It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract word count: 200 (max 200)	Manuscript word count: 1858 (max 2000)
Display items: 2 (max 2)	Reference count: 20 (max 20)

# A comparative medical genomics approach may facilitate the interpretation of rare missense variation

Bushra Haque<sup>1,2,\*</sup>, George Guirguis<sup>1,2,\*</sup>, Meredith Curtis<sup>1,2</sup>, Hera Mohsin<sup>2</sup>, Susan Walker<sup>3</sup>, Michelle M. Morrow<sup>4</sup>, Gregory Costain, MD, PhD, FRCPC<sup>1-3,5</sup>

<sup>1</sup>Program in Genetics and Genome Biology, SickKids Research Institute, Toronto, ON, Canada

<sup>2</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

<sup>3</sup>The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada

<sup>4</sup>GeneDx, Gaithersburg, Maryland, USA

<sup>5</sup>Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, and Department of Paediatrics, University of Toronto, Toronto, ON, Canada

\*These authors contributed equally

Address correspondence to: Gregory Costain, gregory.costain@sickkids.ca

Short title: Comparative medical genetics

Keywords: genetic testing; variant interpretation; rare disease; veterinary medicine

It is made available under a CC-BY-NC-ND 4.0 International license .

### ABSTRACT

**Purpose:** To determine the degree to which likely causal missense variants of single-locus traits in non-human livestock and domestic species have features suggestive of pathogenicity in a human genomic context.

**Methods:** We extracted missense variants from the Online Mendelian Inheritance in Animals database for nine animals (cat, cattle, chicken, dog, goat, horse, pig, rabbit, sheep), mapped coordinates to the human reference genome, and annotated variants using genome analysis tools. We also searched a private commercial laboratory database of genetic testing results from >400,000 individuals with suspected rare disorders.

**Results:** Of 339 variants that were mappable to the same residue and gene in the human genome, 56 had been previously classified with respect to pathogenicity: 31 (55.4%) pathogenic/likely pathogenic, 1 (1.8%) benign/likely benign, and 24 (42.9%) uncertain/other. The odds ratio for a pathogenic/likely pathogenic classification in ClinVar was 7.0 (95% confidence interval: 4.1-12.0, p<0.0001), compared to all other germline missense variants in these same 220 genes. The remaining 283 variants disproportionately had allele frequencies and REVEL scores that supported pathogenicity.

**Conclusion:** Cross-species comparisons could facilitate the interpretation of rare missense variation. These results provide further support for comparative medical genomics approaches that connect big data initiatives in human and veterinary genetics.

It is made available under a CC-BY-NC-ND 4.0 International license .

## **INTRODUCTION**

Genome-wide sequencing provides a near-comprehensive view of the exonic single nucleotide variants (SNVs) that contribute to a breadth of human diseases.<sup>1</sup> Distinguishing pathogenic missense variants that cause rare disease from missense variants that are benign or have no functional consequence remains a major challenge in human genetics.<sup>2</sup> In the companion field of veterinary medicine, genetic testing of economically important and domesticated animals for Mendelian traits and diseases is increasingly common.<sup>3</sup> Many first- and second-generation *in silico* prediction tools for missense variants incorporate evolutionary conservation of the amino acid residue,<sup>2 4</sup> and a new wave of bioinformatic tools are learning from "tolerated" variation in non-human primates and other animals.<sup>5 6</sup> There is also a long-standing history of isolated observations that specific variants are pathogenic in both human and non-human orthologs (e.g., <sup>7</sup>). However, the generalizability of a "comparative medical genomics" approach that correlates disease associations of specific variants across species is unknown.

Modeled after Online Mendelian Inheritance in Man (OMIM; omim.org/), the Online Mendelian Inheritance in Animals (OMIA; omia.org/) database includes "likely causal variants" of "single-locus traits" in non-human animal species that were expertly curated from the published literature.<sup>3</sup> Beginning in 1995, by 2023 OMIA listed 1,577 likely causal variants from 485 species.<sup>3</sup> Many of the "traits" studied in a veterinary medicine context are severe phenotypes reminiscent of human genetic diseases (e.g., connective tissue, haematologic, skeletal, or neurogenetic disorders; Supplemental Figure 1). We hypothesized that leveraging these easily accessible but previously siloed OMIA data to identify additional evidence for pathogenicity could aid in the interpretation of human missense variants. The purpose of this study was to systematically assess the degree to

It is made available under a CC-BY-NC-ND 4.0 International license .

which pathogenic missense variation observed in a non-human animal genome possess features suggestive of pathogenicity in a human genome context.

