# Will variants of uncertain significance still exist in 2030?

Douglas M. Fowler[1,2,3,*] and Heidi L. Rehm[4,5,*]


## Summary

In 2020, the National Human Genome Research Institute (NHGRI) made ten "bold predictions," including that "the clinical relevance of all encountered genomic variants will be readily predictable, rendering the diagnostic designation 'variant of uncertain significance (VUS)' obsolete." We discuss the prospects for this prediction, arguing that many, if not most, VUS in coding regions will be resolved by 2030. We outline a confluence of recent changes making this possible, especially advances in the standards for variant classification that better leverage diverse types of evidence, improvements in computational variant effect predictor performance, scalable multiplexed assays of variant effect capable of saturating the genome, and data-sharing efforts that will maximize the information gained from each new individual sequenced and variant interpreted. We suggest that clinicians and researchers can realize a future where VUSs have largely been eliminated, in line with the NHGRI's bold prediction. The length of time taken to reach this future, and thus whether we are able to achieve the goal of largely eliminating VUSs by 2030, is largely a consequence of the choices made now and in the next few years. We believe that investing in eliminating VUSs is worthwhile, since their predominance remains one of the biggest challenges to precision genomic medicine.


## Introduction

In 2020, the National Human Genome Research Institute (NHGRI) made ten "bold predictions," including that "the clinical relevance of all encountered genomic variants will be readily predictable, rendering the diagnostic designation 'variant of uncertain significance (VUS)' obsolete." On September 16, 2021, we talked about this prediction in the NHGRI Bold Predictions seminar series, and here, we update and expand upon these talks (accessible on the NHGRI website at https://www.genome.gov/event-calendar/Bold-Predictions-for-Human-Genomics-by-2030). This prediction appears to be the boldest of all ten as VUSs deposited in ClinVar have increased by ~5-fold from 2020 when the prediction was published to 2023 (ClinVar; https://www.ncbi.nlm.nih.gov/clinvar/; date accessed August 1, 2023). However, despite this daunting increase, we are hopeful that many if not most VUSs awaiting classification will in fact be resolved by 2030.

Variants observed during clinical genetic testing must be classified for pathogenicity with respect to one or more diseases and modes of inheritance for a given gene in order to be used to diagnose or guide the treatment of disease. Classification depends on integration of different types of evidence, including observations in individuals with disease, segregation in a family, population frequency, functional data, and computational predictions.[1] Unfortunately, evidence is often lacking, especially for rare variants, leading to the VUS classification. VUSs are difficult to apply in the clinic, often representing a dead end that can only be overcome by the collection of additional evidence. One way we track the state of understanding of human variation is through the ClinVar database[2] where laboratories voluntarily share their classified variants. Over 90% of submissions come from clinical laboratories and therefore represent a glimpse of the state of variation being observed in patients. Currently, the ClinVar database contains over 3 million submissions on over 2 million unique variants from over 2,500 submitters in 89 countries (accessed August 1, 2023). While ClinVar represents a major success in community data sharing, the classification of each of those >2 million variants reveals the challenge: 36% are VUSs and 5% are conflicting. 15% are classified as pathogenic (10%) or likely pathogenic (5%) with the remainder classified as benign (18%), likely benign (25%), or other (1%). This predominance of VUSs is reflected in individual genetic testing reports returned to patients: in 2020–2021, 19 laboratories in North America offering multigene panels, exomes, and genomes reported that 32% of individuals received an inconclusive test report due to one or more VUSs in the absence of any clear explanation for disease.[3] In addition, only 11% of reports contained causal pathogenic or likely pathogenic variants, suggesting that most individuals sent for diagnostic testing did not obtain insight into their condition. Furthermore, this picture largely derives from the interrogation of the coding regions of genes, which represent the ~2% of the genome where we are most able to interpret variants but excludes most of the non-coding portion of the genome. Moreover, current testing is focused on single nucleotide variants and is only just beginning to

