D1172–D1178 Nucleic Acids Research, 2022, Vol. 50, Database issue Published online 28 October 2021
https://doi.org/10.1093/nar/gkab939

ConVarT: a search engine for matching human genetic variants with variants from non-human species

Mustafa S. Pir[1], Halil I. Bilgin[2], Ahmet Sayici[2], Fatih Coşkun[2], Furkan M. Torun[1], Pei Zhao[3], Yahong Kang[3], Sebiha Cevik[1] and Oktay I. Kaplan[1,*]

[1]Rare Disease Laboratory, School of Life and Natural Sciences, Abdullah Gul University, Kayseri 38090, Turkey, [2]Department of Computer Engineering, Abdullah Gul University, Kayseri 38090, Turkey and [3]SunyBiotech Co., Ltd, Fuzhou 35000, China

Received July 06, 2021; Revised September 07, 2021; Editorial Decision September 22, 2021; Accepted October 14, 2021

ABSTRACT

The availability of genetic variants, together with phenotypic annotations from model organisms, facilitates comparing these variants with equivalent variants in humans. However, existing databases and search tools do not make it easy to scan for equivalent variants, namely ‘matching variants’ (MatchVars) between humans and other organisms. Therefore, we developed an integrated search engine called ConVarT (http://www.convart.org/) for matching variants between humans, mice, and *Caenorhabditis elegans*. ConVarT incorporates annotations (including phenotypic and pathogenic) into variants, and these previously unexploited phenotypic MatchVars from mice and *C. elegans* can give clues about the functional consequence of human genetic variants. Our analysis shows that many phenotypic variants in different genes from mice and *C. elegans*, so far, have no counterparts in humans, and thus, can be useful resources when evaluating a relationship between a new human mutation and a disease.

INTRODUCTION

The progress of the genomics field is accelerating with the aid of next-generation sequencing technologies, most recently exemplified by the launch of the Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD). These genomic databases were gathered from >140 000 randomly selected individuals, providing large-scale collections of human genetic variants (1,2). Humans possess over 700 million genetic variants in the protein-encoding and non-protein-coding regions of the genome. However, the growing number of human genetic variants does not automatically mean that our understanding of the functional effects of these variants concurrently increases (3). Due to their clinical importance, functional consequences are only available for a handful of genetic variants, while biological impacts are unavailable for most human genetic variants (4). The biological understanding of these genetic variants is crucial to diagnose diseases better and make therapeutic decisions regarding diseases, particularly noting each variant's especially discerning contribution to rare and complex diseases.

Though examining the functional effects of variants is an efficient strategy for categorizing them as being either pathological or benign, a considerable workload and financial demand are limiting factors for such experimental efforts (5–9). In this regard, model organisms are regularly used to examine the functional influence of clinically relevant variants, but unprecedented quantities of variant and phenotype data from model organisms largely go unnoticed. We assume that when an amino acid, where it shows a similarity with other homologous sequences, undergoes an intolerable change in non-human species, as observed with phenotype-changing variants, human-equivalent variants will likely disrupt the function of genes in humans. The basis for this claim is that the function of proteins is likely to be compromised when the positions of highly conserved amino acids are altered (10). Despite various orthology available search, there is no existing search engine that: (1) allows users to scan for equivalent variants between human and non-human species; (2) incorporates available experimental data from model organisms into corresponding variant positions and (3) enables users to submit the empirical variant data (11–15).

Here, we develop an integrated search tool and database called ConVarT (**Con**gruent clinical **Var**iation Visualization **T**ool) to search for equivalent variants (pathological variants, phenotypic variants, and variants of uncertain significance) between human and non-human species. As a proof of concept, we matched amino acid-changing variants from humans with variants from mice and the nematode *Caenorhabditis elegans* and generated a list of equivalent variants called matching variants (MatchVar) between

[*]To whom correspondence should be addressed. Tel: +90 541 716 5849; Email: oktay.kaplan@agu.edu.tr

human and mice or *C. elegans*. Furthermore, we integrated available annotations (pathological, phenotypic, etc.) for variants into the database, typically making it easy for users to search for MatchVars with phenotypic data. Through the use of MatchVars associated with functional data, our comprehensive database can provide an additional layer of proof for the functional interpretation of human variants.

