

UCSF

UC San Francisco Previously Published Works

Title

An Atlas of Variant Effects to understand the genome at nucleotide resolution

Permalink

<https://escholarship.org/uc/item/2zn7v2wt>

Journal

Genome Biology, 24(1)

ISSN

1474-760X

Authors

Fowler, Douglas M

Adams, David J

Gloyn, Anna L

et al.

Publication Date

2023

DOI

10.1186/s13059-023-02986-x

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

CORRESPONDENCE

Open Access



An Atlas of Variant Effects to understand the genome at nucleotide resolution

Douglas M. Fowler^{1,2,3*} , David J. Adams⁴ , Anna L. Gloyn^{5*} , William C. Hahn^{6,7}, Debora S. Marks^{7,8} , Lara A. Muffley¹ , James T. Neal^{7,9} , Frederick P. Roth^{10,11} , Alan F. Rubin^{12,13} , Lea M. Starita^{1,2,3} and Matthew E. Hurles^{4*}

*Correspondence:
dfowler@uw.edu; agloyn@stanford.edu; meh@sanger.ac.uk

¹ Department of Genome Sciences, University of Washington, Seattle, WA, USA

⁴ Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK

⁵ Department of Pediatrics & Department of Genetics, Division of Endocrinology, Stanford School of Medicine, Stanford University, Stanford, CA, USA
Full list of author information is available at the end of the article

Abstract

Sequencing has revealed hundreds of millions of human genetic variants, and continued efforts will only add to this variant avalanche. Insufficient information exists to interpret the effects of most variants, limiting opportunities for precision medicine and comprehension of genome function. A solution lies in experimental assessment of the functional effect of variants, which can reveal their biological and clinical impact. However, variant effect assays have generally been undertaken reactively for individual variants only after and, in most cases long after, their first observation. Now, multiplexed assays of variant effect can characterise massive numbers of variants simultaneously, yielding variant effect maps that reveal the function of every possible single nucleotide change in a gene or regulatory element. Generating maps for every protein encoding gene and regulatory element in the human genome would create an 'Atlas' of variant effect maps and transform our understanding of genetics and usher in a new era of nucleotide-resolution functional knowledge of the genome. An Atlas would reveal the fundamental biology of the human genome, inform human evolution, empower the development and use of therapeutics and maximize the utility of genomics for diagnosing and treating disease. The Atlas of Variant Effects Alliance is an international collaborative group comprising hundreds of researchers, technologists and clinicians dedicated to realising an Atlas of Variant Effects to help deliver on the promise of genomics.

Keywords: Multiplexed assay of variant effect, Genome interpretation, Variant effect, Saturation mutagenesis, Functional genomics, Global alliance

Introduction

Two decades after sequencing the first human genome, millions of human exomes and genomes have been sequenced. Interpreting the effects of the hundreds of millions of variants thus discovered has become a central challenge for genomics. The genomes of the 8 billion people alive today collectively contain nearly all ~9 billion possible single nucleotide genetic variants compatible with life, as well as numerous



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

insertions, deletions and other types of variants [1, 2]. Moreover, within the trillions of cells of each individual, every possible single nucleotide genetic variant will have arisen through somatic mutation. The functional impact of genetic variants has primarily been determined by asking if the variant co-occurs with a disease, disorder or other trait, an approach which has collectively characterised the functional impact of less than 1% of genetic variation. Moreover, our knowledge of variant effects is focused on the best-understood 1–2% of our DNA—the genes that encode proteins. For non-coding variation, the situation is even less certain, because the location of most known non-coding functional elements has only been recently identified [3]. Moreover, non-coding elements are not as highly conserved and their functions are often cell type and development stage specific [4].

