Clinical Genetics

WILEY

CLINICAL GENETICS

SHORT REPORT OPEN ACCESS

The Excess of Carriers in Rare Disorders Suggests a Nonpathogenic Effect for Most Variants of Uncertain Significance

Stefano Medaglia[1,2] | Jaume Reig-Palou[1,3] | Ariadna Bellés[4] | Nerea Moreno-Ruiz[5] | Jairo Rodríguez[6] | Xavier Armengol[6] | Juan Ignacio Aróstegui[7,8] | Lluís Armengol[6] | Juan José Guillén[4] | Hafid Laayouni[5,9] | Ferran Casals[1,3,10]

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain | [2]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy | [3]Institut de Biomedicina de la Universitat de Barcelona (IBUB), Universitat de Barcelona, Barcelona, Spain | [4]Clínica EUGIN, Barcelona, Spain | [5]Institut de Biologia Evolutiva (UPF-CSIC), department of Medicine and Life Sciences, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, Barcelona, Spain | [6]Quantitative Genomic Medicine Laboratories (qGenomics), Esplugues del Llobregat, Barcelona, Spain | [7]Departament d'Immunologia, Hospital Clínic–Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain | [8]Escola de Medicina, Universitat de Barcelona, Barcelona, Spain | [9]BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain | [10]Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain

Correspondence: Ferran Casals (ferrancasals@ub.edu)

Received: 1 August 2024 | **Revised:** 3 October 2024 | **Accepted:** 22 October 2024

Funding: This work was supported by grants RTI2018-096824-B-C22 (F.C.), PID2021-125106OB-C32 (F.C.), RTI2018-096824-B-C21(J.I.A.), and PID2021-125106OB-C31 (J.I.A.) funded by MCIN/AEI /10.13039/501100011033/ and FEDER Una manera de hacer Europa; Agència de Gestió d'Ajuts Universitaris I de Recerca-Generalitat de Catalunya (2021SGR01093) (F.C.). N.M.-R. was supported by grant 2021 FI_B_00296 from Agència de Gestió d'Ajuts Universitaris i de Recerca, Generalitat de Catalunya.

Keywords: carrier screening | functional annotation | recessive disorders | VOUS

ABSTRACT

Functional annotation and interpretation of genetic variants are a critical step in genetic diagnosis, as it may lead to personalized therapeutic options and genetic counseling. While the number of confirmed pathogenic genetic variants in an individual is relatively low, the number of variants of uncertain significance (VOUS) can be considerably higher, increasing the number of potential carriers of genetic disorders. Thus, reducing uncertainty and assessing the real effect of VOUS are crucial for clinical and medical genetics. In this study, we evaluated the efficacy of genetic screening technologies in accurately predicting pathogenic variants and their corresponding disease prevalence in a cohort of over 6000 healthy individuals involved in assisted reproduction programs. Using data from 305 genes associated with recessive disorders, we determined the frequency of carriers of pathogenic variants and VOUS in our dataset and compared the predicted prevalence based on this information with reported population prevalence data. The higher predicted prevalence in some disorders when considering VOUS suggests a mostly benign effect.

1 | Introduction

The advent of massive sequencing technologies and their rapid evolution in the last decade have multiplied the possibilities of genetic analyses for diagnosis, especially for rare disorders [1]. It also allows screening for carriers of hundreds of genetic recessive conditions, which integrates effectively with assisted reproduction techniques to significantly reduce the risk of monogenic disorders in the offspring [2]. Functional annotation of genetic variants to identify truly pathogenic variation among the thousands of genetic variants in an individual is therefore a critical step for genetic diagnosis and carrier detection [3], and

the need for reliable prediction of the pathogenicity of genetic variants goes in parallel with the demand for standardized classification. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) established the ACMG/AMP guidelines, which represented a cornerstone in this direction [4]. However, in many cases, the predicted variant effect remains inconclusive, especially when considering clinical decisions or genetic counseling. Moreover, difficulties in meeting some of the ACMG/AMP criteria, particularly in large-scale screenings via which new variants are frequently discovered, may lead to these variants being classified as variants of uncertain significance (VOUS), further influencing the predictive accuracy of genetic analyses [5].

