ConVarT: a search tool for orthologous variants:

A method and server for functional inference of human genetic variants

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], Oktay I. Kaplan[1]†

1- Rare Disease Laboratory, School of Life and Natural Sciences, Abdullah Gul University, Kayseri, Turkey;

2- Department of Computer Engineering, Abdullah Gul University, Kayseri, Turkey;

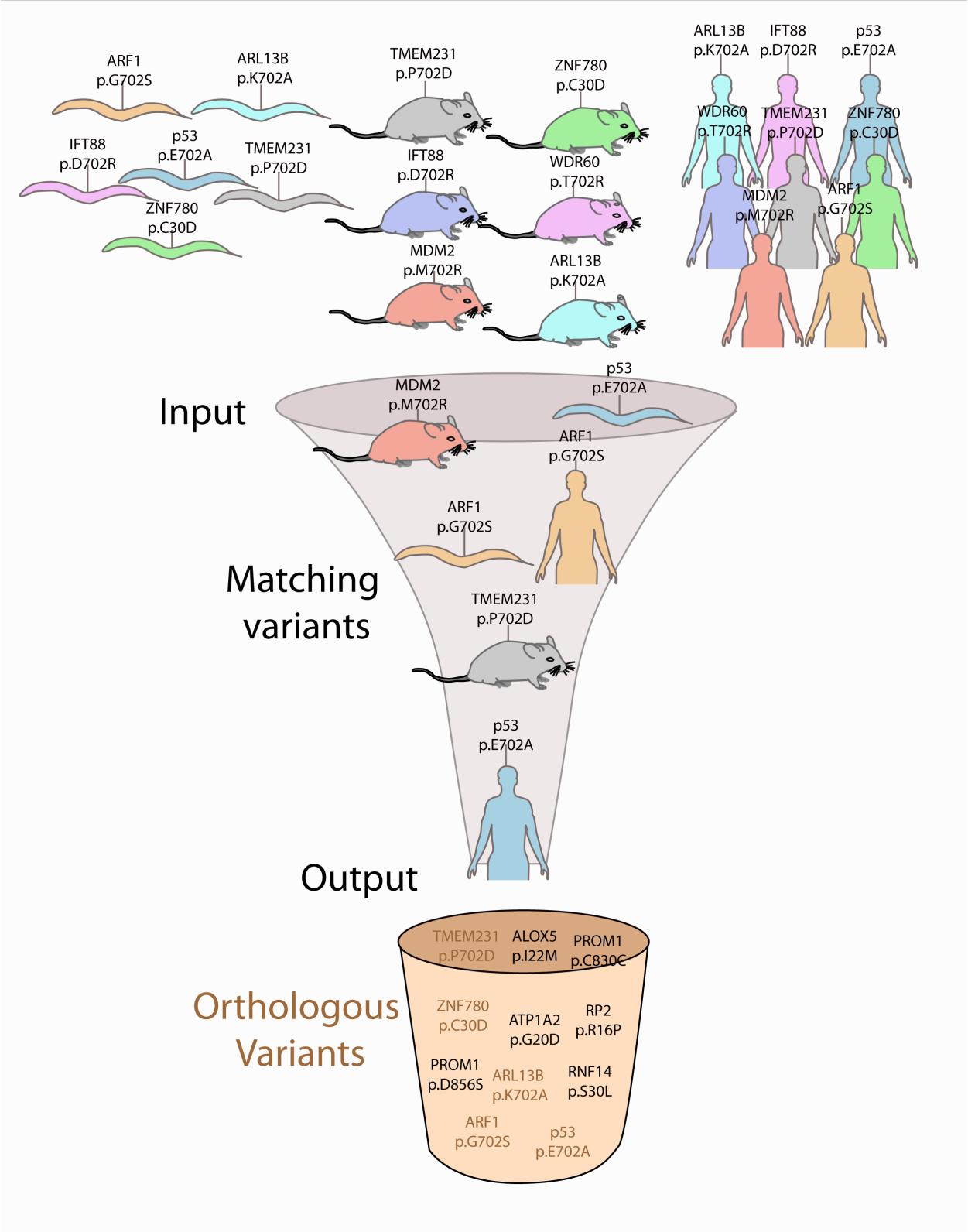
3- SunyBiotech Co., Ltd, Fuzhou 35000, China;

† Correspondence to: oktay.kaplan@agu.edu.tr

Abstract

The availability of genetic variations, together with phenotypic annotations from model organisms, facilitates comparing these variants with equivalent variants. However, existing databases and search tools do not make it easy to scan for equivalent variants, namely "orthologous variants," between humans and other organisms. Therefore, we developed an integrated search engine called ConVarT (http://www.convart.org/) for orthologous variants between humans, mice, and *C. elegans*. ConVarT incorporates annotations (including phenotypic and pathogenic) into variants, and these previously unexploited phenotypic OrthoVars 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.

Graphical Abstract



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 more than 140,000 randomly selected individuals, providing large-scale collections of human genetic variants (Lek et al. 2016; Karczewski et al. 2020). Humans possess over 400 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 (Taliun et al. 2019). 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 (Shendure and Akey 2015). 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.