Our lack of information about the effect of variation found through genetic testing or genome sequencing is the major barrier to the use of sequence information for diagnosing genetic disease. This lack of information limits the effectiveness of genetic precision medicine and hinders our ability to understand genome function. Even when a variant in a well-annotated functional element is known to increase disease risk, the mechanism by which it does so is often unknown. A solution lies in our ability to assess the functional effect of variants using in vitro or cell-based assays, which can provide strong evidence to interpret their biological and clinical impact and can, in principle, be applied to any variant. However, owing to the resource- and time-intensive nature of traditional variant effect assays, they have generally been undertaken reactively for individual variants only after and, in most cases long after, the first observation of the variant. Now, multiplexed assays of variant effect (MAVEs) enable the generation of ‘variant effect maps’ characterising aspects of the function of every possible single nucleotide change in a gene or functional element of interest. Because variant effect maps are comprehensive, they profile all previously observed variants, as well as those that might be found in the future. Generating variant effect maps for every protein encoding gene and regulatory element in the human genome would create an ‘Atlas’ of variant effect maps that would transform our understanding of genetics by ushering in a new era of nucleotide-resolution functional knowledge of the genome.

The generation of an Atlas of Variant Effects (AVE) would have major impact across multiple areas of basic and translational research and, importantly, for clinical care. Any effort to determine whether a variant alters function would be transformed by having an Atlas, including in the following high impact areas (Fig. 1):

- **Precision genomic medicine.** Variant effect maps of functional elements known to harbour disease-causing variation can drive more accurate, rapid and inexpensive genetic diagnostic testing. Variant effect maps can also enhance our understanding of penetrance and variable expressivity and potentially even reveal compensatory genetic perturbations. For a wide variety of genetically driven disorders, knowledge of disease risk variants allows screening within families or even populations for early detection and thus early intervention [5].
- **Disease association studies.** Just as targeted variant functional assays have assisted discovery and validation of associations between specific rare genetic

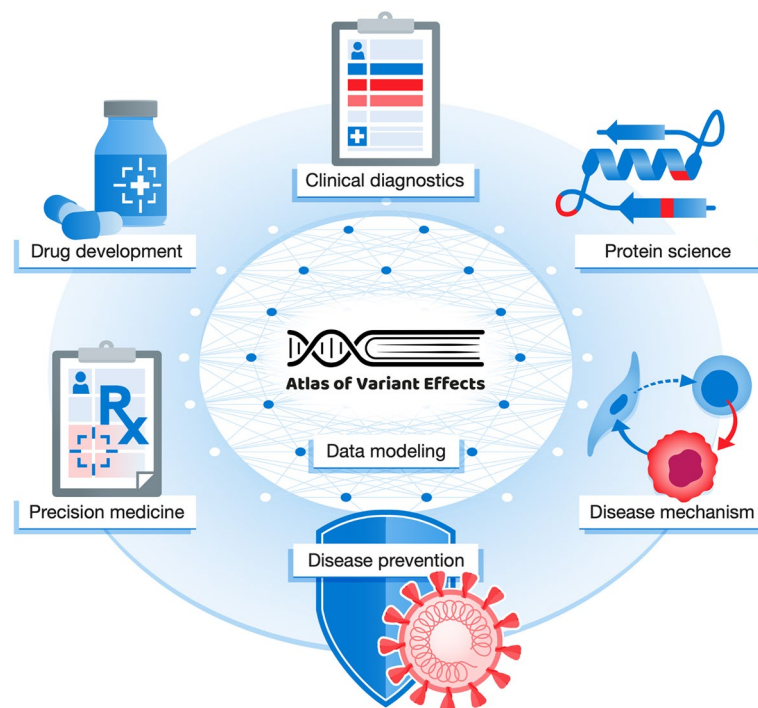


Fig. 1 Schematic representation of areas of high impact resulting from an Atlas of Variant Effects

variants and disease risk, variant effect maps can enable this approach broadly, at scale [6, 7].