Even considering that most VOUS do not have a pathogenic effect, often confirmed by reclassification studies for specific disorders [6], in many cases they still influence patient treatment and genetic counseling [7]. In assisted reproduction, strategies to minimize genetic disease risk might include conservative measures, such as avoiding pairings between individuals carrying combinations of VOUS and pathogenic mutations. This approach could artificially increase the proportion of carriers among donors, which may limit the number of compatible donors. Therefore, a better characterization of VOUS may be critical for the implementation of expanded carrier screenings with an increasing number of assessed genes [5]. This is also relevant in prenatal testing, as the appropriateness and consequences of communicating such findings to parents are also controversial [8]. In this study, we aimed to indirectly evaluate the accuracy of pathogenicity prediction for variants in genes associated with recessive disorders by comparing the predicted prevalence based on the carriers frequency with the reported disease prevalence in the population [9]. We hypothesized that discrepancies between the predicted and reported prevalence values might indicate inaccuracies in variant annotation [10]. To explore this, we analyzed genetic screening data of over 300 genes associated with autosomal and X-linked recessive disorders in more than 6000 healthy individuals enrolled in assisted reproduction programs.

2 | Materials and Methods

2.1 | Genetic Data

We analyzed DNA sequence data for 6461 anonymous individuals in 305 genes related to recessive disorders included in the qCarrier test (https://qgenomics.com/wp-content/uploads/genes-qCarrierTest.pdf). DNA extraction from peripheral blood or saliva was performed using commercial methods based on magnetic beads following manufacturer instructions (Promega Inc., Madison, WI). We performed capture enrichment of target regions including coding exons of known disease-related recessive and X-linked genes before massive parallel sequencing (Illumina Inc., San Diego, USA). Quality control, alignment, variant calling, and annotation were performed using an in-house implementation of standard bioinformatic tools and methods and published elsewhere [11]. Variant pathogenicity was determined using a standardized procedure based on ACMG guidelines [4] after manual assessment of variant annotations and a review of current literature, followed by a revision conducted by a different operator. We performed a principal component analysis (PCA) on 8702 biallelic genetic variants (frequency > 0.1%) in 258 autosomal genes. We detected 1858 outliers, defined as samples with PC1 values more than three standard deviations from the mean, and removed them accordingly. We next performed an analysis to exclude possible related individuals, based on the estimation of the kinship coefficient with variants at frequencies higher than 1% using KING [12] and the percentage of shared rare variants between individuals, which resulted in the removal of 56 individuals.

2.2 | Frequency and Prevalence Analysis

First, we calculated the allele frequency for each variant. We then calculated the predicted prevalence of each disorder based on the frequency of the alternative allele for all variants annotated as pathogenic or VOUS detected in the associated gene. When several genes have been associated with a given disease, predicted prevalence values from genetic variants in those genes were aggregated into a single prevalence value. Reported prevalence values of each disease were retrieved from *Orphanet* point prevalence data (https://www.orpha.net, updated until June 2022). Minimum and maximum values were assigned to each prevalence range. Geographic point prevalence was selected in order of preference: European or Worldwide (Table S1). Reported prevalence values for disorders lacking annotation on *Orphanet* were retrieved manually from *Medline* (https://medlineplus.gov). The statistical significance of the ratios between predicted and reported prevalence was determined using a *Z*-test, applying Bonferroni's multiple-test correction.

3 | Results and Discussion

We performed PCA and kinship analyses (see Section 2) to remove potential outliers that could artificially affect carrier frequency and predicted prevalence estimation, generating a final cohort of 4547 European individuals (2219 females and 2328 males). We identified a total of 2433 pathogenic variants and 13 854 VOUS across the analyzed genes. These variants were unequally distributed and, as expected, correlated mostly with the protein-coding length (Figure S1) and belonged predominantly to low-frequency classes [13] (Figure S2). The average number of variants per individual was 7.59, with a mean of 1.56 pathogenic variants and 6.03 VOUS, consistent with previous studies [14, 15].

We calculated two values of predicted prevalence for each disorder (see Section 2): one considering only pathogenic variants and the other considering both pathogenic variants and VOUS found in the analyzed genes. The predicted prevalence values were then compared to the reported prevalence of the 146 disorders with available epidemiological estimation (Table S1). The results were expressed as a ratio of predicted to reported prevalence of each disorder. For most of the diseases, the reported prevalence was expressed as a range and, therefore the minimum and maximum values were used to calculate the ratio, obtaining "ratio-min" and "ratio-max" values (Figure 1, Table S1). When considering pathogenic variants alone, both ratio-min and ratio-max values spanned from −4 to 2 (in base-10 logarithmic scale),