New Approach

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 (Hieter and Boycott 2014; Yamamoto et al. 2014; McGurk et al. 2015; Chao et al. 2017; Wang et al. 2017). 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 a conserved amino acid 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 (Vaser et al. 2016). 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 (Kasprzyk 2011; Powell et al. 2014; Altenhoff et al. 2015; Linard et al. 2015; Sonnhammer and Östlund 2015).

Here, we develop an integrated search tool and database called ConVarT (**Con**served clinical **Var**iation Visualization **T**ool) to search for equivalent variants (pathological, phenotypic, among others) 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 orthologous variants (OrthoVar) between human and mice or *C. elegans*. Furthermore, we integrated available annotations (pathological, phenotypic, among others) for variants into the database, typically making it easy for users to search for OrthoVars with phenotypic data. Our comprehensive database may help deduce the biological relevance of human variants from OrthoVars associated with functional data.

ConVarT: a search tool for "orthologous 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)** (Sherry et al. 1999; Wang et al. 2015; Harris et al. 2019; Tate et al. 2019; Karczewski et al. 2020; Landrum et al. 2020). We initially derived the human gene

74 orthologs for *M. musculus* and *C. elegans*. We then collected amino acid-changing variants from
75 gnomAD, ClinVar, COSMIC, and dbSNP for humans, from WormBase for *C. elegans*, and from
76 the Australian Phenome Bank and Mutagenetix for mice. gnomAD describes over 16,179,380
77 variations, while the ClinVar database deposits over 500,000 variants, including pathogenic and
78 benign variants (Karczewski et al. 2020; Landrum et al. 2020). COSMIC and dbSNP contain over
79 6,842,627 variations and 1,086,546 variations, respectively (Sherry et al. 1999; Tate et al. 2019).
80 Our analysis revealed that gnomAD, ClinVar, COSMIC, and dbSNP have identical and non-
81 identical variants (**Figure 2A)**. Mutagenetix and WormBase present almost 800,000 variants from
82 these organisms (Wang et al. 2015; Harris et al. 2019). To compare human variants with those
83 from mice and *C. elegans*, we matched the protein sequences of the ortholog genes, followed by
84 extensive multiple sequence alignments (MSAs) of these protein sequences to ensure alignment of
85 the corresponding amino acid positions. The schematic workflow of ConVarT is presented in
86 **Figure 1**. We then integrated genetic variants and variant-specific annotations (such as
87 pathogenicity, phenotypes, and allele frequency) from humans, mice, and *C. elegans* into
88 corresponding amino acid positions in each gene. We restricted our comparisons to amino acid-
89 changing variants, as amino acids are more conserved than non-amino acid-changing variants, and
90 protein function is likely to be affected by an alteration in the corresponding amino acids. For
91 visualization of MSAs together with variants from humans, mice, and *C. elegans*, we used in-
92 house software (Torun F&Bilgin H et al. Manuscript in preparation). Previous studies revealed
93 that many human variants overlap with amino acids undergoing post-translational modifications
94 (PTMs) and are enriched at the protein domains. Therefore, Pfam protein domains for humans and
95 383k PTMs from PhosphoSitePlus were systematically integrated into corresponding positions
96 (Bateman 2004; Hornbeck et al. 2015).
97

98 Our analysis revealed that many equivalent amino acid substitutions exist between humans and
99 these two organisms (**Figure 2B**). For example, both human AP1M1 and *C. elegans* UNC-101,
100 the orthologue of AP1M1, have the identical residue change at the corresponding positions
101 (Human NP_001123996.1: p.P375S, Variant ID: rs1490425955 vs *C. elegans* NP_001040675:
102 p.P362S, Variant ID: WBVar00679460). We have named these equivalent variations "orthologous
103 variants" (OrthoVars) (**Figure 2B, and Table S1, S2**). For a variant to be considered an OrthoVar,
104 the amino acid residue at the corresponding positions must be conserved between the human gene
105 and the ortholog gene. The Materials and Methods provides a detailed explanation. There are
106 50430 OrthoVars between humans and mice, while the number of OrthoVars between humans and
107 *C. elegans* is 42614 (**Figure 2B, and Table S1, S2**). Finally, to make the OrthoVars available to
108 the community, we created a web-based search engine and database called ConVarT
109 (http://www.convart.org/) to facilitate consistent and rapid visualization of OrthoVars and variant
110 annotations on MSAs together with human PTMs and protein domains. ConVarT reduces the time
111 needed to find the phenotypic and unknown significance of OrthoVars in the genomes of three
112 organisms from hours to seconds.
113

114 **Usage of ConVarT to predict the functional effects of variants in disease analysis**

115 Disease-causing variants are annotated as pathogenic in humans, whereas variants with phenotypic
116 annotations in non-human species indicate disease-like conditions for them. Both types of
117 annotations represent the disruption of the protein function, suggesting that the effect of mutations
118 on protein in human and non-human species is similar. Expectedly, compared to the fraction of
119 phenotypic OrthoVars coinciding with benign variants of ClinVar, the proportion of phenotypic
120 OrthoVars of *C. elegans* overlapping with pathogenic variants of ClinVar is calculated to be