RESULTS

ConVarT: a search tool for 'matching variants'

The availability of organism-specific databases (WormBase for *C. elegans* and Mutagenetix for mice) containing non-naturally occurring variants and phenotype data for *C. elegans* and mice enables us to equate them with human genetic variants from gnomAD, Clinical Variants (ClinVar), COSMIC and dbSNP (Figure 1) (2,16–21). We initially derived the human gene orthologs for *M. musculus* and *C. elegans*. We then collected amino acid-changing variants from gnomAD, ClinVar, COSMIC and dbSNP for humans, from WormBase for *C. elegans*, and from the Australian Phenome Bank and Mutagenetix for mice. gnomAD describes over 16 179 380 variations, while the ClinVar database deposits over 500 000 variants, including pathogenic and benign variants (2,20). COSMIC and dbSNP contain over 6 842 627 variations and 1,086,546 variations, respectively (16,19). Our analysis revealed that gnomAD, ClinVar, COSMIC and dbSNP have identical and non-identical variants (Figure 2A). Mutagenetix and WormBase present almost 800 000 variants from these organisms (17,18). To compare human variants with those from mice and *C. elegans*, we matched the protein sequences of the ortholog genes, followed by performing 2 379 397 multiple sequence alignments (MSAs) of all combinations of these protein sequences to ensure alignment of the corresponding amino acid positions. The schematic workflow of ConVarT is presented in Figure 1. We then integrated genetic variants and variant-specific annotations (such as pathogenicity, phenotypes, and allele frequency) from humans, mice, and *C. elegans* into corresponding amino acid positions in each gene. We restricted our comparisons to amino acid-changing variants, as amino acids are more conserved than non-amino acid-changing variants, and protein function is likely to be affected by an alteration in the corresponding amino acids. For visualization of MSAs together with variants from humans, mice, and *C. elegans*, we used in-house software (22). Previous studies revealed that many human variants overlap with amino acids undergoing post-translational modifications (PTMs) and are enriched at the protein domains. Therefore, Pfam protein domains for humans and 383k PTMs from PhosphoSitePlus were systematically integrated into corresponding positions (23,24).

Our analysis revealed that many equivalent amino acid substitutions exist between humans and these two organisms (Figure 2B). For example, both human AP1M1 and *C. elegans* UNC-101, the homologue of AP1M1, have the identical residue change at the corresponding positions (Human NP_001123996.1: p.P375S, Variant ID: rs1490425955 versus *C. elegans* NP_001040675: p.P362S, Variant ID: WBVar00679460). We have named these equivalent variants 'matching variants' (MatchVars) (Figure 2B, and Supplementary Tables S1 and S2). For a variant to be considered a MatchVar, the amino acid residue at the corresponding positions must be matched between the human gene and the ortholog genes. The Materials and Methods provide a detailed explanation. There are 50430 MatchVars between humans and mice, while the number of MatchVars between humans and *C. elegans* is 42614 (Figure 2B, and Supplementary Tables S1 and S2). Finally, to make the MatchVars available to the community, we created a web-based search engine and database called ConVarT (http://www.convart.org/) to facilitate consistent and rapid visualization of MatchVars and variant annotations on MSAs together with human PTMs and protein domains. ConVarT offers a protein sequence similarity index based on the matching residues for three species below part of MSAs so that the researchers utilize this section to gain more insight into the conservation of the sequence in the model organism compared to humans.

ConVarT reduces the time needed to find the phenotypic and unknown significance of MatchVars in the genomes of three organisms from hours to seconds. A comprehensive list of identifiers that can be used to search for genes on ConVarT can be found in Supplementary Table S17. Furthermore, users can now submit any protein sequence as an input, and ConVarT will perform sequence comparison and alignment to find the closest human orthologue of that protein sequence, as well as display human variants on the pairwise sequence alignment of human protein and submitted protein sequence (Supplementary Figure S2). This is especially useful for model organism researchers, such as *S. cerevisiae,* who have yet to be incorporated into ConVarT.

Usage of ConVarT to conjecture about the functional effects of variants in disease analysis

Disease-causing variants are annotated as pathogenic in humans, whereas variants with phenotypic annotations in non-human species indicate disease-like conditions for them, suggesting that both types of annotations represent the disruption of the protein function. The majority of human genetic variants in the ClinVar database are single nucleotide variants (SNVs), and variant of uncertain significance (VUS) make up the majority of these SNVs (Supplementary Figure S1A and B). Expectedly, compared to the fraction of phenotypic MatchVars coinciding with benign variants of ClinVar, the proportion of phenotypic MatchVars of *C. elegans* overlapping with pathogenic variants of ClinVar is calculated to be significantly higher ($P < 0.0001$; chi-square χ^2 test), which is consistent with the equality of pathogenic with phenotypic variants. For phenotypic MatchVars from mice, the mouse phenotypic MatchVars/ClinVar pathogenic variants ratio is slightly higher than the ratio of the mouse phenotypic MatchVars/ClinVar benign variants ($P < 0.05$; χ^2 test) (Figure 2C and D). ConVarT can serve researchers and human geneticists in several ways. First, they can assess whether variants of interest from humans currently have MatchVars already associated with the phenotype in the mice or *C. elegans,* potentially adding a new layer of support to their study. For example, the MatchVar p.G368E