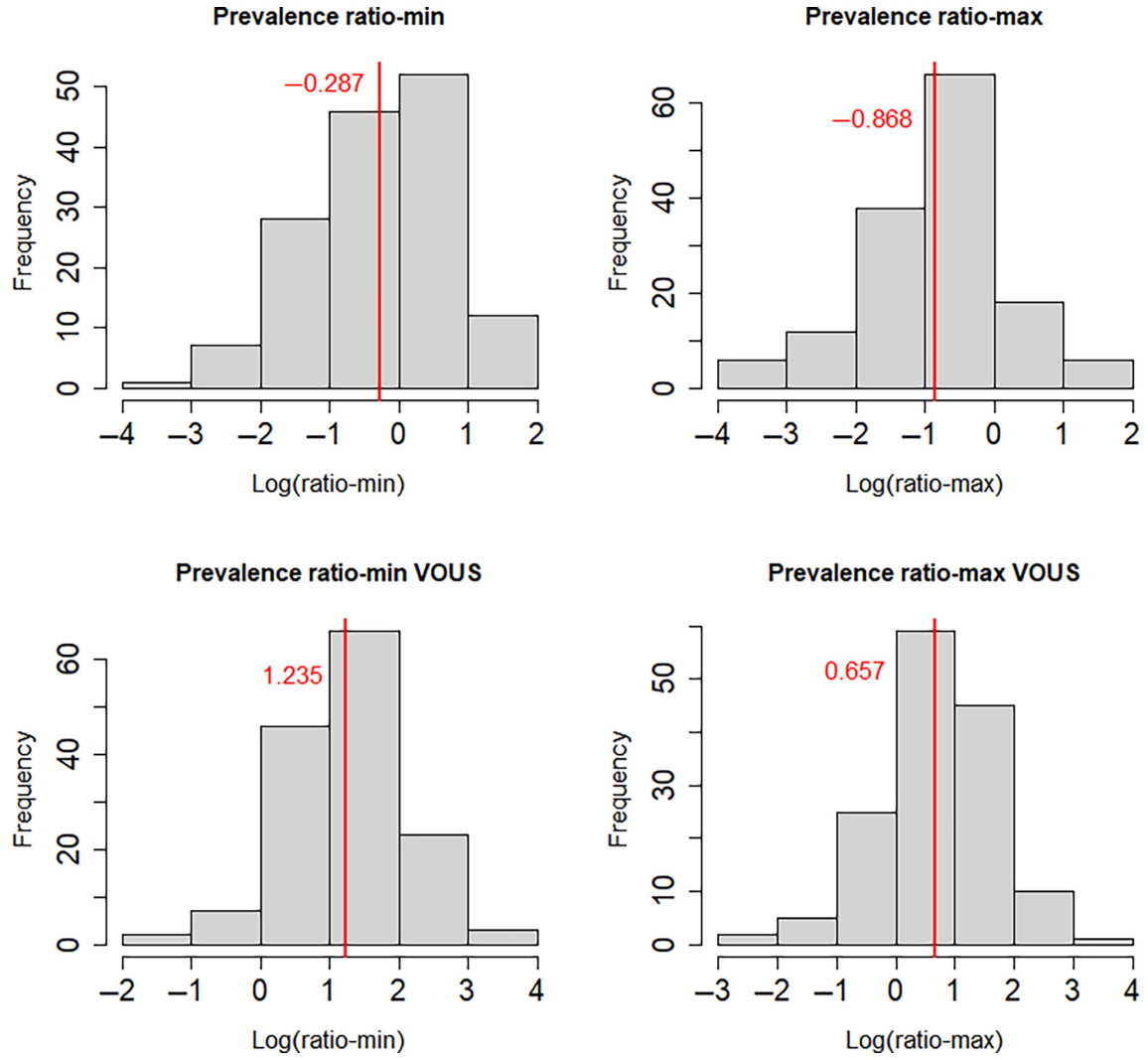

FIGURE 1 | Comparison between the predicted and the reported prevalence in 146 recessive disorders. Distributions of the different ratio-min and ratio-max for each one of the 146 analyzed recessive disorders. The first row shows the ratios of predicted against reported prevalence considering only pathogenic variants, while the second refers to the analysis considering both VOUS and pathogenic variants. *X*-axes show the base-10 logarithmic scale of the ratios; *y*-axes show the number of disorders displaying a particular ratio. The red lines and labels represent the mean value of the ratio for each plot. Values close to or equal to zero denote a predicted prevalence consistent with the reported one. Negative values correspond to predicted prevalence values lower than the reported ones, and vice versa for positive values. [Colour figure can be viewed at wileyonlinelibrary.com]

with a global average of −0.29 and −0.87 respectively (Figure 1). For many disorders, we did not detect significant discrepancies between predicted and reported prevalence, suggesting that functional prediction and annotation of the genetic variants effects are mostly accurate, which makes genetic screening efficient and effective for these pathologies. Some disorders showed strong discrepancies with much lower predicted prevalence values than reported, and although the statistical significance could not be assessed due to the low number of observations, this could suggest a lack of knowledge about the genetic etiology of these diseases.