121 significantly higher ($p < 0.0001$; Chi-square χ^2 test), which is consistent with the equality of
122 pathogenic with phenotypic variants. For phenotypic OrthoVars from mice, the mouse phenotypic
123 OrthoVars/ClinVar pathogenic variants ratio is slightly higher than the ratio of the mouse
124 phenotypic OrthoVars/ClinVar benign variants ($p < 0.05$; χ^2 test) (**Figure 2C and 2D**). ConVarT
125 can serve researchers and human geneticists in several ways. First, they can assess whether variants
126 of interest from humans currently have OrthoVars already associated with the phenotype in the
127 mice or *C. elegans,* potentially adding a new layer of support to their study. For example, the
128 OrthoVar p.G368E (allele name n4435 and NP_495455) in *C. elegans* KAT-1 (Human ACAT1)
129 was identified as phenotypic (short life span) in 2010. The p.G388E variant (NP 000010) in human
130 ACAT1 was pathogenic seven years later, resulting in human acetyl-CoA acetyltransferase
131 deficiency (Berdichevsky et al. 2010; Paquay et al. 2017). We presented all the phenotypic OrthoVars
132 of mice and *C. elegans*, overlapping with pathogenic and other variants in ClinVar, COSMIC,
133 dbSNP, and gnomAD (**Figure 2E, F, G, H, I and J, and Table S1- S17**). Second, there may
134 currently be no OrthoVars from humans for the phenotypic OrthoVars of the mice or *C. elegans*.
135 Where there may be the appearance of a new human variant associated with a disease, scientists
136 and human geneticists may look for the phenotypic OrthoVars in ConVarT, thereby helping to
137 assign certain variants as being potentially detrimental. Finally, the Million Mutation Project and
138 Mutagenetix have many strain collections that bear unique OrhoVar mutations. Many of these
139 mutations are poorly characterized, such as *C. elegans* and mice mutants bearing OrthoVars (*C.*
140 *elegans* DYF-18/CDK7 NP_502232: p.P245L and p.R26K; *C. elegans* OSM-3/KIF17
141 NP_001023308: p.T89I, p.P161L, p.S208L, p.A464T, and p.W572*; Mice Ros1, NP_025412:
142 p.T220A and p.V2050A; Mice Tg, NP_033401: p.S164G, p.T1555S, and p.C1993R) (**Figure 3A,**

143 **B, C, and D**) (Thompson et al. 2013; Wang et al. 2015). Therefore, the functional significance of
144 these OrthoVars is currently unknown. These mutants can be obtained from the distribution centers
145 for functional analysis (Thompson et al. 2013; Wang et al. 2015). Researchers can also easily
146 submit their variants and phenotype data to ConVarT, thereby sharing their phenotypic and variant
147 data with the community.
148

149 **Functional inference from phenotypic variants currently with no OrthoVars in humans**

150 Eight missense variants in *C. elegans ben-1* (Human *TUBB4B*) encoding tubulin beta chain
151 (NP_497728: p.E69G, p.Q131L, p.S145F, p.S145F, p.A185P, p.F200Y, p.M257I, and p.D404N)
152 were implicated in displaying resistance to benzimidazole (BZ) and were labeled as putative loss-
153 of-function (LoF) variants (Hahnel et al. 2018; Kitchen et al. 2019) (**Figure 3E**). We found that
154 *C. elegans* missense residues were conserved in humans and mice, except for p.F200. Next, we
155 investigated whether these conserved residues had OrthoVars in humans, which revealed no
156 OrthoVars for p.E69G, p.Q131L, p.S145F, p.A185P, and p.D404N in humans. This result
157 suggested that these variants were under negative selection in the human population, potentially
158 due to their detrimental effects. All these variants were consistently predicted to be deleterious by
159 SIFT (**Table S3)**. However, p.M257I and p.F167Y in C. elegans ben-1 (NP_497728) are the
160 OrthoVars of p.M257I and p.F167Y (variant IDs: COSM3929966 and COSM6451925) in human
161 TUBB4B (NP_006079.1), respectively, predicted to be pathogenic by the FATHMM (**Figure 3E**)
162 (Tate et al. 2019). Indeed, Indeed, the absence of the p.M257I and p.F167Y mutations in human
163 TUBB4B in gnomAD and TopMed suggests that they were likely negatively selected. This data
164 indicates that the experimental data already available from *C. elegans* and mice may provide
165 additional evidence to deduce the functional implication of human variants together with allele
166 frequency.

Conclusion

ConVarT is a search tool for OrthoVars and a database comparably displaying the pathogenic or phenotypic and unknown significance of variants from humans, mice, and *C. elegans.* ConVarT was designed to expand and include variants from other organisms, such as Zebrafish and Drosophila. The model organism community can easily submit new variant data and expand ConVarT to become an open platform for exchanging information about new OrhoVars, thereby aiding the presentation of empirical evidence from model organisms for the functional inference of human variants.