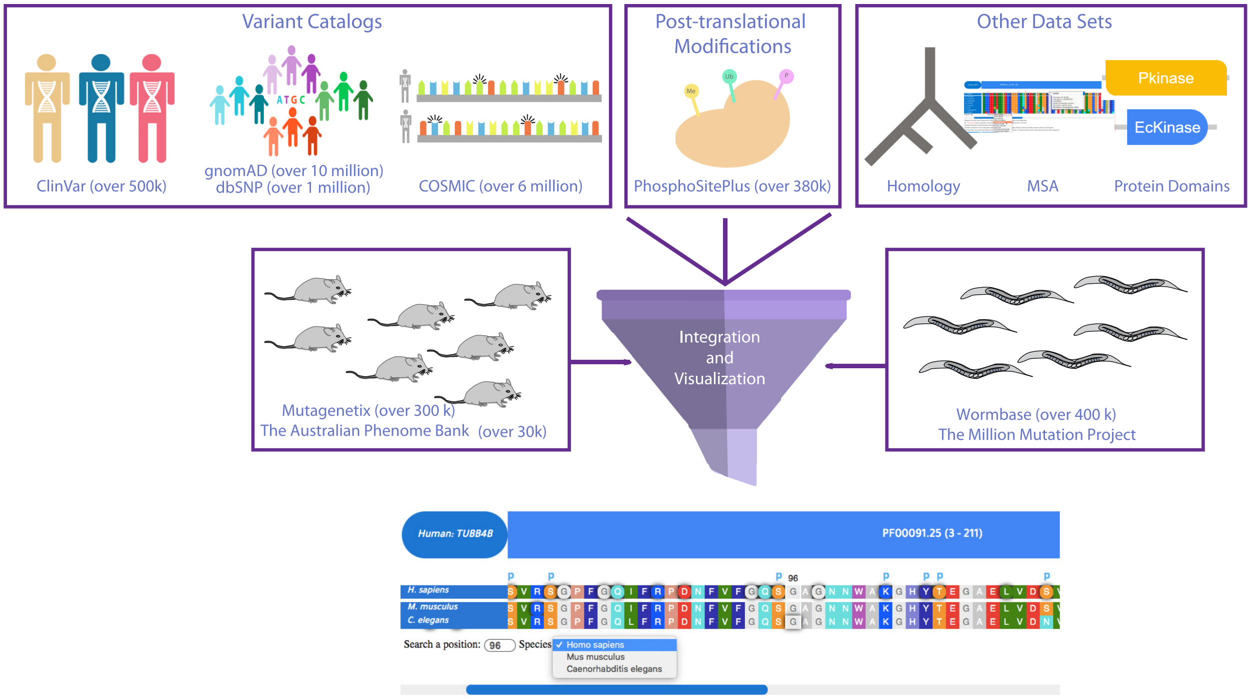

Figure 1. The schematic flow of ConVarT, a search engine for matching variants—gene orthology was determined, followed by MSAs and the incorporation of human variants from gnomAD, ClinVar, COSMIC and dbSNP, and variants from the Australian Phenome Bank (Mice), Mutagenetix (Mice) and WormBase (*C. elegans*). Human PTMs and human protein domains are integrated for co-visualization with the MSA and variants.

(allele name n4435 and NP_495455) in *C. elegans* KAT-1 (Human ACAT1) was identified as phenotypic (short life span) in 2010. The p.G388E variant (NP 000010) in human ACAT1 was pathogenic seven years later, resulting in human acetyl-CoA acetyltransferase deficiency (25,26). We presented all the phenotypic MatchVars of mice and *C. elegans*, overlapping with pathogenic and other variants in ClinVar, COSMIC, dbSNP, and gnomAD (Figure 2E–J, and Supplementary Tables S1–S16). Second, there may currently be no MatchVars from humans for the phenotypic MatchVars of the mice or *C. elegans*. Where there may be the appearance of a new human variant associated with a disease, scientists and human geneticists may look for the phenotypic MatchVars in ConVarT, thereby helping to assign certain variants as being potentially detrimental. Finally, the Million Mutation Project and Mutagenetix have many strain collections that bear unique MatchVar mutations. Many of these mutations are poorly characterized, such as *C. elegans* and mice mutants bearing MatchVars (*C. elegans* DYF-18/CDK7 NP_502232: p.P245L and p.R26K; *C. elegans* OSM-3/KIF17 NP_001023308: p.T89I, p.P161L, p.S208L, p.A464T, and p.W572*; Mice Ros1, NP_025412: p.T220A and p.V2050A; Mice Tg, NP_033401: p.S164G, p.T1555S, and p.C1993R) (Figure 3A–D) (17,27). Therefore, the functional significance of these MatchVars is currently unknown. These mutants can be obtained from the distribution centers for functional analysis (17,27). Researchers can also easily submit their variants and phenotype data to ConVarT, thereby sharing their phenotypic and variant data with the community.