Interestingly, 32 disorders showed statistically significant differences with higher predicted than reported prevalence values when considering both pathogenic variants and VOUS, even when considering the maximum reported prevalence with ratios ranging from 0.59 to 3.79 (Table S1). However, several factors could potentially inflate the predicted prevalence values. For instance, pathogenic variants in pleiotropic genes have an effect on multiple disorders. In the case of *CFTR* in Cystic Fibrosis, for example, we noticed that predicted and reported values were similar if we considered only unequivocal cystic fibrosis mutations, excluding those related to other phenotypes such as congenital absence of the vas deferens [16]. Additionally, the classification of genetic variants can be biased toward pathogenic or VOUS categories (i.e., considering the most pathogenic prediction). Moreover, linkage disequilibrium can have an effect: two or more VOUS or pathogenic variants co-occurring in the same individual could increase the predicted prevalence value of a given disorder. To refine our estimates, we recalculated the predicted values considering only variants catalogued at least once as pathogenic or VOUS in ClinVar [17]. Furthermore, we limited our analysis to only one variant per gene per individual to avoid the effect of linkage disequilibrium, and we excluded the disorders affected by genes with pleiotropic effects. Following these adjustments, 10 disorders still showed a significantly higher predicted prevalence ($p < 0.001$, Table 1), confirming the suspicion that most VOUS may not have a pathogenic effect [18], in

TABLE 1 | Disorders with a larger discrepancy between predicted and reported prevalence.

Disease	Genes (OMIM)	Reported prevalence[a]	Predicted prevalence	Carriers
Autosomal recessive polycystic kidney disease	PKHD1	1/20000 (0.00005)	0.00239	625
Primary ciliary dyskinesia	DNAH5	1/16000 (0.0000625)	0.0025	640
Familial thyroid dyshormonogenesis	DUOX2, TG, IYD, DUOXA2, SLC5A5, TPO	9/100000 (0.00009)	0.00292	1259 (400 + 522 + 77 + 26 + 71 + 163)
Hemochromatosis type 3	TFR2	1/1000000 (0.000001)	0.00017	160
Fanconi anemia	FANCA, FANC	9/1000000 (0.000009)	0.00049	358 (267 + 91)
Glycine encephalopathy	AMT, GLDC	9/1000000 (0.000009)	0.00036	296 (75 + 221)
Alpha-mannosidosis	MAN2B1	9/1000000 (0.000009)	0.00032	222
Propionic acidemia	PCCB, PCCA	9/1000000 (0.000009)	0.00025	274 (160 + 114)
Wilson disease	ATP7B	9/100000 (0.00009)	0.00086	369
Carnitine palmitoyl transferase 1A deficiency	CPT1A	< 1/1000000 (0.000001)	0.00007	104

[a]Maximum reported value.

line with the results obtained in VOUS reclassification analyses [6]. Alternatively, although we considered the highest reported prevalence to identify significant deviations, this result may also suggest an underestimation of the actual prevalence of the disorder in the general population, or it may be driven by the presence of variants with incomplete penetrance [19, 20].

Our study introduces a rapid and reproducible method to detect possible mismatches between the predicted and the actual number of pathogenic variants for specific disorders. This approach can be used to refine screening programs in areas where reliable prevalence values are available. The general approach maintained in our study opens to a wide spectrum of follow-up investigations, from functional assessments of genetic variants, and a refinement of the prevalence estimation for some disorders, to the adjustment of gene panels used in carrier tests.

Author Contributions

J.J.G., H.L., and F.C. conceived and designed the study. A.B., J.R., X.A., L.A., and J.J.G. provided the data and performed the variant annotation. S.M., J.R.-P., A.B., N.M.-R., J.I.A., H.L., and F.C. performed the analysis. All the authors participated in the discussion and contributed to manuscript writing.

Acknowledgements

We appreciate the participation of patients and donors. We thank Anna Holloway (Eugin) for the English language and grammar editing of the article.

Ethics Statement

The study was approved by The *Eugin Ethics Committee* with the protocol number VARPOP/PD-55-06-ES.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Peer Review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/cge.14642.

References

1. Z. Liu, L. Zhu, R. Roberts, and W. Tong, "Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We?," *Trends in Genetics* 35 (2019): 852–867.