Material and Methods

Materials and methods are available in the supplementary file.

Code availability

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

190 **Figure 1: The schematic flow of ConVarT, a search tool for orthologous variants –** Gene
191 orthology was determined, followed by MSAs and the incorporation of human variants from
192 gnomAD, ClinVar, COSMIC and dbSNP, and variants from the Australian Phenome Bank (Mice),
193 Mutagenetix (Mice) and WormBase (*C. elegans*). Human PTMs and human protein domains are
194 integrated for co-visualization with the MSA and variants.
195
196 **Figure 2: Orthologous variants among human, mice, and *C. elegans* – A)** Venn diagram
197 showing the number of overlapping and non-overlapping variants between gnomAD, ClinVar,
198 COSMIC, and dbSNP **B)** the total number of OrthoVars with phenotypes and unknown
199 significance from mice and *C. elegans* were displayed in the two distinct Venn diagrams. **C), D),**
200 Bar plots display the fraction of phenotypic or unknown significance of OrthoVars from mouse
201 and *C. elegans* with either ClinVar pathogenic or benign variants. A statistically significant
202 difference between two proportions was displayed for mouse and *C. elegans.* $\chi 2$= Chi Square **E),**
203 **F) G), H, I and J)** Bar plots display comparisons of phenotypic and unknown OrthoVars from
204 mice and *C. elegans* (percentage) with a range of human variants databases ClinVar, COSMIC,
205 gnomAD, and dbSNP. N: number of OrthoVars between human, mice or *C. elegans*.
206
207 **Figure 3: Distribution of orthologous and pathological/phenotypic variants per gene –**
208 Graphical representations of data (lollipop plots) exhibits pathological/phenotypic, orthologous,
209 variants of unknown significance, and benign or likely benign variants (annotated amino acid
210 changes) at their corresponding positions per gene. The length of each protein is shown, together
211 with specific functional protein domains (different colors). Protein IDs and gene names are
212 displayed next to the plots. **A), C), and E)** Comparison of *C. elegans* genes with human

counterparts. In Figure 3A, different colors for the asterisk reflect OrthoVars among humans and *C. elegans*. **B) and D)** Mice and human genes are displayed.

Acknowledgment

We thank Dr. Seref Gul for critical reading of manuscripts. We gratefully thank Mutagenetix and Wormbase for generously sharing variants from Mutagenetix and Wormbase. Work in the Kaplan laboratory is supported by Abdullah Gul University Scientific Research Project (TOA-2018-110).

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 Figure 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

Footnotes

The authors declare no conflict of interest.

References

Altenhoff AM, 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.

Bateman A. 2004. The Pfam protein families database. *Nucleic Acids Res.* 32:138D – 141.

Berdichevsky A, Nedelcu S, Boulias K, Bishop NA, Guarente L, Horvitz HR. 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.

240 Chao H-T, Liu L, Bellen HJ. 2017. Building dialogues between clinical and biomedical research
241 through cross-species collaborations. *Semin. Cell Dev. Biol.* 70:49–57.
242 Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. 2018. Extreme
243 allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural
244 resistance to benzimidazoles.Martin RJ, editor. *PLOS Pathog.* 14:e1007226.
245 Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore
246 R, et al. 2019. WormBase: a modern Model Organism Information Resource. *Nucleic*
247 *Acids Res.*:gkz920.
248 Hieter P, Boycott KM. 2014. Understanding Rare Disease Pathogenesis: A Grand Challenge for
249 Model Organisms. *Genetics* 198:443–445.
250 Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. 2015.
251 PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.*
252 43:D512–D520.
253 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia
254 KM, Ganna A, Birnbaum DP, et al. 2020. The mutational constraint spectrum quantified
255 from variation in 141,456 humans. *Nature* 581:434–443.
256 Kasprzyk A. 2011. BioMart: driving a paradigm change in biological data management.
257 *Database* 2011:bar049–bar049.
258 Kitchen S, Ratnappan R, Han S, Leasure C, Grill E, Iqbal Z, Granger O, O'Halloran DM,
259 Hawdon JM. 2019. Isolation and characterization of a naturally occurring multidrug-
260 resistant strain of the canine hookworm, Ancylostoma caninum. Int. J. Parasitol. 49:397–
261 406.
262 Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, Hoffman D, Jang W, Kaur K, Liu
263 C, et al. 2020. ClinVar: improvements to accessing data. *Nucleic Acids Res.* 48:D835–
264 D844.
265 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH,
266 Ware JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic
267 variation in 60,706 humans. *Nature* 536:285–291.
268 Linard B, Allot A, Schneider R, Morel C, Ripp R, Bigler M, Thompson JD, Poch O, Lecompte
269 O. 2015. OrthoInspector 2.0: Software and database updates. *Bioinformatics* 31:447–448.
270 McGurk L, Berson A, Bonini NM. 2015. *Drosophila* as an *In Vivo* Model for Human
271 Neurodegenerative Disease. *Genetics* 201:377–402.
272 Paquay S, Bourillon A, Pichard S, Benoist J-F, de Lonlay P, Dobbelaere D, Fouilhoux A, Guffon
273 N, Rouvet I, Labarthe F, et al. 2017. Mitochondrial acetoacetyl-CoA thiolase deficiency:
274 basal ganglia impairment may occur independently of ketoacidosis. *J. Inherit. Metab.*
275 *Dis.* 40:415–422.
276 Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei
277 T, Creevey C, Kuhn M, et al. 2014. eggNOG v4.0: nested orthology inference across
278 3686 organisms. *Nucleic Acids Res.* 42:D231–D239.
279 Shendure J, Akey JM. 2015. The origins, determinants, and consequences of human mutations.
280 *Science* 349:1478–1483.
281 Sherry ST, Ward M, Sirotkin K. 1999. dbSNP-database for single nucleotide polymorphisms and
282 other classes of minor genetic variation. *Genome Res.* 9:677–679.
283 Sonnhammer ELL, Östlund G. 2015. InParanoid 8: orthology analysis between 273 proteomes,
284 mostly eukaryotic. *Nucleic Acids Res.* 43:D234–D239.
285 Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A,