Utilizing phenotypic variants from non-human species to add one layer support for understanding human variants

Eight missense variants in *C. elegans ben-1* (Human *TUBB4B*) encoding tubulin beta chain (NP_497728: p.E69G, p.Q131L, p.S145F, p.F167Y, p.A185P, p.F200Y, p.M257I and p.D404N) were implicated in displaying resistance to benzimidazole (BZ) and were labeled as putative loss-of-function (LoF) variants (28,29) (Figure 3E). We found that *C. elegans* missense residues were conserved in humans and mice, except for p.F200. Next, we investigated whether these conserved residues had MatchVars in humans, which revealed no MatchVars for p.E69G, p.Q131L, p.S145F, p.A185P and p.D404N in humans. This result suggested that these variants were under negative selection in the human population, potentially due to their detrimental effects. All these variants were consistently predicted to be deleterious by SIFT (Supplementary Table S3). However, p.M257I and p.F167Y in *C. elegans* ben-1 (NP_497728) are the MatchVars of p.M257I and p.F167Y (variant IDs: COSM3929966 and COSM6451925) in human TUBB4B (NP_006079.1), respectively, predicted to be pathogenic by the FATHMM (Figure 3E) (19). Indeed, the absence of the p.M257I and p.F167Y mutations in human TUBB4B in gnomAD and TOPMed suggests that they were likely negatively selected. This data indicates that the experimental data already available from *C. elegans* and mice may provide additional evidence to deduce the functional implication of human variants together with allele frequency.