## **METHODS**

#### Identification of non-human animal variants and their human equivalents

We extracted all 442 missense single nucleotide variants (SNVs) from OMIA (accessed: January 2023) that were classified as likely causal for Mendelian phenotypes in one of nine animals (cat, cattle, chicken, dog, goat, horse, pig, rabbit, sheep) and where genomic coordinates were reported in OMIA or in the original scientific publication reporting the variant. Four hundred variants were successfully mapped to the human reference genome (GRCh38) using UCSC's LiftOver tool. We manually inspected each genomic region in the UCSC human genome browser to confirm that (i) the reference amino acid residue in the non-human animal species with the OMIA variant entry matched the reference amino acid residue in humans, and (ii) the alternate amino acid residue would be created through the specific nucleotide change reported in the non-human species; this resulted in the exclusion of an additional 61 variants.

#### Searching in public and private databases

We annotated the remaining 339 human genome variants using ANNOVAR<sup>8</sup> and custom R scripts, including for allele frequency in gnomAD (v3.1.2 and v2.1.1),<sup>9</sup> presence in ClinVar (download date: August 29, 2022),<sup>10</sup> phyloP conservation score,<sup>11</sup> REVEL score,<sup>12</sup> and AlphaMissense score.<sup>13</sup> Variants were also searched in Leiden Open Variation Database (LOVD v3.0)<sup>14</sup> and via Franklin by Genoox (franklin.genoox.com) in winter 2023. We reviewed the primary literature regarding each variant to determine whether the classification or reporting of the variant in one

It is made available under a CC-BY-NC-ND 4.0 International license .

species was directly informed by findings in another species and, for a subset, to summarize the evidence type(s) contributing to a reported association between a specific missense variant and a phenotype in a non-human animal species. Last, we cross-referenced these 339 variants with a large private commercial genetic laboratory (GeneDx) database of genetic testing results from >400,000 individuals with suspected rare Mendelian disorders and their family members. The de-identified data was assessed in accordance with an IRB-approved protocol (WIRB #20171030).

### Statistical methods

Standard descriptive statistics, and parametric and non-parametric tests, were performed using R statistical software, version 4.1.0 (R Foundation for Statistical Computing) with two-tailed statistical significance set at p < 0.05.

#### RESULTS

In total, 339 missense SNVs across 220 different genes, initially identified in non-human animal species from OMIA, were studied in humans (Figure 1; Supplemental Table 1). There were no variants present in two or more of the non-human species. 159 (72.3%) of the genes were associated with germline Mendelian disease(s) in OMIM (Supplemental Table 2). All but one variant was rare, with a minor allele frequency <0.001 in gnomAD. Most variants (n=228; 67.3%) had REVEL scores  $\geq$ 0.644 (supporting or greater evidence for pathogenicity) and only 13.6% variants had REVEL scores  $\leq$ 0.290 (supporting or greater evidence for benign-ness).<sup>15</sup>

### Non-human animal variants were often classified as pathogenic when seen in humans

Of the 339 variants, 56 variants in 49 genes had been previously seen in humans and classified with respect to pathogenicity in ClinVar: 31 (55.4%) as P/LP, 24 (42.9%) as variants of uncertain significance or with conflicting interpretations of pathogenicity (VUS/CIP), and 1 (1.8%) as benign/likely benign (B/LB) (Figure 1). The human Mendelian disease phenotypes associated with these 49 genes were typically concordant with the phenotypes observed in the corresponding nonhuman animal species (Supplemental Figures 1-2), and included both common (e.g., classic Ehlers-Danlos syndrome) and rare (e.g., geleophysic dysplasia) genetic diseases familiar to medical geneticists. Similarly, 23 of the variants were detected across a total of 172 different families that underwent testing at GeneDx: 15 were classified as P/LP, 7 as VUS, and 1 as B/LB. No additional variants were found in LOVD. The odds ratio for these variants having a P/LP classification in ClinVar was 7.0 (95% confidence interval: 4.1-12.0, p<0.0001), when compared to all other germline missense variants with ClinVar entries in the 220 genes (n=45,925) (Supplemental Figure 3). Determinations of a "likely causal" genotype-phenotype association in a non-human animal species were informed by similar principles as are used in medical genetics practice (Supplemental Figure 4), with an average of 2.75 categories of evidence applied in each report. We found direct references to non-human animal findings informing assessment of the human variant in only 10 cases and noted that the term "OMIA" appears once (accession number VCV000016168.16) in the ClinVar database of 3,373,166 records (search date: June 2023). Conversely, we found references to an equivalent human variant in the initial publication and assessment of the non-human animal variant in 28 instances.

