10.1186/1471-2105-9-208*

Pardo, P. S. and Boriek, A. M. (2020), 'SIRT1 regulation in ageing and obesity', *Mechanisms of Ageing and Development* **188**, 111249.
URL: *https://doi.org/10.1016/j.mad.2020.111249*

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L. and Lerer, A. (2017), Automatic differentiation in pytorch, *in* 'NIPS-W'.

Patton, M. A. and Afzal, A. R. (2002), 'Robinow syndrome', *Journal of Medical Genetics* **39**(5), 305.
URL: *http://jmg.bmj.com/content/39/5/305.abstract*

Pesquita, C., Faria, D., Falcão, A. O., Lord, P. and Couto, F. M. (2009), 'Semantic similarity in biomedical ontologies', *PLoS Computational Biology* **5**(7), e1000443.
URL: *https://doi.org/10.1371/journal.pcbi.1000443*

Peterson, K. A. and Murray, S. A. (2021), 'Progress towards completing the mutant mouse null resource', *Mammalian Genome* .
URL: *https://doi.org/10.1007/s00335-021-09905-0*

Prykhozhij, S. V. and Berman, J. N. (2018), 'Zebrafish knock-ins swim into the mainstream', *Disease Models & Mechanisms* **11**(10).
URL: *https://doi.org/10.1242/dmm.037515*

Resnik, P. (1995), Using information content to evaluate semantic similarity, *in* 'Proceedings of the 14th International Joint Conference on Artificial Intelligence', pp. 448–453.

Resnik, P. (1999), 'Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language', *Journal of artificial intelligence research* **11**, 95–130.

Ringwald, M., Richardson, J. E., Baldarelli, R. M., Blake, J. A., Kadin, J. A., Smith, C. and Bult, C. J. (2021), 'Mouse genome informatics (MGI): latest news from MGD and GXD', *Mammalian Genome* .
URL: *https://doi.org/10.1007/s00335-021-09921-0*

Schriml, L. M., Munro, J. B., Schor, M., Olley, D., McCracken, C., Felix, V., Baron, J. A., Jackson, R., Bello, S. M., Bearer, C., Lichenstein, R., Bisordi, K., Dialo, N. C., Giglio, M. and Greene, C. (2021), 'The human disease ontology 2022 update', *Nucleic Acids Research* **50**(D1), D1255–D1261.
URL: *https://doi.org/10.1093/nar/gkab1063*

Schriml, L. M., Munro, J. B., Schor, M., Olley, D., McCracken, C., Felix, V., Baron, J. A., Jackson, R., Bello, S. M., Bearer, C. et al. (2022), 'The human disease ontology 2022 update', *Nucleic acids research* **50**(D1), D1255–D1261.

Shefchek, K. A., Harris, N. L., Gargano, M., Matentzoglu, N., Unni, D., Brush, M., Keith, D., Conlin, T., Vasilevsky, N., Zhang, X. A. et al. (2020), 'The monarch initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species', *Nucleic acids research* **48**(D1), D704–D715.

Smaili, F., Gao, X. and Hoehndorf, R. (2018*a*), 'Opa2vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction', *Bioinformatics* .

Smaili, F. Z., Gao, X. and Hoehndorf, R. (2018*b*), 'Onto2vec: joint vector-based representation of biological entities and their ontology-based annotations', *Bioinformatics* **34**(13), i52–i60.
URL: *https://doi.org/10.1093/bioinformatics/bty259*

Smedley, D., Jacobsen, J. O. B., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Buske, O. J., Washington, N. L., Bone, W. P., Haendel, M. A. and Robinson, P. N. (2015), 'Next-generation diagnostics and disease-gene discovery with the exomiser', *Nature Protocols* **10**(12), 2004–2015.
URL: *https://doi.org/10.1038/nprot.2015.124*

Smedley, D., Oellrich, A., Kohler, S., Ruef, B., Westerfield, M., Robinson, P., Lewis, S. and and, C. M. (2013), 'PhenoDigm: analyzing curated annotations to associate animal models with human diseases', *Database* **2013**(0), bat025–bat025.
URL: *https://doi.org/10.1093/database/bat025*

Smedley, D. and Robinson, P. N. (2015), 'Phenotype-driven strategies for exome prioritization of human mendelian disease genes', *Genome Medicine* **7**(1).
URL: *https://doi.org/10.1186/s13073-015-0199-2*