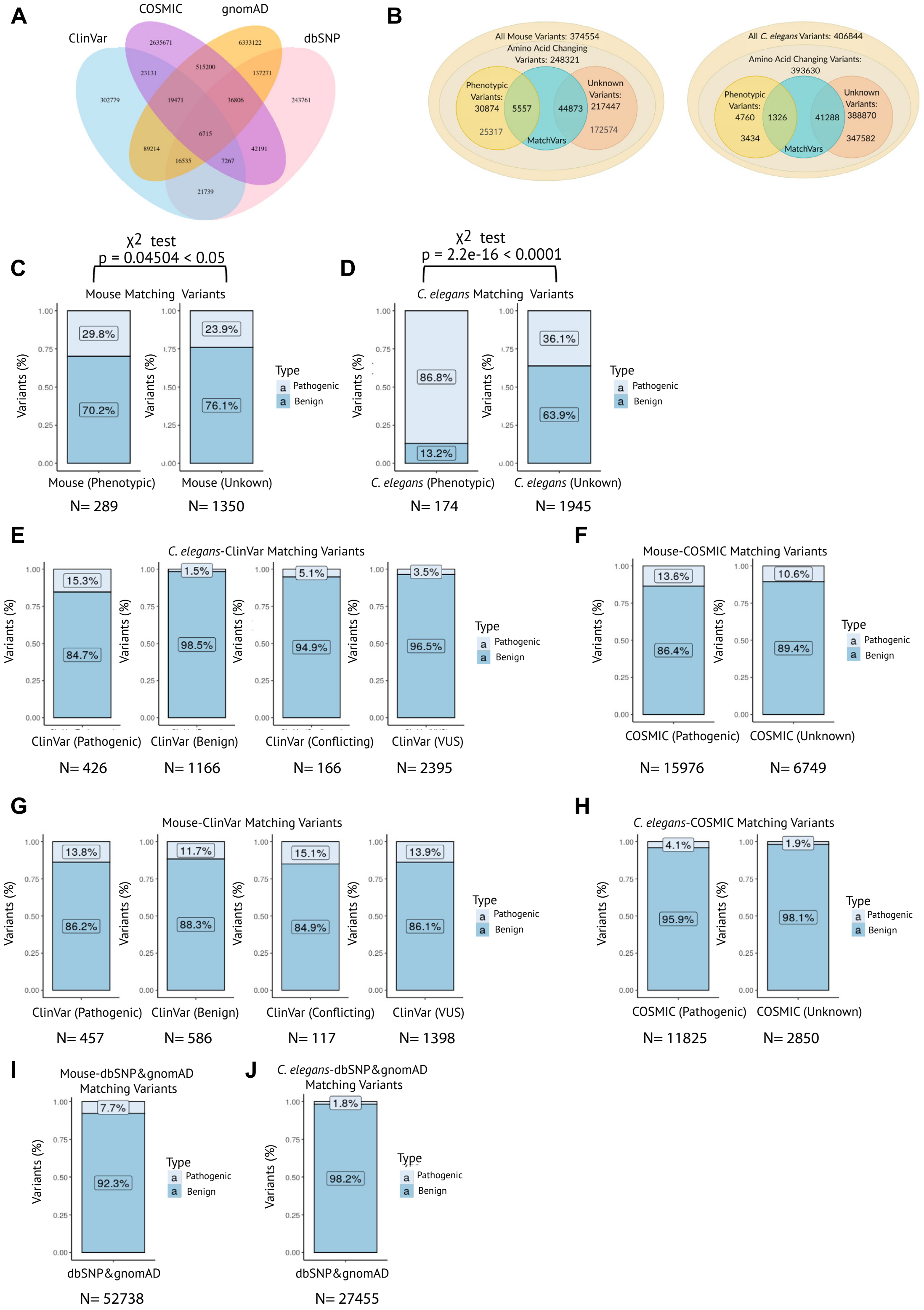

Figure 2. Matching variants among human, mice and *C. elegans*. (**A**) Venn diagram showing the number of overlapping and non-overlapping variants between gnomAD, ClinVar, COSMIC and dbSNP (**B**) the total number of MatchVars with phenotypes and unknown significance from mice and *C. elegans* were displayed in the two distinct Venn diagrams. (**C, D**), Bar plots display the fraction of phenotypic or unknown significance of MatchVars from mouse and *C. elegans* with either ClinVar pathogenic or benign variants. A statistically significant difference between two proportions was displayed for mouse and *C. elegans.* χ^2 = chi Square (**E–J**) Bar plots display comparisons of phenotypic and unknown MatchVars from mice and *C. elegans* (percentage) with a range of human variants databases ClinVar, COSMIC, gnomAD and dbSNP. N: number of MatchVars between humans, mice or *C. elegans*.