- **Therapeutic development and pharmacogenetics.** Variant effect maps can shed light on disease mechanisms and may identify novel potential targets for drugs or other therapeutics [8], help predict the safety and efficacy of modulating specific targets, reveal routes of resistance and identify patients likely to respond favourably in clinical trials. Variant effect maps of pharmacogenes, where genetic variation can influence the activity or metabolism of drugs, could reveal the optimal dose for an individual or identify predispositions to adverse reactions. Variant effect maps could also enable the systematic study of genetic dose–response curves through functional and clinical correlations.
- **Sequence/structure/function relationships.** Understanding the relationship between sequence and function is fundamental to biology [9] and remains difficult to predict. Variant effect maps can illuminate this relationship, for example by improving or benchmarking computational variant effect prediction; revealing protein function, allostery or structure; and discerning the composition and mechanisms of regulatory elements [10–19].
- **Evolutionary genetics.** Differences in the biology of species, including those of commercial interest, is genomically encoded. Variant effect maps can highlight the subset of genetic differences between species that have functional consequences, probe inferred ancestral sequences [20] and improve phylogenetic inference [21–23].

- **Pathogen biology.** Genetic variation in pathogen genomes influences key characteristics of pathogen biology, including virulence, transmission, immune evasion and drug resistance. Variant effect maps can inform the surveillance of pathogen evolution [24] and provide opportunities to respond more rapidly, as well as revealing drug resistance and immune evasion variants [25].

By comprehensively capturing the impact of variants in functional elements throughout the genome, an Atlas of Variant Effects would accelerate and empower biological research, drug discovery and clinical practice. Systematic variant analysis, unbiased by allele frequency in any population, would empower equitable interpretation and reduce healthcare disparities [26]. Building and implementing a coherent Atlas of Variant Effects will necessarily be a collective endeavour, drawing together diverse expertise from different communities, including patients, patient advocates, researchers, clinicians, diagnostics companies and drug developers.

MAVEs can measure the effect of genetic variants at the scale necessary to compile an Atlas of Variant Effects

MAVEs are a rapidly growing family of methods that involve mutagenesis of a DNA-encoded protein or regulatory element followed by a multiplexed assay for some aspect of function [9, 27–29]. High-throughput DNA sequencing is used to read out each variant's effect in the assay (Fig. 2A). MAVEs encompass both assays of protein function, often called deep mutational scans, and of regulatory elements, often called massively parallel reporter assays. Early MAVEs were applied to small protein domains and short regulatory elements [14, 15, 30] generally querying single 'sub-functions' of an element such as promoter activity [14, 15], protein–ligand interactions [30–32] or stability [33, 34]. Other early efforts focused on the ability of an element to perform its overall cellular function in a cell-based growth assay [35]. Subsequently, MAVEs have been developed for a variety of functions and have been used to generate multiple variant effect maps examining different functions for the same element [9, 28, 36]. Now, MAVEs have been scaled up and optimised to enable routine application to entire genes, measuring the relative functional impact of tens of thousands of variants in a single controlled experiment.

To date, variant effect maps have been generated for hundreds of functional elements encompassing over 11 million total variants (Fig. 2B). However, existing variant effect maps cover < 1% of the known clinically relevant human genome and are largely focused on single nucleotide variants, as these are the type of variants most often encountered in current human genome sequencing and clinical testing. No functional element has been mapped in a diverse panel of cell types or across developmental stages. However, even at this very early stage in the development of a comprehensive Atlas of Variant Effects, multiplexed variant functional data are proving to be powerful. In particular, variant effect maps are beginning to reshape how human variants found in clinical genetic testing are interpreted and also to redefine our understanding of the mapping between DNA sequence and molecular, cellular and organismal phenotype.