286 Gogarten SM, Kang HM, et al. 2019. Sequencing of 53,831 diverse genomes from the
287 NHLBI TOPMed Program. Genomics Available from:
288 http://biorxiv.org/lookup/doi/10.1101/563866
289 Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG,
290 Creatore C, Dawson E, et al. 2019. COSMIC: the Catalogue Of Somatic Mutations In
291 Cancer. *Nucleic Acids Res.* 47:D941–D947.
292 Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, Au V, Chaudhry I,
293 Fernando L, Hutter H, et al. 2013. The million mutation project: A new approach to
294 genetics in Caenorhabditis elegans. *Genome Res.* 23:1749–1762.
295 Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. SIFT missense predictions for
296 genomes. *Nat. Protoc.* 11:1–9.
297 Wang J, Al-Ouran R, Hu Y, Kim S-Y, Wan Y-W, Wangler MF, Yamamoto S, Chao H-T,
298 Comjean A, Mohr SE, et al. 2017. MARRVEL: Integration of Human and Model
299 Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome.
300 *Am. J. Hum. Genet.* 100:843–853.
301 Wang T, Zhan Xiaowei, Bu C-H, Lyon S, Pratt D, Hildebrand S, Choi JH, Zhang Z, Zeng M,
302 Wang K, et al. 2015. Real-time resolution of point mutations that cause phenovariance in
303 mice. *Proc. Natl. Acad. Sci.* 112:E440–E449.
304 Yamamoto S, Jaiswal M, Charng W-L, Gambin T, Karaca E, Mirzaa G, Wiszniewski W,
305 Sandoval H, Haelterman NA, Xiong B, et al. 2014. A Drosophila Genetic Resource of
306 Mutants to Study Mechanisms Underlying Human Genetic Diseases. *Cell* 159:200–214.