#### Understanding discordance between human and non-human variant consequences

It is made available under a CC-BY-NC-ND 4.0 International license .

We reviewed the variant classified in humans as B/LB, to look for plausible reason(s) for discordant classifications between humans and other animals. The B/LB variant in ClinVar is NM\_002386.4(MC1R):c.274G>A;p.(Val92Met) (ClinVar Accession: VCV000014308.11), which has an allele frequency of >0.05 in many human populations. This variant in the melanocyte stimulating hormone receptor gene is a well-studied functional polymorphism associated with pigmentation in humans and in pigs.<sup>16 17</sup> The impact of the variant is therefore concordant across species, in line with our initial hypothesis. The B/LB variant in the GeneDx dataset is NM\_004006.3(DMD):c.5869C>T;p.(Arg1957Trp) and was identified in the hemizygous state in an adult. The REVEL score is 0.22 and hemizygotes (n=3) are reported in the ExAC and ABraOM population databases.<sup>18</sup> Subsequent review of the original study in pigs<sup>19</sup> revealed that the variant was associated with "stress syndrome" rather than a muscular dystrophy phenotype, and it remains possible that the variant is a risk factor for similar decompensation in humans.

### In silico scores for the remaining variants were suggestive of pathogenicity in humans

The remaining 283 variants absent from ClinVar were found across 190 genes (Figure 1). Compared with all missense SNVs in ClinVar in these same genes (n=33,526), the REVEL scores of the OMIA variants (median: 0.78; 69.8%  $\geq$ 0.644) more closely resembled P/LP variants (median: 0.93; 78.4%  $\geq$ 0.644) than B/LB variants (median: 0.30; 16.2%  $\geq$ 0.644). To account for potential confounding by residue conservation scores, we restricted to variants with phyloP score  $\geq$ 0.8. OMIA variants continued to resemble P/LP variants more than B/LB variants with respect to REVEL scores (Figure 2). Findings were similar using AlphaMissense (Supplemental Figure 5).

It is made available under a CC-BY-NC-ND 4.0 International license .

#### DISCUSSION

Massive amounts of genomic data are being generated in human populations, but also in domesticated animals like dogs and cattle. While comparative genomics is a well-established field,<sup>20</sup> and evolutionary conservation of amino acid residues is a foundational component of *in silico* tools for rare variant interpretation, the degree to which the impact of specific rare missense variants at shared residues is concordant across species ("comparative medical genomics") was understudied. Our results suggest that the missense variants in human genomes that correspond to likely causal variants of single-locus traits in other animals are more likely to be pathogenic (functionally significant). The presence of a variant in OMIA suggests that additional published evidence from a non-human species exists that may be relevant to understanding the impact of the orthologous variant in humans. These observations have implications for rare variant interpretation in medical and veterinary medicine contexts.

While our focus in this report was on the application of non-human animal genomic data to interpreting variation in humans, we suspect that the association is bi-directional. Similarly, although there were no variants present in two or more of the non-human species, we hypothesize that missense variants at conserved residues across two or more species are likely to have a high degree of correlation with respect to functional consequences. There are also several limitations of this study. Consideration of this approach requires conservation of the gene and the residue. There are no consensus variant classification guidelines yet for veterinary medicine / non-human contexts, and the types and quality of evidence supporting a "likely causal" attribution in OMIA will vary by entry. The possibility of circularities in the variant interpretation and reporting process (i.e., that variants were deemed pathogenic in a non-human species and included in OMIA because

It is made available under a CC-BY-NC-ND 4.0 International license .

of pre-existing findings in humans) cannot be excluded. The presence of an orthologous variant in OMIA is suspected but has not been shown to be independent of other lines of evidence typically used in human variant interpretation. There is complete overlap between the ClinVar and GeneDx databases, and so our results should not be overinterpreted as having replicated the findings in two independent cohorts. The added value of the data from GeneDx derives from the relative homogeneity in how evidence was applied for variant interpretation and the ability to look at total number of variant observations/families potentially impacted by OMIA entries.