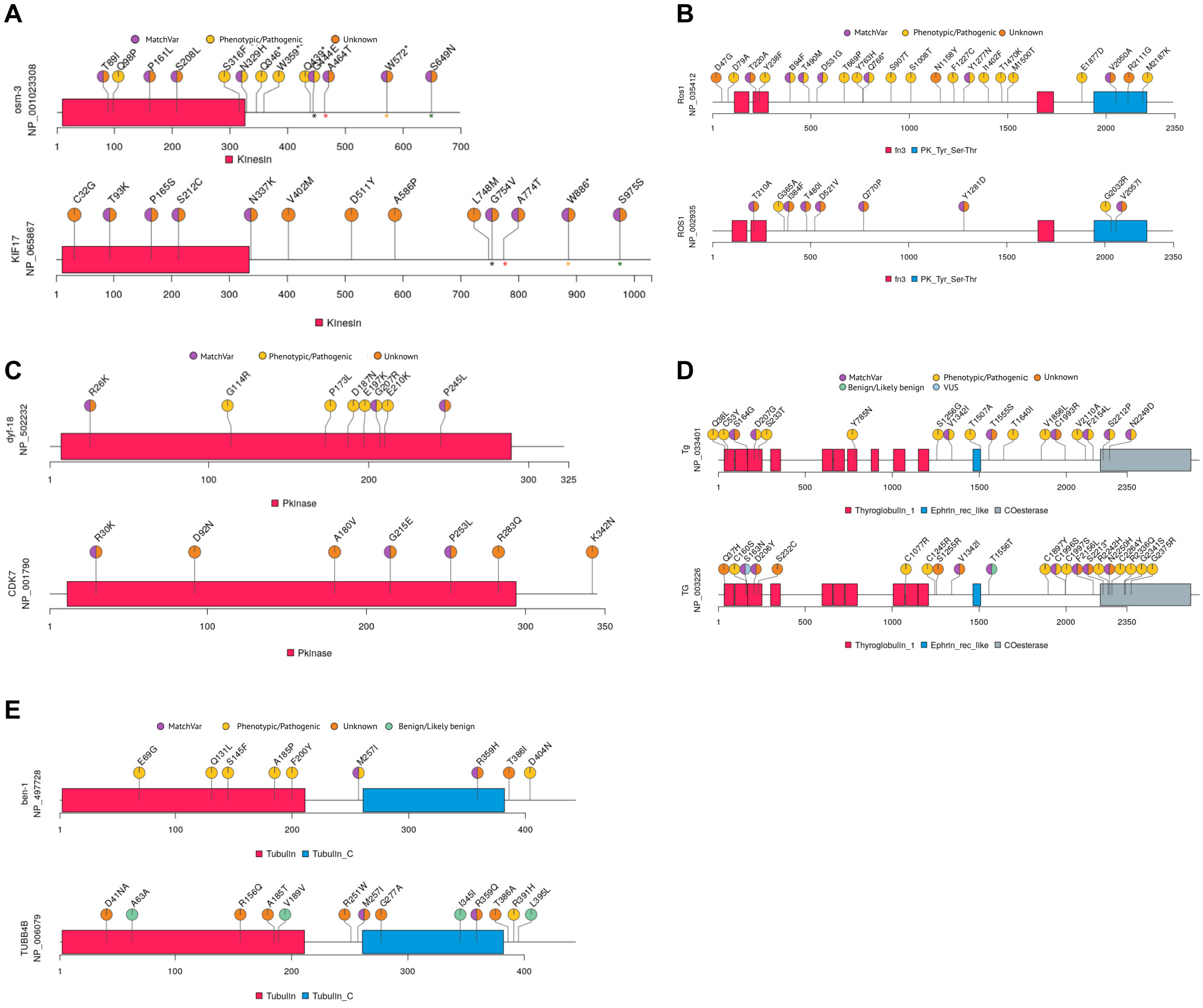

Figure 3. Distribution of matching and pathological/phenotypic variants per gene – Graphical representations of data (lollipop plots) exhibits pathological/phenotypic, orthologous, variants of unknown significance, and benign or likely benign variants (annotated amino acid changes) at their corresponding positions per gene. The length of each protein is shown, together with specific functional protein domains (different colors). Protein IDs and gene names are displayed next to the plots. (**A, C** and **E**) Comparison of *C. elegans* genes with human counterparts. In (A), different colors for the asterisk reflect MatchVars among humans and *C. elegans*. (**B** and **D**) Mice and human genes are displayed.

DISCUSSION

Model organisms have contributed significantly to our knowledge of human biology and diseases. For this reason, the Alliance of Genome Resources recently launched an integrated database that shares genetic and functional genomic data from a variety of model organisms in order to aid scientists in their understanding of human biology and disorders, though they did not provide a search engine for MatchVars (30). This current work fills this gap by presenting a search engine for MatchVars and a database comparably displaying the pathogenic or phenotypic and unknown significance of variants from humans, mice, and *C. elegans.* ConVarT was designed with the goal of expanding and incorporating variants from other organisms, such as Zebrafish and Drosophila. ConVarT recently added 24 658 526 human variants from TOPMed to its database (Supplementary Table S18) (3). The model organism community can easily submit new variant data and expand ConVarT to become an open platform for exchanging information about new MatchVars, thereby aiding the presentation of empirical evidence from model organisms to provide one more layer of support for interpretation of human variants.

MATERIAL AND METHODS

The gene homology list for humans and non-human species

The gene homology list for human and non-human organisms was first build using organism-specific orthology resources such as DIOPT, MGI, and the Zebrafish Information Network (ZFIN) (20,31–32). In addition, as we uncovered inaccuracies and missing data in the orthology data, we manually corrected them using a reciprocal BLAST (33) analysis to find the counterparts of human genes. For example, the reciprocal BLAST analysis reveals that

C. elegans F42G8.19 (ceph-41) is the most likely ortholog of human CEP41, *C. elegans* F42G8.19 does not emerge as an ortholog of human CEP41 based on orthology searches from *C. elegans* gene orthology database (34,35).

Download of genetic variants for humans, mice and *C. elegans* and posttranslational modification (PTMs)

The human variants and organisms specific variants were downloaded from the following resources: ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/ (ClinVar), http://gnomad.broadinstitute.org/ (Aggregation Database (gnomAD v2.03), https://cancer.sanger.ac.uk/download (COSMIC), https://ftp.ncbi.nih.gov/snp/ (dbSNP) and https://www.phosphosite.org/Supplemental_Files (PhosphoSitePlus) (2,16,19–20,24). Mutagenetix (https://mutagenetix.utsouthwestern.edu/about.cfm) and Wormbase generously shared variants and phenotypic data for Mouse and *C. elegans,* respectively (17,18). We obtained mice phenotypic data and variants from the Australian Phenomics website (the webpage for download is https://pb.apf.edu.au/phenbank/homePage.html).

Multiple sequence alignments

We retrieved human protein sequences and human homolog genes from chimp, Macaca, rat, mouse, zebrafish, *C. elegans, D. melanogaster,* and frog, and used ClustalW v2.1 pairwise alignment to perform 2,379,397 distinct combinations of multiple sequence alignments (MSAs) of these protein sequences (36). The final files were stored in MySQL together with variants from humans and non-human species.

Data visualization and website development

The recently developed MSA visualization tool was used to co-visualize MSAs along with annotations (variants and phenotypic data) (22). We constructed the ConVarT website and deposited all data for the users. Furthermore, we shared all of the necessary code (ConVarT Web https://github.com/thekaplanlab/ConVarT_Web) on GitHub.