[1]Department of Genome Sciences, University of Washington, Seattle, WA, USA; [2]Department of Bioengineering, University of Washington, Seattle, WA, USA; [3]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA; [4]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; [5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
*Correspondence: dfowler@uw.edu (D.M.F.), hrehm@mgh.harvard.edu (H.L.R.)
https://doi.org/10.1016/j.ajhg.2023.11.005.

systematically include moderately sized copy number variants below the resolution of chromosomal microarrays, as well as more complex insertions, deletions, and other genomic changes. Thus, as we look into the abyss of the non-coding portions of our genome, grapple with structural variation, and examine genes not yet linked to disease, the task of discerning the few impactful causal variants from the plethora of other variants is challenging. As such, we clearly have much work to do!

So, how will we tackle this seemingly insurmountable VUS challenge? Is there hope over the remaining seven years to come close to achieving the NHGRI's audacious prediction? We believe so. While we may not eliminate all VUSs, much progress is being made, and there is a path to a brighter future where most human variants are understood. But arriving at this future will require bold new thinking, creative technology development, a massive research effort, and community building. Below, we outline key advances and explore how they are creating change and shedding light on a path forward.

## Standards for variant classification

Over 8 years ago, the American College of Medical Genetics (ACMG) and Genomics and the Association for Molecular Pathology (AMP) released more rigorous standards for the classification of genetic variation with respect to monogenic disease.[1] There was widespread appreciation that many variants had been falsely classified as pathogenic and needed to be reclassified as VUSs. The new standards improved the accuracy of variant classification, yet also highlighted the paucity of knowledge we had to interpret most variants. However, since these standards were published, the Clinical Genome Resource (ClinGen) Sequence Variant Interpretation working group has been releasing interim guidance, both clarifying the application of the ACMG/AMP standards as well as building on them, most recently allowing more missense and splice variants to receive higher levels of evidence based on improved analytical algorithms that, in some cases, use machine learning to better predict the impact of the variation.[4,5] We anticipate that the next sequence variant classification standards, currently under development, will continue to improve the accuracy of variant classification and also enhance our ability to use available evidence in adjudicating variant effects. Additional guidance is also being generated around evaluating variants with lower penetrance, as well as differentially evaluating variants in genes associated with multiple diseases and pathogenic mechanisms. Furthermore, as we discover more of the causes of rare disease, for which thousands of gene-disease relationships remain undiscovered,[6] and better understand the mechanisms and relevant domains in which variation in each gene leads to disease, we will have better frameworks to evaluate all variants in those genes and their attendant non-coding regions.

## Tools for predicting variant effects

Computational variant effect predictors use information about a variant, especially from multiple sequence alignments and protein structure, to predict the variant's effect. A key advantage of variant effect predictors is that predictions can be generated for many or most variants across the human genome, meaning that predictions are available for nearly all single-nucleotide variants. However, predictor accuracy was modest and difficult to fairly assess, meaning that predictor-derived evidence had little weight and thus a small influence on clinical variant interpretation.[1] Recently, dramatic innovations have begun to overcome both of these limitations. Prediction quality has improved for a variety of reasons, one of which is the pairing of more sophisticated machine learning approaches with the increased amount of example pathogenic and benign variants that can be used for training predictive models (e.g., Wu et al.[7]). Another example is the rise of "metapredictors" that combine multiple modestly performing predictions (e.g., Ioannidis et al.[8]). Lastly, approaches for incorporating the features that drive predictions have been developed, particularly in the case of multiple sequence alignments (e.g., Gao et al.[9]). Another key driver of predictor improvement has been an appreciation of the problems posed by supervised model circularity (e.g., using known benign and pathogenic variants both to train a model and evaluate its effects) and the development of solutions to this problem (e.g., rigorous definition of test sets and use of unsupervised models; Frazer et al. and Cheng et al.[10,11]). Thus, today, high-quality variant effect predictions are available from numerous models.