The value of functional evidence for informing clinical variant interpretation is already well appreciated and has been incorporated within current professional guidelines for

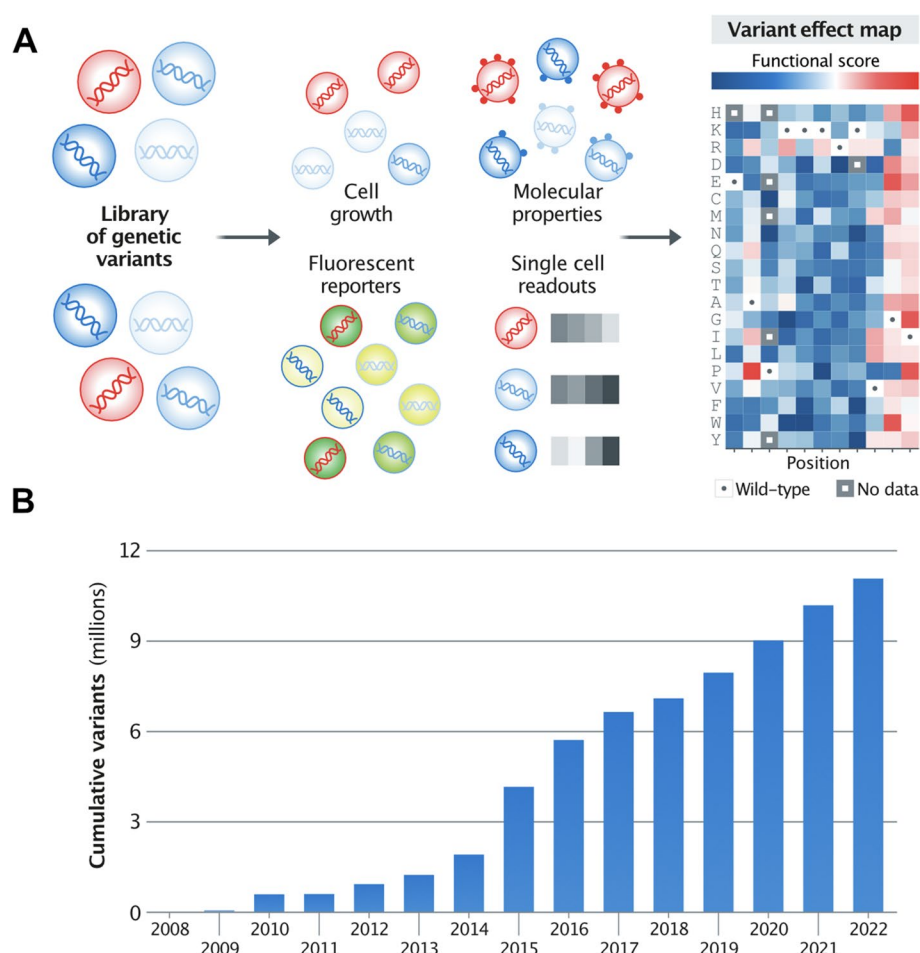


Fig. 2 **A** (top panel) MAVEs can measure a wide variety of protein and regulatory DNA functions, and they produce comprehensive variant effect maps representing the effects of nearly all possible nucleotide or amino acid variants in the scanned functional element. A variant effect map is shown for a small region of a protein-coding gene; each column in the map is a position in a gene and each row is an amino acid substitution. Tiles are coloured based on the measured effect of the variant. **B** (bottom panel) MAVEs have been applied to hundreds of functional elements and, collectively, ~11 million variant effect measurements have been made with MAVEs. Data available at [10.5281/zenodo.7662580](https://zenodo.org/record/7662580)

genetic diagnosis that are used internationally [37, 38]. MAVE-derived variant functional data has numerous advantages as compared to functional data derived from traditional, low-throughput assays. Unlike testing variants in small batches using different methods in different labs, MAVEs can determine the effects of thousands of variants simultaneously, not only improving reproducibility but allowing assessment of variants in the context of the functional effects of all of the variants in that gene, including the effects of known pathogenic and benign variants. Thus, MAVE-derived functional data can be used to eliminate many, if not most, of the uncertain, clinically observed variants in monogenic disease genes demonstrating the power of functional data to help deliver more definitive genetic test results to patients and clinicians [39–41].