Our results support the generalizability of a comparative medical genomics approach. The existence of potentially relevant information from a non-human species (as indicated by the variant appearing in OMIA) should prompt a dedicated review of the primary literature supporting the OMIA entry, in situations where additional evidence is needed to interpret the human orthologous variant. The benefits to human medical genetics of leveraging data from veterinary medicine could extend beyond establishing spontaneous large animal disease models for human conditions. Our work is intended to foster new collaborations that bridge previously siloed areas of Mendelian diagnostics, as the many lessons learned from formalizing human DNA variant classifications over the past decade could inform veterinary genetics practice. We anticipate that the coming decade will result in orders of magnitude more genome-wide sequencing in domestic animals and livestock species, through large-scale coordinated academic projects, industry-sponsored research, and direct-to-consumer pet testing. The degree to which the exponential growth of human and veterinary genomic datasets can be integrated and harnessed to improve variant interpretation across both contexts warrants additional study.

It is made available under a CC-BY-NC-ND 4.0 International license .

### ACKNOWLEDGEMENTS

The authors thank the many contributors to the public and private databases used in this study.

### **CONTRIBUTORS**

GC and SW conceptualized the study. BH, GG, MC, HM, and MMM curated the data. BH and GG conducted formal analysis of the data. BH and GC acquired the funding for the study. GC supervised the study. BH contributed to all visualization of results. BH, GG, and GC drafted the manuscript, and MC, HM, SW, and MMM were given the opportunity to revise it critically for important intellectual content. All authors give final approval of the submitted version and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## FUNDING

SickKids Research Institute, Canadian Institutes of Health Research, and the University of Toronto McLaughlin Centre. The funders had no role in the design and conduct of the study.

#### **COMPETING INTERESTS**

SW is an employee of Genomics England Limited. MMM is an employee of GeneDx, LLC. The remaining authors have no potential conflicts of interest to declare.

### PATIENT CONSENT FOR PUBLICATION

Not applicable.

It is made available under a CC-BY-NC-ND 4.0 International license .

## **ETHICS APPROVAL**

This secondary use data study was approved by the Research Ethics Board at the Hospital for Sick Children. The de-identified data from GeneDx was assessed in accordance with an IRB-approved protocol (WIRB #20171030).

# DATA AVAILABILITY STATEMENT

All data relevant to the study are available in a public, open access repository (OMIA; omia.org/) and/or are included in the article and uploaded as supplementary information.

It is made available under a CC-BY-NC-ND 4.0 International license .

## REFERENCES

- 1. Costain G, Cohn RD, Scherer SW, Marshall CR. Genome sequencing as a diagnostic test. *CMAJ* 2021;193(42):E1626-E29.
- 2. Costain G, Andrade DM. Third-generation computational approaches for genetic variant interpretation. *Brain* 2023;146(2):411-12.
- 3. Nicholas FW. Online Mendelian Inheritance in Animals (OMIA): a record of advances in animal genetics, freely available on the Internet for 25 years. *Anim Genet* 2021;52(1):3-9.
- 4. Spielmann M, Kircher M. Computational and experimental methods for classifying variants of unknown clinical significance. *Cold Spring Harb Mol Case Stud* 2022;8(3)
- Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, Gal Y, Marks DS. Disease variant prediction with deep generative models of evolutionary data. *Nature* 2021;599(7883):91-95.
- 6. Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, Yang Y, Dietrich ASD, Fiziev PP, Kuderna LFK, Sundaram L, Wu Y, Adhikari A, Field Y, Chen C, Batzoglou S, Aguet F, Lemire G, Reimers R, Balick D, Janiak MC, Kuhlwilm M, Orkin JD, Manu S, Valenzuela A, Bergman J, Rousselle M, Silva FE, Agueda L, Blanc J, Gut M, de Vries D, Goodhead I, Harris RA, Raveendran M, Jensen A, Chuma IS, Horvath JE, Hvilsom C, Juan D, Frandsen P, de Melo FR, Bertuol F, Byrne H, Sampaio I, Farias I, do Amaral JV, Messias M, da Silva MNF, Trivedi M, Rossi R, Hrbek T, Andriaholinirina N, Rabarivola CJ, Zaramody A, Jolly CJ, Phillips-Conroy J, Wilkerson G, Abee C, Simmons JH, Fernandez-Duque E, Kanthaswamy S, Shiferaw F, Wu D, Zhou L, Shao Y, Zhang G, Keyyu JD, Knauf S, Le MD, Lizano E, Merker S, Navarro A, Bataillon T, Nadler T, Khor CC, Lee J, Tan P, Lim WK, Kitchener AC, Zinner D, Gut I, Melin A, Guschanski K, Schierup MH, Beck RMD, Umapathy G, Roos C, Boubli JP, Lek M, Sunyaev S, O'Donnell-Luria A, Rehm HL, Xu J, Rogers J, Marques-Bonet T, Farh KK. The landscape of tolerated genetic variation in humans and primates. *Science* 2023;380(6648):eabn8153.
- 7. Van Poucke M, Martle V, Van Brantegem L, Ducatelle R, Van Ham L, Bhatti S, Peelman LJ. A canine orthologue of the human GFAP c.716G>A (p.Arg239His) variant causes Alexander disease in a Labrador retriever. *Eur J Hum Genet* 2016;24(6):852-6.
- 8. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38(16):e164.
- 9. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581(7809):434-43.
- 10. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K,