However, the development of high-quality predictions did not immediately alter the 2015 ACMG/AMP standards specifying that evidence from computational variant effect predictors could be used at the lowest "supporting" level of evidence for clinical variant classification as genes needed large sets of variants with pathogenicity validated by independent evidence to develop stronger evidence levels. Eventually, numerous studies hinting at the fact that predictors could and should yield stronger evidence combined with rules articulated by the ClinGen Sequence Variant Interpretation working group for rigorously evaluating evidence strength[12] opened the door for a reassessment of predictor evidence. Responding to these factors, the ClinGen Sequence Variant Interpretation working group has recently articulated a new framework for calibrating the strength of evidence provided by computational variant effect predictors.[4] This framework hinges on the analysis of predictor performance using known pathogenic and benign variants from across the genome and has resulted in a dramatic shift in the strength of evidence that

can be generated by predictors. In fact, some predictors can generate strong evidence for variant classification, both for pathogenic effects and benign effects. Because predictions are available for a large number of VUSs, this sea change in predictor evidence strength will likely considerably reduce the number of VUSs in the coming years. Moving forward, dramatic improvements in machine-learning algorithms along with increased numbers of known benign and pathogenic variants will drive continued improvement in predictor accuracy.

## Tools for measuring variant function

Functional assays can be used to measure the effect of a variant on molecular, cellular, or model organism phenotypes. A key strength of functional assays is that, with the inclusion of sufficient control pathogenic and benign variants, the effect of the variant as compared to controls can inform on the clinical consequences of the variant. Indeed, the 2015 ACMG/AMP standards specified that variant functional data could be used as strong evidence for clinical classification if rigorously validated. However, variant functional data were rarely available and often difficult to validate because most datasets contained very limited numbers of control variants. Moreover, calibration of evidence strength from functional assays was challenging, with little guidance initially provided for how to evaluate evidence of assays performed outside of licensed clinical labs.

The advent of multiplexed assays of variant effect has enabled both of these limitations to be addressed. In a multiplexed assay of variant effect, thousands of variant effects are measured simultaneously, generally using high-throughput DNA sequencing as a readout.[13–15] A first generation of multiplexed assays have been developed that encompass a wide variety of molecular and cellular phenotypes ranging from protein activity and abundance to promoter and enhancer activity to cell growth, morphology, and transcriptomic state.[16–18] Multiplexed functional data offer many advantages as compared to their low-throughput predecessors. Most importantly, nearly all possible single-nucleotide or amino acid variants can be simultaneously assessed in a target region of the genome. The result is a variant effect map that can provide evidence for previously identified variants, as well as all variants that will be found in the future. Thus, like computational variant effect predictions, multiplexed functional data are proactive. Moreover, each multiplexed functional experiment includes internal control variants (e.g., synonymous and nonsense variants within a coding region), enabling each dataset to be evaluated for quality according to rigorous standards.[19,20] Multiplexed assays of variant effect are becoming widely adopted, with over 10 million variants assessed as of mid-2023.[21]

However, despite the potential utility of multiplexed functional data, their uptake for clinical variant classification was initially slow. Like for computational predictions, the reason was because existing standards did not specify a rigorous, universally applicable way to calibrate the strength of evidence provided by each functional dataset. Just before the NHGRI issued its bold predictions, the ClinGen Sequence Variant Interpretation workgroup articulated such standards.[12] Application of these new standards to multiplexed functional data for three key cancer-risk genes, *BRCA1*, *TP53*, and *PTEN*, revealed that the inclusion of properly calibrated functional data from multiplexed assays could lead to the reclassification of many, if not most VUSs.[22] Thus, multiplexed functional data are poised to play an important role in reducing the number of VUSs.