Multiplexed variant functional data can also transform our understanding of how variants encode molecular and cellular function and how sequence dictates biological structure. For example, multiplexed measurements of variant abundance and ligand binding

in SH3 and PDZ domains, combined with a model, enabled a comprehensive accounting of allostery within each domain [16]. Multiplexed variant functional data can be used to validate proposed protein structures [17, 42] or, where variant combinations are assayed, even infer them de novo [12, 13]. Knowledge of the precise mechanism of variant effects opens the door for variant-guided therapies designed to ameliorate protein misfolding or aggregation, aberrant splicing and more.

Existing variant effect maps for human genes have been generated by a range of different technologies, from yeast complementation assays to CRISPR-based saturation genome editing in human cells. Each technology has specific advantages and disadvantages. For example, yeast complementation assays are only applicable to a minority of human genes [43] and would not be appropriate for identifying some variant effects, such as those that affect functions beyond those needed for complementation or those that disrupt splicing. CRISPR-based saturation genome editing of an endogenous locus is costly and practical only for growth-based assays. Thus, no single technology can currently be used to generate maps of variant effects for all functional elements. Indeed, even within a single gene, multiple assays may be required to assess different pathophysiological mechanisms. Current MAVEs require appreciable effort, and the time and cost needed to develop new assays can be considerable. Moreover, some variant effects may only be well-modelled in terminally differentiated cell types or in multicellular systems or by assaying variant effects on complex phenotypes like cell morphology or transcriptional state. Thus, the existing portfolio of MAVE technologies can be applied to a substantial fraction of the genome, but more technology development is required to achieve comprehensive coverage of genomic functional elements and to identify the mechanism by which most variants act.

The AVE Alliance provides international coordination to create, disseminate and implement an Atlas of Variant Effects

Compiling a complete Atlas of Variant Effects for all 20,000 human genes, not to mention potentially hundreds of thousands of noncoding regulatory elements, will require an international collaborative effort involving thousands of researchers, clinicians and technologists. Comparing this initiative to some of the landmark genomic collaborative achievements of the past 30 years highlights some of the key challenges to be addressed. The Human Genome Project (HGP) required a small number of centres generating data at unprecedented scales, in a highly coordinated and centralised fashion. By contrast, the Protein Data Bank (PDB) contains structures for thousands of human proteins, generated by thousands of researchers, in a largely uncoordinated and decentralised fashion [44]. Despite their differences, both HGP and PDB succeeded in generating an enduring and sustainable knowledge base and depended, crucially, on robust data standards, community-agreed quality metrics and centralised data deposition and dissemination. Moreover, a strong community ethos was essential for the development and adoption of these core standards and infrastructure. Some of the critical informatics infrastructure needed to support the AVE has already been developed, for example the MaveDB repository [45, 46], initial standards [47] for MAVE datasets and a MAVE project registry [48].

We envisage that the AVE will sit between the extremes exemplified by HGP and PDB, with a combination of a small number of centres generating variant effect maps at scale

using generalisable assays and a large number of laboratories generating small numbers of maps, using bespoke assays, leveraging their expertise in investigating particular genes and biological pathways. Integration of variant effect data for the same gene, generated using different MAVEs, will in some cases be required to achieve accurate and comprehensive characterisation of different functional effects [39, 49–52]. The computational prediction of variant effect maps using AI/ML methods will continue to improve and will leverage growing numbers of experimentally determined variant effect maps, analogous to the advances in computational prediction of protein structures based on thousands of experimentally determined protein structures (Fig. 3). With these expectations in mind, we can identify some of the key challenges that realising the AVE vision will face and some of the likely solutions on the critical path to success:

- **Diverse expertise.** Developing new experimental technologies that reflect the complexity of biology and disease, scaling existing technologies, processing and managing complex data, and translating knowledge into clinical benefits requires a broad range of expertise, interests and competencies, working collaboratively. No one centre or community will be able to create the AVE in isolation. Technology developers, geneticists, cell biologists, protein scientists, data scientists, software engineers, clinicians will need to work together, aligned around a common vision, language and values.
- **Technology development and scaling.** Generating variant effect maps for all 20,000 genes will require both the scaling of existing technologies that can be applied to many genes, and the development of new technologies that will extend coverage of

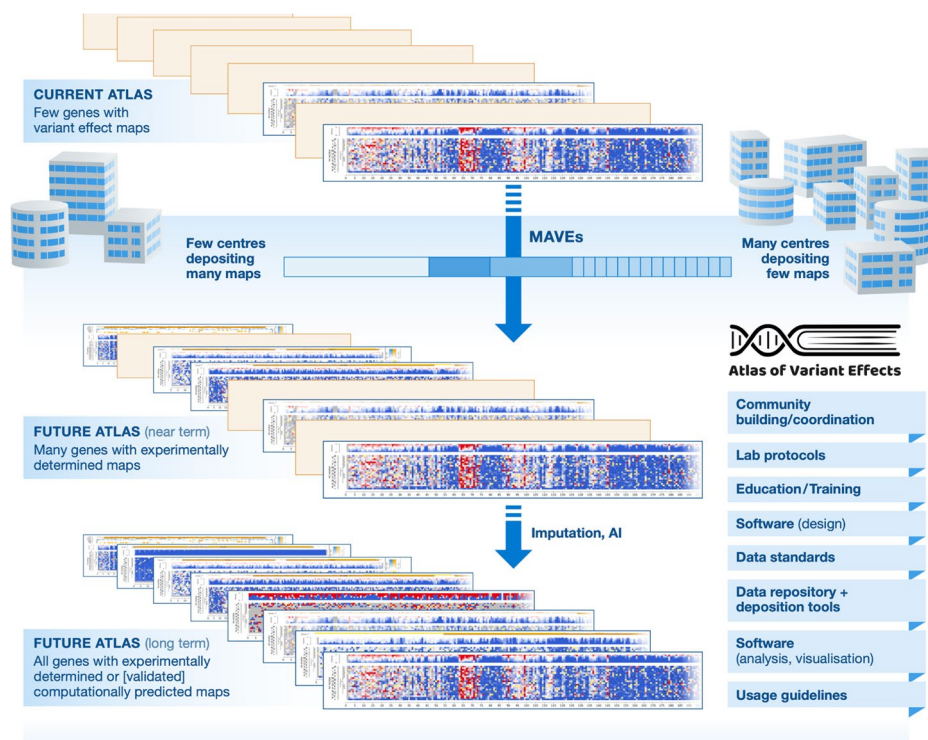


Fig. 3 Stages of Atlas of Variant Effects completion

MAVE-compatible assays to all functional elements. Moreover, new approaches will be needed to assess variant effects in more complex contexts, such as specific cell types or in development, and for more complex phenotypes, such as cell morphology and behaviour.

- **Democratisation of technology.** Completing the AVE will require a major expansion in the numbers of researchers and organisations actively performing MAVEs. Readily accessible training materials, protocols, experimental resources (e.g. cell lines, libraries) and easy-to-use and flexible software will all be crucial, as will advocacy and support to facilitate researchers with expertise in informative assays to adopt MAVE technologies.
- **Data standards and coordination.** Data standards, community-agreed quality standards, centralised data deposition, open dissemination and a FAIR ethos [53] are all necessary but not sufficient for compiling the Atlas of Variant Effects. The existing informatics infrastructure needs to evolve, become integrated into the wider clinical and biological data ecosystem and be actively sustained for long term impact. Moreover, community-wide adoption of best practices with regard to data and meta-data deposition are critical for data integration.
- **Ensuring trustworthy clinical adoption.** The potential clinical impact of the Atlas of Variant Effects can only be achieved through rigorous and clinician-trusted integration into diagnostic workflows. Co-development of quality standards and guidelines with clinical communities will help to build trust, as will starting conservatively. Integration with existing clinical decision support software (e.g. DECIPHER [54]) and data resources (e.g. ClinVar [55]), as opposed to requiring diagnosticians to use new systems, will facilitate rapid adoption.