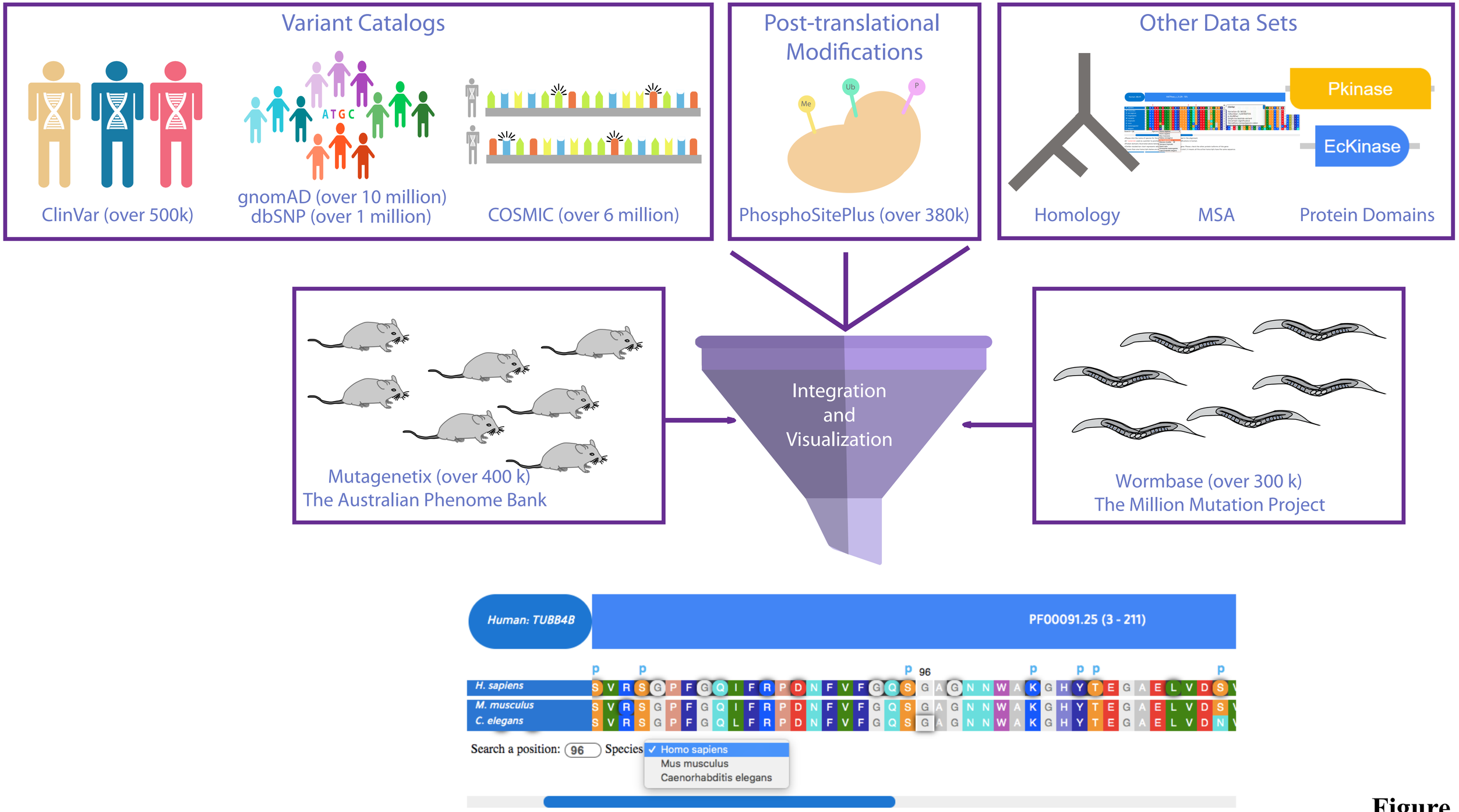

Variant Catalogs
ATGC
ClinVar (over 500k)
gnomAD (over 10 million)
dbSNP (over 1 million)
COSMIC (over 6 million)
Post-translational Modifications
Me
Ub
P
PhosphoSitePlus (over 380k)
Other Data Sets
Pkinase
EcKinase
Homology
MSA
Protein Domains
Mutagenetix (over 400 k)
The Australian Phenome Bank
Integration and Visualization
Wormbase (over 300 k)
The Million Mutation Project
Human: TUBB4B
PF00091.25 (3 - 211)
H. sapiens
M. musculus
C. elegans
Search a position: 96
Species
Homo sapiens
Mus musculus
Caenorhabditis elegans