Moving forward, several key challenges remain. Multiplexed functional data are currently expensive to generate and available for only a small minority of clinically related genes, meaning that community-scale efforts to generate such data are required. The first generation of multiplexed functional assays, while powerful, cannot be extended to every clinically related gene because many of these genes function in processes that cannot currently be assayed or in specialized cell types for which there are not multiplexable models. Existing technologies are also largely unable to account for genetic and environmental context, both of which can impact variant effects. Lastly, existing technologies are largely focused on single-nucleotide or amino acid variants, and few approaches exist for querying insertions, deletions, or more complex events at scale. Thus, continued development of multiplexed assay technology is needed to reduce costs and overcome these challenges.

Many details remain to be resolved surrounding the use of functional data in clinical variant classification. Evaluating the strength of evidence generated by a functional dataset is difficult for genes where few control pathogenic and benign variants are available. Classification of more pathogenic and benign variants, particularly in partnership with a variant curation expert panel, is one possible solution. Even if clinical control variants are unavailable, standards for judging functional data quality and reproducibility have been articulated and should be used.[19] Another challenge is how multiple datasets available for a gene should be integrated and how computational predictions, functional data, and other types of data can be combined. Here, comparison of multiple predictor and functional datasets could yield insights into the molecular and cellular mechanisms by which variants act, greatly improving our understanding of disease. However, model-based integration and weighting of predictor and functional datasets, with appropriate benchmarking using control variants, will be needed to account for the fact that different functional assays evaluate distinct processes (e.g., expression vs. splicing vs. protein activity) and to avoid double counting of datasets with the same information content. Lastly, a clinical-facing resource for discovering and assessing the quality of functional datasets from multiplexed assays is urgently needed. The ideal resource would give clinicians a variant-level view while also

highlighting the quality of the functional dataset and the strength of evidence generated.

## Data sharing

Most variation identified and interpreted in rare disease testing is unique to an individual and may never be observed again by a given laboratory, making the challenge of variant classification even harder. Indeed, 78% of variants submitted to ClinVar have only been submitted by one laboratory, primarily because they have been observed only once in a single individual. Analysis of the the Genome Aggregation Database (gnomAD) dataset suggests that every individual harbors on average 27 unique and 200 very rare variants in their coding sequencing alone.[23] Thus, finding additional evidence for these rare variants from other individuals in the population will require massive, widespread data sharing. This data sharing will need to come in multiple forms, including allowing genomic datasets from diverse populations to be aggregated and released to inform allele frequencies (AFs) of all variation across every population, aiding in the ability to rule out variation as causal based on subpopulation AFs too high to be disease causing. With tens to hundreds of millions of human genomes aggregated, we could also saturate the observation of most types of human genetic variation, allowing not only the elimination of causal variation through high frequency but also the ability to infer pathogenicity through the absence of variation. So far, we have only begun to approach saturation for single-nucleotide variants at highly mutable CpG sites[24] and will need much larger datasets to reach saturation of all variation types and in other genomic contexts. But with thousands of cohorts and biobanks being assembled across the globe, already totaling over 50 million enrolled participants as of 2020,[25] and standards being developed by the Global Alliance for Genomics and Health to enable interoperable data sharing,[26] we anticipate major progress in access to massive genomic datasets to inform the classification of genetic variation. And, while the genomic data alone can yield some insights into variant effects, most variants will still require additional evidence to classify. As such, all shared genomic datasets must be accompanied by clinical data, including phenotype, *de novo* occurrence, and segregation data, to enable variant classification. Efforts are underway to build federated variant-level querying capabilities to make it easy to access these data for a given rare variant, speeding the interpretation of monogenic-disease genetic testing and gene-disease discovery.[27]