Smedley, D., Schubach, M., Jacobsen, J. O., Köhler, S., Zemojtel, T., Spielmann, M., Jäger, M., Hochheiser, H., Washington, N. L., McMurry, J. A., Haendel, M. A., Mungall, C. J., Lewis, S. E., Groza, T., Valentini, G. and Robinson, P. N. (2016), 'A whole-genome analysis framework for effective identification of pathogenic regulatory variants in mendelian disease', *The American Journal of Human Genetics* **99**(3), 595–606.
URL: *https://doi.org/10.1016/j.ajhg.2016.07.005*

Smedley, D., Smith, K. R., Martin, A., Thomas, E. A., McDonagh, E. M., Cipriani, V., Ellingford, J. M., Arno, G., Tucci, A., Vandrovcova, J. et al. (2021), '100,000 genomes pilot on rare-disease diagnosis in health care — preliminary report', *New England Journal of Medicine* **385**(20), 1868–1880.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L. and Rosse, C. (2005), 'Relations in biomedical ontologies', *Genome biology* **6**(5), 1–15.

Smith, C. L. and Eppig, J. T. (2012), 'The mammalian phenotype ontology as a unifying standard for experimental and high-throughput phenotyping data', *Mammalian Genome* **23**(9-10), 653–668.
URL: *https://doi.org/10.1007/s00335-012-9421-3*

Sundberg, J. P., Roopenian, D. C., Liu, E. T. and Schofield, P. N. (2013), 'The cinderella effect: searching for the best fit between mouse models and human diseases.', *The Journal of investigative dermatology* **133 11**, 2509–2513.

Thiele, F., Cohrs, C. M., Flor, A., Lisse, T. S., Przemeck, G. K. H., Horsch, M., Schrewe, A., Gailus-Durner, V., Ivandic, B., Katus, H. A., Wurst, W., Reisenberg, C., Chaney, H., Fuchs, H., Hans, W., Beckers, J., Marini, J. C. and de Angelis, M. H. (2012), 'Cardiopulmonary dysfunction in the osteogenesis imperfecta mouse model aga2 and human patients are caused by bone-independent mechanisms', *Human Molecular Genetics* **21**(16), 3535–3545.
URL: *https://doi.org/10.1093/hmg/dds183*

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., Matthews, B. B., Millburn, G., Antonazzo, G., Trovisco, V. et al. (2018), 'Flybase 2.0: the next generation', *Nucleic acids research* **47**(D1), D759–D765.

Train, C.-M., Glover, N. M., Gonnet, G. H., Altenhoff, A. M. and Dessimoz, C. (2017), 'Orthologous matrix (oma) algorithm 2.0: more robust to asymmetric evolutionary rates and more scalable hierarchical orthologous group inference', *Bioinformatics* **33**(14), i75–i82.

Van Slyke, C. E., Bradford, Y. M., Westerfield, M. and Haendel, M. A. (2014), 'The zebrafish anatomy and stage ontologies: representing the anatomy and development of danio rerio', *Journal of biomedical semantics* **5**(1), 12.

Varela, F., Maturana, H. and Uribe, R. (1974), 'Autopoiesis: The organization of living systems, its characterization and a model', *Biosystems* **5**(4), 187–196.
URL: *https://doi.org/10.1016/0303-2647(74)90031-8*

Végh, M. and Basler, K. (2003), 'A genetic screen for hedgehog targets involved in the maintenance of the drosophila anteroposterior compartment boundary', *Genetics* **163**(4), 1427–1438.

Wagner, G. P. and Zhang, J. (2011), 'The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms', *Nature Reviews Genetics* **12**(3), 204–213.
URL: *https://doi.org/10.1038/nrg2949*

Wangler, M. F., Yamamoto, S., Chao, H.-T., Posey, J. E., Westerfield, M., Postlethwait, J., Hieter, P., Boycott, K. M., Campeau, P. M. and and, H. J. B. (2017), 'Model organisms facilitate rare disease diagnosis and therapeutic research', *Genetics* **207**(1), 9–27.
URL: *https://doi.org/10.1534/genetics.117.203067*

Washington, N. L., Haendel, M. A., Mungall, C. J., Ashburner, M., Westerfield, M. and Lewis, S. E. (2009), 'Linking human diseases to animal models using ontology-based phenotype annotation', *PLoS Biology* **7**(11), e1000247.
URL: *https://doi.org/10.1371/journal.pbio.1000247*

Weinreich, S. S., Mangon, R., Sikkens, J., Teeuw, M. and Cornel, M. (2008), 'Orphanet: a european database for rare diseases', *Nederlands tijdschrift voor geneeskunde* **152**(9), 518–519.

Willyard, C. (2018), 'New human gene tally reignites debate', *Nature* **558**(7710), 354–356.