Figure 1

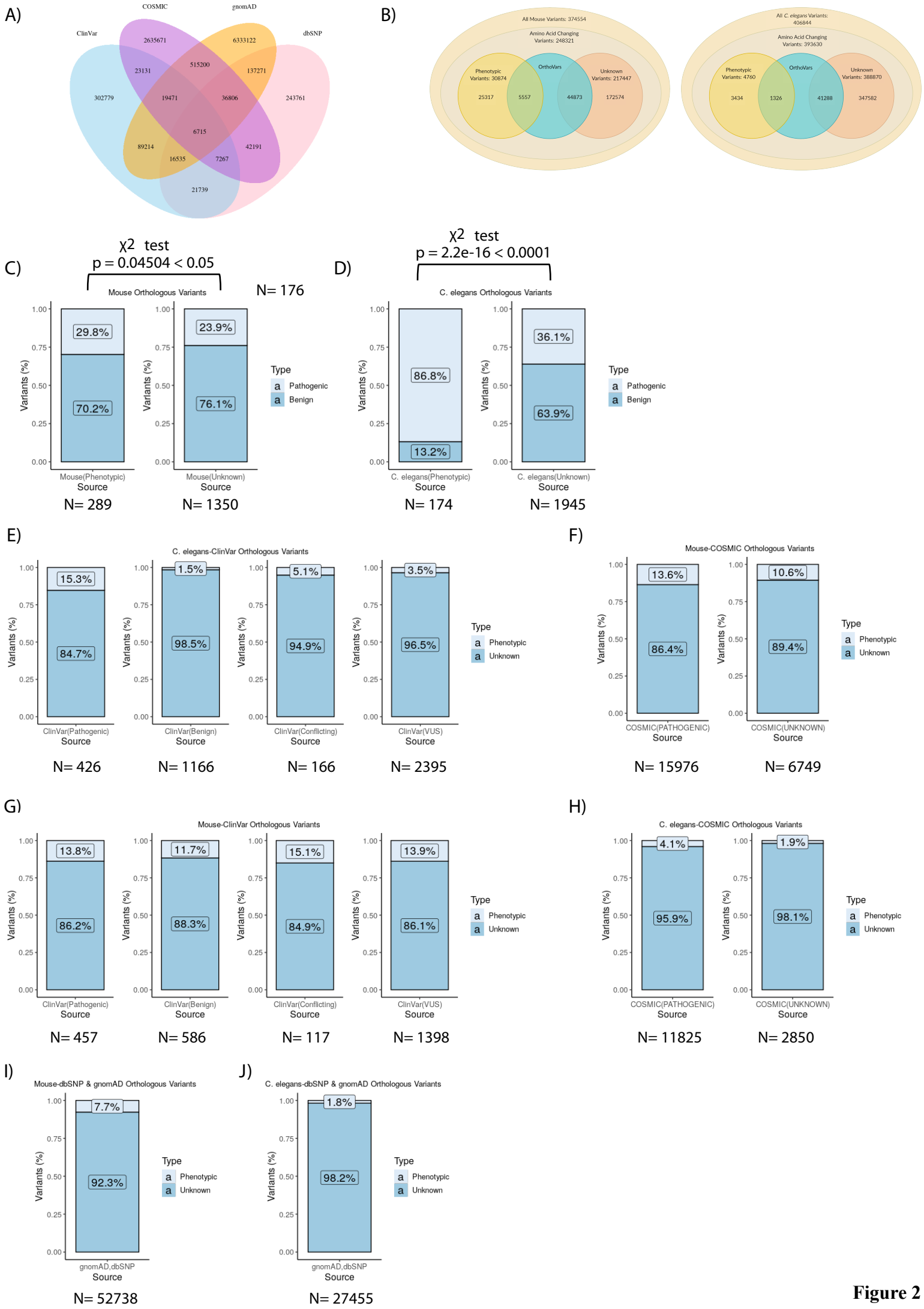

A)
COSMIC
gnomAD
ClinVar
dbSNP
2635671
6333122
23131
515200
137271
302779
19471
36806
243761
6715
89214
42191
16535
7267
21739
B)
All Mouse Variants: 374554
Amino Acid Changing Variants: 248321
Phenotypic Variants: 30874
OrthoVars
Unknown Variants: 217447
25317
5557
44873
172574
All C. elegans Variants: 406844
Amino Acid Changing Variants: 393630
Phenotypic Variants: 4760
OrthoVars
Unknown Variants: 388870
3434
1326
41288
347582
C)
χ2 test
p = 0.04504 < 0.05
Mouse Orthologous Variants
N= 176
29.8%
70.2%
23.9%
76.1%
Variants (%)
Mouse(Phenotypic)
Mouse(Unknown)
Source
N= 289
N= 1350
Type
Pathogenic
Benign
D)
χ2 test
p = 2.2e-16 < 0.0001
C. elegans Orthologous Variants
86.8%
13.2%
36.1%
63.9%
C. elegans(Phenotypic)
C. elegans(Unknown)
N= 174
N= 1945
E)
C. elegans-ClinVar Orthologous Variants
15.3%
84.7%
1.5%
98.5%
5.1%
94.9%
3.5%
96.5%
ClinVar(Pathogenic)
ClinVar(Benign)
ClinVar(Conflicting)
ClinVar(VUS)
N= 426
N= 1166
N= 166
N= 2395
Type
Phenotypic
Unknown
F)
Mouse-COSMIC Orthologous Variants
13.6%
86.4%
10.6%
89.4%
COSMIC(PATHOGENIC)
COSMIC(UNKNOWN)
N= 15976
N= 6749
G)
Mouse-ClinVar Orthologous Variants
13.8%
86.2%
11.7%
88.3%
15.1%
84.9%
13.9%
86.1%
N= 457
N= 586
N= 117
N= 1398
H)
C. elegans-COSMIC Orthologous Variants
4.1%
95.9%
1.9%
98.1%
N= 11825
N= 2850
I)
Mouse-dbSNP & gnomAD Orthologous Variants
7.7%
92.3%
gnomAD,dbSNP
N= 52738
J)
C. elegans-dbSNP & gnomAD Orthologous Variants
1.8%
98.2%
N= 27455