## A realistic picture for 2030

We are hopeful that many, if not most, VUSs in coding regions will be resolved by 2030. Our hope is driven by the confluence of recent changes we outlined above, especially advances in the standards for variant classification that better leverage diverse types of evidence, improvements in computational variant effect predictor performance, scalable multiplexed assays of variant effect capable of saturating the genome, and data-sharing efforts that will maximize the information gained from each new individual sequenced and variant interpreted. Indeed, with high-quality predictions and functional data available for a gene, along with information from previously sequenced individuals, the overwhelming majority of single-nucleotide variants could be classified to a level that would enable a clinician to act on genetic information in the care of a patient, as well as identify borderline VUSs for which an incremental amount of additional evidence could push variants over to likely pathogenic. However, we acknowledge that progress is likely to be highly uneven across genes, even for single nucleotide variants. For a handful of genes, including *BRCA1* and *TP53*, all the pieces are in place, and we predict that VUS will largely be eliminated in the next few years. For other genes, perhaps on the order of several hundred, high-quality predictions exist and functional data are forthcoming, suggesting that VUSs will largely be resolved by 2030. For structural variation and intergenic variation, the picture is bleaker because effective methods are largely missing to make high-quality variant effect predictions or to collect multiplexed functional data.

Even for single-nucleotide variants within genes and their well-understood regulatory elements, progress depends on a variety of factors. First and foremost, the rate of progress will depend on the clinical community's willingness to revisit VUS and to engage with researchers generating computational variant effect predictors and collecting multiplexed functional data. High engagement will help the research community to focus on delivering the highest-value predictions and variant effect measurements for the genes where they can make the most difference. While such engagement is increasing, as reflected by the recent advent of the Atlas of Variant Effects Alliance,[21] NHGRI Impact of Genomic Variation on Function consortium (https://igvf.org/), and other scalable variant effect efforts, much more is needed to maximize the utility of research efforts. For research, more investment will be required to realize the goal of having high-quality predictions and variant effect measurements for every clinically relevant gene in the genome, nearly 5,000 to date (www.omim.org; accessed July 30, 2023), and to develop scalable methods capable of dealing with intergenic and structural variation. On the clinical side, resources are lacking for supporting efficient data sharing of primary evidence, as well as revisiting unsolved cases over time, neither of which are incentivized sufficiently.

Thus, for clinicians and researchers there exists a path to a world where VUSs are largely eliminated, in line with the

NHGRI's bold prediction. The pace at which we walk this path, and thus whether we are able to achieve the goal of largely eliminating VUSs by 2030, is largely a consequence of the choices made now and in the next few years. We believe that investing in eliminating VUSs is worthwhile, since their predominance remains one of the biggest challenges to precision genomic medicine.


## Acknowledgments

We thank Shawn Fayer for assistance in querying ClinVar.

H.L.R. was funded in part by the National Human Genome Research Institute under awards U24HG006834, U24HG011450, and U01HG011755. D.M.F. was funded in part by the National Human Genome Research Institute under awards RM1HG010461 and UM1HG011969 and the National Heart, Lung, and Blood Institute under award R01HL152066.


## Declaration of interests

H.L.R. is an associate editor for *AJHG*.