2. S. A. Kraft, D. Duenas, B. S. Wilfond, and K. A. B. Goddard, "The Evolving Landscape of Expanded Carrier Screening: Challenges and Opportunities," *Genetics in Medicine* 21 (2019): 790–797.

3. S. E. Antonarakis, "Carrier Screening for Recessive Disorders," *Nature Reviews. Genetics* 20 (2019): 549–561.

4. S. Richards, N. Aziz, S. Bale, et al., "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and

the Association for Molecular Pathology," *Genetics in Medicine* 17 (2015): 405–423.

5. H. Fridman, D. M. Behar, S. Carmi, and E. Levy-Lahad, "Preconception Carrier Screening Yield: Effect of Variants of Unknown Significance in Partners of Carriers With Clinically Significant Variants," *Genetics in Medicine* 22 (2020): 646–653.

6. E. Hahn, C. Mighton, Y. Fisher, et al., "Variant Classification Changes Over Time in the Clinical Molecular Diagnostic Laboratory Setting," *Journal of Medical Genetics* 61 (2024): 788–793.

7. N. Walsh, A. Cooper, A. Dockery, and J. J. O'Byrne, "Variant Reclassification and Clinical Implications," *Journal of Medical Genetics* 61 (2024): 207–211.

8. G. Watts and A. J. Newson, "To Offer or Request? Disclosing Variants of Uncertain Significance in Prenatal Testing," *Bioethics* 35 (2021): 900–909.

9. E. Vintschger, D. Kraemer, P. Joset, et al., "Challenges for the Implementation of Next Generation Sequencing-Based Expanded Carrier Screening: Lessons Learned From the Ciliopathies," *European Journal of Human Genetics* 31 (2023): 953–961.

10. E. V. Minikel, S. Vallabh, M. Lek, et al., "Quantifying Penetrance in a Dominant Disease Gene Using Large Population Control Cohorts," *Science Translational Medicine* 8 (2016): 322ra9.

11. A. Abulí, M. Boada, B. Rodríguez-Santiago, et al., "NGS-Based Assay for the Identification of Individuals Carrying Recessive Genetic Mutations in Reproductive Medicine," *Human Mutation* 37 (2016): 516–523.

12. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, and W. M. Chen, "Robust Relationship Inference in Genome-Wide Association Studies," *Bioinformatics* 26 (2010): 2867–2873.

13. Y. Kobayashi, S. Yang, K. Nykamp, J. Garcia, S. E. Lincoln, and S. E. Topper, "Pathogenic Variant Burden in the ExAC Database: An Empirical Approach to Evaluating Population Data for Clinical Variant Interpretation," *Genome Medicine* 9 (2017): 13, https://doi.org/10.1186/S13073-017-0403-7.

14. A. Capalbo, R. A. Valero, J. Jimenez-Almazan, et al., "Optimizing Clinical Exome Design and Parallel Gene-Testing for Recessive Genetic Conditions in Preconception Carrier Screening: Translational Research Genomic Data From 14,125 Exomes," *PLoS Genetics* 15 (2019): e1008409.

15. H. Fridman, H. G. Yntema, R. Mägi, et al., "The Landscape of Autosomal-Recessive Pathogenic Variants in European Populations Reveals Phenotype-Specific Effects," *American Journal of Human Genetics* 108 (2021): 608–619.

16. J. L. Deignan, C. Astbury, G. R. Cutting, et al., "CFTR Variant Testing: A Technical Standard of the American College of Medical Genetics and Genomics (ACMG)," *Genetics in Medicine* 22 (2020): 1288–1295.

17. M. J. Landrum, J. M. Lee, G. R. Riley, et al., "ClinVar: Public Archive of Relationships Among Sequence Variation and Human Phenotype," *Nucleic Acids Research* 42 (2014): D980–D985.

18. A. C. M. Joynt, M. M. Axford, L. Chad, and G. Costain, "Understanding Genetic Variants of Uncertain Significance," *Paediatrics and Child Health* 27 (2022): 10–11.

19. R. Kingdom and C. F. Wright, "Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts," *Frontiers in Genetics* 13 (2022): 920390, https://doi.org/10.3389/FGENE.2022.920390.

20. J. van Rooij, P. Arp, L. Broer, et al., "Reduced Penetrance of Pathogenic ACMG Variants in a Deeply Phenotyped Cohort Study and Evaluation of ClinVar Classification Over Time," *Genetics in Medicine* 22 (2020): 1812–1820.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.