Figure 2

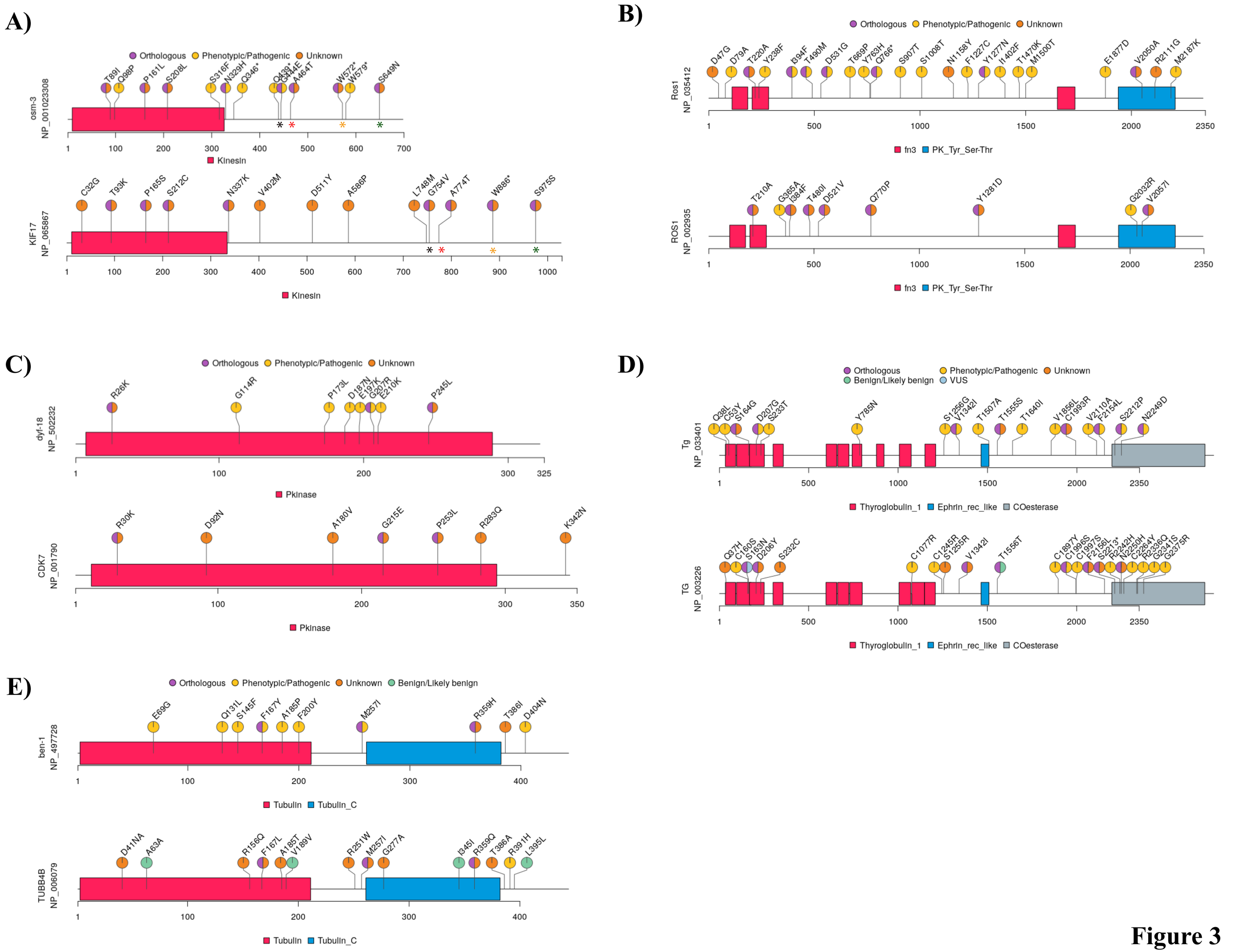

A)
Orthologous
Phenotypic/Pathogenic
Unknown
osm-3
NP_001023308
T89I
Q98P
P161L
S208L
S316F
N329H
Q346*
Q439*
G444E
A464T
W572*
W579*
S649N
Kinesin
KIF17
NP_065867
C32G
T93K
P165S
S212C
N337K
V402M
D511Y
A586P
L748M
G754V
A774T
W886*
S975S
Kinesin
B)
Orthologous
Phenotypic/Pathogenic
Unknown
Ros1
NP_035412
D47G
D79A
T220A
Y238F
I394F
T490M
D531G
T669P
Y763H
Q766*
S907T
S1008T
N1158Y
F1227C
Y1277N
I1402F
T1470K
M1500T
E1877D
V2050A
R2111G
M2187K
fn3
PK_Tyr_Ser-Thr
ROS1
NP_002935
T210A
G365A
I384F
T480I
D521V
Q770P
Y1281D
G2032R
V2057I
fn3
PK_Tyr_Ser-Thr
C)
Orthologous
Phenotypic/Pathogenic
Unknown
dyf-18
NP_502232
R26K
G114R
P173L
D187N
E197K
G207R
E210K
P245L
Pkinase
CDK7
NP_001790
R30K
D92N
A180V
G215E
P253L
R283Q
K342N
Pkinase
D)
Orthologous
Phenotypic/Pathogenic
Unknown
Benign/Likely benign
VUS
Tg
NP_033401
Q38L
C53Y
S164G
D207G
S233T
Y785N
S1256G
V1342I
T1507A
T1555S
T1640I
V1856L
C1993R
V2110A
F2154L
S2212P
N2249D
Thyroglobulin_1
Ephrin_rec_like
COesterase
TG
NP_003226
Q37H
C160S
S163N
D206Y
S232C
C1077R
C1245R
S1255R
V1342I
T1556T
C1897Y
C1996S
C1997S
F2156L
S2213*
R2242H
N2250H
C2264Y
R2336Q
G2341S
G2375R
Thyroglobulin_1
Ephrin_rec_like
COesterase
E)
Orthologous
Phenotypic/Pathogenic
Unknown
Benign/Likely benign
ben-1
NP_497728
E69G
Q131L
S145F
F167Y
A185P
F200Y
M257I
R359H
T386I
D404N
Tubulin
Tubulin_C
TUBB4B
NP_006079
D41NA
A63A
R156Q
F167L
A185T
V189V
R251W
M257I
G277A
I345I
R359Q
T386A
R391H
L395L
Tubulin
Tubulin_C

Figure 3