## References


1. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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. Genet. Med. *17*, 405–424.
2. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980–D985.
3. Rehm, H.L., Alaimo, J.T., Aradhya, S., Bayrak-Toydemir, P., Best, H., Brandon, R., Buchan, J.G., Chao, E.C., Chen, E., Clifford, J., et al. (2023). The landscape of reported VUS in multigene panel and genomic testing: Time for a change. Genet. Med. *25*, 100947.
4. Pejaver, V., Byrne, A.B., Feng, B.-J., Pagel, K.A., Mooney, S.D., Karchin, R., O'Donnell-Luria, A., Harrison, S.M., Tavtigian, S.V., Greenblatt, M.S., et al. (2022). Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am. J. Hum. Genet. *109*, 2163–2177.
5. Walker, L.C., Hoya, M.d.l., Wiggins, G.A.R., Lindy, A., Vincent, L.M., Parsons, M.T., Canson, D.M., Bis-Brewer, D., Cass, A., Tchourbanov, A., et al. (2023). Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup. Am. J. Hum. Genet. *110*, 1046–1067.
6. Bamshad, M.J., Nickerson, D.A., and Chong, J.X. (2019). Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am. J. Hum. Genet. *105*, 448–455.
7. Wu, Y., Li, R., Sun, S., Weile, J., and Roth, F.P. (2021). Improved pathogenicity prediction for rare human missense variants. Am. J. Hum. Genet. *108*, 1891–1906.
8. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. *99*, 877–885.
9. Gao, H., Hamp, T., Ede, J., Schraiber, J.G., McRae, J., Singer-Berk, M., Yang, Y., Dietrich, A.S.D., Fiziev, P.P., Kuderna, L.F.K., et al. (2023). The landscape of tolerated genetic variation in humans and primates. Science *380*, eabn8153.
10. Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J.K., Brock, K., Gal, Y., and Marks, D.S. (2021). Disease variant prediction with deep generative models of evolutionary data. Nature *599*, 91–95.
11. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492.
12. Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. *12*, 3.
13. Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat. Methods *7*, 741–746.
14. Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., and Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat. Biotechnol. *27*, 1173–1175.
15. Gasperini, M., Starita, L., and Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. Nat. Protoc. *11*, 1782–1787.
16. Weile, J., and Roth, F.P. (2018). Multiplexed assays of variant effects contribute to a growing genotype-phenotype atlas. Hum. Genet. *137*, 665–678.
17. Kinney, J.B., and McCandlish, D.M. (2019). Massively Parallel Assays and Quantitative Sequence-Function Relationships. Annu. Rev. Genomics Hum. Genet. *20*, 99–127.
18. Findlay, G.M. (2021). Linking genome variants to disease: scalable approaches to test the functional impact of human mutations. Hum. Mol. Genet. *30*, R187–R197.
19. Gelman, H., Dines, J.N., Berg, J., Berger, A.H., Brnich, S., Hisama, F.M., James, R.G., Rubin, A.F., Shendure, J., Shirts, B., et al. (2019). Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. Genome Med. *11*, 85.
20. Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., and Fowler, D.M. (2017). Variant Interpretation: Functional Assays to the Rescue. Am. J. Hum. Genet. *101*, 315–325.
21. Fowler, D.M., Adams, D.J., Gloyn, A.L., Hahn, W.C., Marks, D.S., Muffley, L.A., Neal, J.T., Roth, F.P., Rubin, A.F., Starita, L.M., and Hurles, M.E. (2023). An Atlas of Variant Effects to understand the genome at nucleotide resolution. Genome Biol. *24*, 147.
22. Fayer, S., Horton, C., Dines, J.N., Rubin, A.F., Richardson, M.E., McGoldrick, K., Hernandez, F., Pesaran, T., Karam, R., Shirts, B.H., et al. (2021). Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. Am. J. Hum. Genet. *108*, 2248–2258.

23. Gudmundsson, S., Singer-Berk, M., Watts, N.A., Phu, W., Goodrich, J.K., Solomonson, M., O'Donnell-Luria, A., Genome Aggregation Database Consortium, Rehm, H.L., and O'Donnell-Luria, A. (2022). Variant interpretation using population databases: Lessons from gnomAD. Hum. Mutat. *43*, 1012–1030.
24. 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.
25. Manolio, T.A., Goodhand, P., and Ginsburg, G. (2020). The International Hundred Thousand Plus Cohort Consortium: integrating large-scale cohorts to address global scientific challenges. Lancet. Digit. Health *2*, e567–e568.
26. Rehm, H.L., Page, A.J.H., Smith, L., Adams, J.B., Alterovitz, G., Babb, L.J., Barkley, M.P., Baudis, M., Beauvais, M.J.S., Beck, T., et al. (2021). GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genom. *1*, 100029.
27. Rodrigues, E.d.S., Griffith, S., Martin, R., Antonescu, C., Posey, J.E., Coban-Akdemir, Z., Jhangiani, S.N., Doheny, K.F., Lupski, J.R., Valle, D., et al. (2022). Variant-level matching for diagnosis and discovery: Challenges and opportunities. Hum. Mutat. *43*, 782–790.