To achieve the AVE vision and tackle these challenges, an international group of diverse researchers, clinicians and diagnosticians established the Atlas of Variant Effects Alliance (www.varianteffect.org). The AVE Alliance currently has over 400 members from over 100 institutions, located in 30 countries, united by the mission to bring the AVE into reality. The AVE Alliance is committed to Open Science and places diversity and inclusion at the heart of its activities. The AVE Alliance organises an annual meeting, the Mutational Scanning Symposium, and a monthly seminar series, the Variant Effect Seminar Series. To tackle the challenges identified above, AVE has established workstreams to:

- Develop, standardise and democratise experimental and computational technologies,
- Develop the infrastructure necessary to ingest, store and disseminate high quality FAIR data,
- Ensure that clinical benefits are realised,
- Expand, coordinate and sustain a diverse and motivated community.

The AVE Alliance provides a ‘front door’ for other organisations and initiatives to work with the diverse AVE community, from complementary large-scale national initiatives such as the NIH-funded Impact of Genomic Variants on Function (IGVF), as

well as research funders and commercial organisations who are keen to engage with the community as a whole. We welcome any and all readers who are interested in building and learning from the Atlas of Variant Effects to join the Alliance and get involved [56, 57].

Acknowledgements

We thank all members of the Atlas of Variant Effects Alliance for their work and contributions. Uta Mackensen helped prepare final figure graphics. We thank Carlos Araya for helpful input. We thank Alex Hopkins for administrative support. To fairly give credit, and for the purposes of PubMed, we list the following AVE Alliance contributing authors, who provided substantial comments and edits to this manuscript.

Atlas of Variant Effects Alliance contributing authors

Nadav Ahituv¹⁴, Orli G. Bahcal¹⁵, Dustin Baldridge¹⁶, Jonathan S. Berg¹⁷, Alice H. Berger¹⁸, Aisha Haley Bianchi¹⁹, Benedetta Bolognesi²⁰, Michael Boutros²¹, Steven Brenner²², Matthew H. Brush²³, Vanessa Bryant²⁴, Carol J. Bult²⁵, Martha Bulyk²⁶, Melissa Call²⁷, Hannah Carter²⁸, Melina Claussnitzer^{28,29}, Feng Chen³⁰, Melissa S. Cline³¹, Josh T. Cuperus¹, Moez Dawood³², Hannah N. De Jong³³, Mafalda Dias³⁴, Michael Dunn⁵, Jesse Engreitz³⁵, Kyle Farh³⁰, Phillip G. Febbo³⁰, Stanley Fields¹, Gregory M. Findlay³⁶, Helen Firth³⁷, James S. Fraser³⁸, Jonathan Frazer³⁴, Mattia Frontini³⁹, Irene Gallego Romero⁴⁰, Andrew M. Glazer⁴¹, Murat Güler²¹, Rasmus Hartmann-Petersen⁴², Richard Houlston⁴³, Kuan-lin Huang⁴⁴, Carolyn M. Hutter⁴⁵, Sujatha Jagannathan^{46,47}, Richard G. James⁴⁸, Martin Kampmann^{49,50}, Rachel Karchin⁵¹, Justin B. Kinney⁵², Alexis C. Komor⁵³, Sriam Kosuri⁵⁴, Ben Lehner^{5,34,55,56}, Kresten Lindorff-Larsen⁴², Zané Lombard⁵⁷, Daniel G. MacArthur⁵⁸, Maria Martin⁵⁹, Ultan McDermott⁶⁰, Shannon M. McNulty⁶¹, Alex N. Nguyen Ba⁶², Anne O'Donnell-Luria^{63,64}, Brian J. O'Roak⁶⁵, Victoria N. Parikh⁶⁶, Leopold Parts⁵, Michael J. Pazin⁴⁵, Tina Pesaran⁶⁷, Slavé Petrovski⁶⁸, Christine Queitsch^{1,3}, David E. Root⁸, Jay Shendure^{1,3}, Amanda B. Spurdle⁶⁹, Kevin L. Taylor⁷⁰, Clare Turnbull⁴³, Judit Villén¹, L.E.L.M. Vissers⁷¹, Alex H. Wagner^{72,73}, Matthew J. Wakefield⁷⁴, Jochen Weile¹⁰, Jenny Xiao⁷⁵