Additional materials and methods are available in the supplementary file.

DATA AVAILABILITY

All codes used for generating data, websites, and figures are accessible from https://github.com/thekaplanlab.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Seref Gul for critical reading of manuscripts. We gratefully thank Mutagenetix and Wormbase for generously sharing variants from Mutagenetix and Wormbase.

Author contributions: O.I.K. and S.C. originated project, H.I.B., F.M.T. and M.S.P. generated MSA data and collected variants, F.M.T. generated the homology table, H.B., F.T., A.S. and F.C. designed and developed web application, M.S.P. analyzed the data and generated Figures 2 and 3 and supplemental tables and figures, P.Z and Y. K. provided reagents, O.I.K. directed the work of H.B., F.M.T., M.S.P., A.S. and F.C. and wrote the original draft

FUNDING

Kaplan Laboratory is supported by Abdullah Gul University Scientific Research Project [TOA-2018-110].

Conflict of interest statement. None declared.

REFERENCES

1. Lek,M., Karczewski,K.J., Minikel,E.V., Samocha,K.E., Banks,E., Fennell,T., O'Donnell-Luria,A.H., Ware,J.S., Hill,A.J., Cummings,B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.
2. Karczewski,K.J., Francioli,L.C., Tiao,G., Cummings,B.B., Alföldi,J., Wang,Q., Collins,R.L., Laricchia,K.M., Ganna,A., Birnbaum,D.P. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.
3. Taliun,D., Harris,D.N., Kessler,M.D., Carlson,J., Szpiech,Z.A., Torres,R., Taliun,S.A.G., Corvelo,A., Gogarten,S.M., Kang,H.M. *et al.* (2021) Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, **590**, 290–299.
4. Shendure,J. and Akey,J.M. (2015) The origins, determinants, and consequences of human mutations. *Science*, **349**, 1478–1483.
5. Hieter,P. and Boycott,K.M. (2014) Understanding rare disease pathogenesis: a grand challenge for model organisms. *Genetics*, **198**, 443–445.
6. Yamamoto,S., Jaiswal,M., Charng,W.-L., Gambin,T., Karaca,E., Mirzaa,G., Wiszniewski,W., Sandoval,H., Haelterman,N.A., Xiong,B. *et al.* (2014) A Drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell*, **159**, 200–214.
7. McGurk,L., Berson,A. and Bonini,N.M. (2015) *Drosophila* as an *in vivo* model for human neurodegenerative disease. *Genetics*, **201**, 377–402.
8. Chao,H.-T., Liu,L. and Bellen,H.J. (2017) Building dialogues between clinical and biomedical research through cross-species collaborations. *Semin. Cell Dev. Biol.*, **70**, 49–57.
9. Wang,J., Al-Ouran,R., Hu,Y., Kim,S.-Y., Wan,Y.-W., Wangler,M.F., Yamamoto,S., Chao,H.-T., Comjean,A., Mohr,S.E. *et al.* (2017) MARRVEL: integration of human and model organism genetic resources to facilitate functional annotation of the human genome. *Am. J. Hum. Genet.*, **100**, 843–853.
10. Vaser,R., Adusumalli,S., Leng,S.N., Sikic,M. and Ng,P.C. (2016) SIFT missense predictions for genomes. *Nat. Protoc.*, **11**, 1–9.
11. Kasprzyk,A. (2011) BioMart: driving a paradigm change in biological data management. *Database*, **2011**, bar049.
12. Powell,S., Forslund,K., Szklarczyk,D., Trachana,K., Roth,A., Huerta-Cepas,J., Gabaldón,T., Rattei,T., Creevey,C., Kuhn,M. *et al.* (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res.*, **42**, D231–D239.
13. Altenhoff,A.M., Škunca,N., Glover,N., Train,C.-M., Sueki,A., Piližota,I., Gori,K., Tomiczek,B., Müller,S., Redestig,H. *et al.* (2015) The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Res.*, **43**, D240–D249.
14. Linard,B., Allot,A., Schneider,R., Morel,C., Ripp,R., Bigler,M., Thompson,J.D., Poch,O. and Lecompte,O. (2015) OrthoInspector 2.0: software and database updates. *Bioinformatics*, **31**, 447–448.
15. Sonnhammer,E.L.L. and Östlund,G. (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic. Acids. Res.*, **43**, D234–D239.
16. Sherry,S.T., Ward,M. and Sirotkin,K. (1999) dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.*, **9**, 677–679.
17. Wang,T., Zhan,X., Bu,C.-H., Lyon,S., Pratt,D., Hildebrand,S., Choi,J.H., Zhang,Z., Zeng,M., Wang,K. *et al.* (2015) Real-time

resolution of point mutations that cause phenovariance in mice. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E440–E449.