It is made available under a CC-BY-NC-ND 4.0 International license .

Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46(D1):D1062-D67.

- 11. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 2010;20(1):110-21.
- 12. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016;99(4):877-85.
- 13. Cheng J, Novati G, Pan J, Bycroft C, Zemgulyte A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, Schneider RG, Senior AW, Jumper J, Hassabis D, Kohli P, Avsec Z. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 2023;381(6664):eadg7492.
- 14. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011;32(5):557-63.
- 15. Pejaver V, Byrne AB, Feng BJ, Pagel KA, Mooney SD, Karchin R, O'Donnell-Luria A, Harrison SM, Tavtigian SV, Greenblatt MS, Biesecker LG, Radivojac P, Brenner SE, ClinGen Sequence Variant Interpretation Working G. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet 2022;109(12):2163-77.
- 16. Xu X, Thornwall M, Lundin LG, Chhajlani V. Val92Met variant of the melanocyte stimulating hormone receptor gene. *Nat Genet* 1996;14(4):384.
- 17. Kijas JM, Wales R, Tornsten A, Chardon P, Moller M, Andersson L. Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* 1998;150(3):1177-85.
- 18. Naslavsky MS, Scliar MO, Yamamoto GL, Wang JYT, Zverinova S, Karp T, Nunes K, Ceroni JRM, de Carvalho DL, da Silva Simoes CE, Bozoklian D, Nonaka R, Dos Santos Brito Silva N, da Silva Souza A, de Souza Andrade H, Passos MRS, Castro CFB, Mendes-Junior CT, Mercuri RLV, Miller TLA, Buzzo JL, Rego FO, Araujo NM, Magalhaes WCS, Mingroni-Netto RC, Borda V, Guio H, Rojas CP, Sanchez C, Caceres O, Dean M, Barreto ML, Lima-Costa MF, Horta BL, Tarazona-Santos E, Meyer D, Galante PAF, Guryev V, Castelli EC, Duarte YAO, Passos-Bueno MR, Zatz M. Wholegenome sequencing of 1,171 elderly admixed individuals from Sao Paulo, Brazil. *Nat Commun* 2022;13(1):1004.
- 19. Nonneman DJ, Brown-Brandl T, Jones SA, Wiedmann RT, Rohrer GA. A defect in dystrophin causes a novel porcine stress syndrome. *BMC Genomics* 2012;13:233.
- 20. Zoonomia C. A comparative genomics multitool for scientific discovery and conservation. *Nature* 2020;587(7833):240-45.

It is made available under a CC-BY-NC-ND 4.0 International license .

## **DISPLAY ITEMS**

**Figure 1.** Missense variation from the OMIA database extrapolated to a human genome context. See main text for details. REVEL score thresholds for supporting evidence for pathogenicity (PP3) and for benign-ness (BP4) are from <sup>15</sup>. B/LB, Benign/Likely benign; P/LP, Pathogenic/Likely pathogenic; VUS/CIP, variant of uncertain significance/conflicting interpretations of pathogenicity. Created with BioRender.com.

Figure 2. Boxplots of REVEL scores for the OMIA missense variants absent from ClinVar, comparing with other missense variation in ClinVar in the same gene set, restricting all variants to those with phyloP scores of  $\geq 0.8$ . B/LB, Benign/Likely benign; P/LP, Pathogenic/Likely pathogenic; VUS/CIP, variant of uncertain significance/conflicting interpretations of pathogenicity.