18. Harris,T.W., Arnaboldi,V., Cain,S., Chan,J., Chen,W.J., Cho,J., Davis,P., Gao,S., Grove,C.A., Kishore,R. *et al.* (2019) WormBase: a modern Model Organism Information Resource. *Nucleic Acids Res.*, **48**, D762–D767.
19. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E. *et al.* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.*, **47**, D941–D947.
20. Landrum,M.J., Chitipiralla,S., Brown,G.R., Chen,C., Gu,B., Hart,J., Hoffman,D., Jang,W., Kaur,K., Liu,C. *et al.* (2020) ClinVar: improvements to accessing data. *Nucleic Acids Res.*, **48**, D835–D844.
21. Bult,C.J., Blake,J.A., Smith,C.L., Kadin,J.A. and Richardson,J.E., (2019) the Mouse Genome Database Group. 2019. Mouse Genome Database (MGD). *Nucleic Acids Res.*, **47**, D801–D806
22. Torun,F.M., Bilgin,H.I. and Kaplan,O.I. (2021), MSABrowser: dynamic and fast visualization of sequence alignments, variations, and annotations, *Bioinformatics Adv.*, **1**, vbab009.
23. Bateman,A. (2004) The Pfam protein families database. *Nucleic. Acids. Res.*, **32**, 138–141.
24. Hornbeck,P.V., Zhang,B., Murray,B., Kornhauser,J.M., Latham,V. and Skrzypek,E. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*, **43**, D512–D520.
25. Berdichevsky,A., Nedelcu,S., Boulias,K., Bishop,N.A., Guarente,L. and Horvitz,H.R. (2010) 3-Ketoacyl thiolase delays aging of *Caenorhabditis elegans* and is required for lifespan extension mediated by sir-2.1. *Proc. Natl. Acad. Sci.*, **107**, 18927–18932.
26. Paquay,S., Bourillon,A., Pichard,S., Benoist,J.-F., de Lonlay,P., Dobbelaere,D., Fouilhoux,A., Guffon,N., Rouvet,I., Labarthe,F. *et al.* (2017) Mitochondrial acetoacetyl-CoA thiolase deficiency: basal ganglia impairment may occur independently of ketoacidosis. *J. Inherit. Metab. Dis.*, **40**, 415–422.
27. Thompson,O., Edgley,M., Strasbourger,P., Flibotte,S., Ewing,B., Adair,R., Au,V., Chaudhry,I., Fernando,L., Hutter,H. *et al.* (2013) The million mutation project: A new approach to genetics in Caenorhabditis elegans. *Genome Res.*, **23**, 1749–1762.
28. Hahnel,S.R., Zdraljevic,S., Rodriguez,B.C., Zhao,Y., McGrath,P.T. and Andersen,E.C. (2018) Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles. *PLoS Pathog.*, **14**, e1007226.
29. Kitchen,S., Ratnappan,R., Han,S., Leasure,C., Grill,E., Iqbal,Z., Granger,O., O'Halloran,D.M. and Hawdon,J.M. (2019) Isolation and characterization of a naturally occurring multidrug-resistant strain of the canine hookworm, *Ancylostoma caninum*. *Int. J. Parasitol.*, **49**, 397–406.
30. The Alliance of Genome Resources Consortium, Agapite,J., Albou,L.-P., Aleksander,S., Argasinska,J., Arnaboldi,V., Attrill,H., Bello,S.M., Blake,J.A., Blodgett,O. *et al.* (2020) Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Res.*, **48**, D650–D658.
31. Ruzicka,L., Howe,D.G., Ramachandran,S., Toro,S., Van Slyke,C.E., Bradford,Y.M., Eagle,A., Fashena,D., Frazer,K., Kalita,P. *et al.* (2019) The Zebrafish Information Network: new support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic. Acids. Res.*, **47**, D867–D873.
32. Hu,Y., Flockhart,I., Vinayagam,A., Bergwitz,C., Berger,B., Perrimon,N. and Mohr,S.E. (2011) An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics*, **12**, 357.
33. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
34. Kim,W., Underwood,R.S., Greenwald,I. and Shaye,D.D. (2018) OrthoList 2: a new comparative genomic analysis of human and *Caenorhabditis elegans* genes. *Genetics*, **210**, 445–461.
35. Cevik,S. and Kaplan,OI. (2021) The Joubert syndrome protein CEP41 is excluded from the distal segment of cilia in *C. elegans*. *MicroPubl. Biol.*, **2021**, https://doi.org/10.17912/micropub.biology.000406.
36. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.