¹⁴ Department of Bioengineering and Therapeutic Sciences and Institute for Human Genetics, University of California San Francisco USA

¹⁵ Cell Genomics, New York, NY USA

¹⁶ Washington University School of Medicine, St. Louis, MO USA

¹⁷ The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹⁸ Fred Hutchinson Cancer Center, Seattle, WA USA

¹⁹ National Institute of Aging, Baltimore, Maryland USA

²⁰ Institute for Bioengineering of Catalunya (IBEC) and The Barcelona Institute for Science and Technology (BIST), Barcelona, Spain

²¹ German Cancer Research Center (DKFZ) and Heidelberg University, Germany

²² University of California, Berkeley, Berkeley, CA USA

²³ Department of Biomedical Informatics, University of Colorado, Aurora, CO USA

²⁴ The Walter and Eliza Hall Institute of Medical Research/Department of Medical Biology, University of Melbourne/Dept Clinical Immunology, Royal Melbourne Hospital, Australia

²⁵ The Jackson Laboratory/Mouse Genome Informatics (MGI) consortium

²⁶ Brigham & Women's Hospital and Harvard Medical School, USA

²⁷ The Walter and Eliza Hall Institute of Medical Research/Department of Medical Biology, University of Melbourne, Australia

²⁸ Department of Medicine, University of California San Diego, La Jolla CA USA

²⁹ The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease USA

³⁰ Illumina USA

³¹ UC Santa Cruz Genomics Institute, Santa Cruz, CA USA

³² Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

³³ Department of Genetics, Stanford University School of Medicine, Stanford, CA USA

³⁴ Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST) and Universitat Pompeu Fabra (UPF), Barcelona, Spain

³⁵ Department of Genetics, Stanford University School of Medicine, CA USA

³⁶ The Francis Crick Institute, London, UK

³⁷ Sanger/Cambridge Hospitals UK

³⁸ Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA USA

³⁹ Department of Clinical and Biomedical Sciences, University of Exeter Medical School, Faculty of Health and Life Sciences, RILD Building, Barrack Road, Exeter, EX2 5DW.

⁴⁰ School of BioSciences, University of Melbourne, Parkville, Australia

⁴¹ Department of Medicine, Vanderbilt University Medical Center, Nashville, TN USA

⁴² Department of Biology, University of Copenhagen, Copenhagen, Denmark

⁴³ Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey UK

⁴⁴ Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY USA

⁴⁵ Division of Genome Sciences, NHGRI, Bethesda, MD, USA

⁴⁶ Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

⁴⁷ RNA Bioscience Initiative, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

⁴⁸ Seattle Children's Research Institute, Seattle WA USA

⁴⁹ Department of Biochemistry & Biophysics, University of California, San Francisco CA USA

⁵⁰ Institute for Neurodegenerative Diseases, University of California, San Francisco CA USA

⁵¹ Johns Hopkins University, Baltimore, Maryland USA

⁵² Cold Spring Harbor Laboratory, NY USA

- ⁵³ Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA USA
- ⁵⁴ Octant Inc USA
- ⁵⁵ University Pompeu Fabra (UPF), Barcelona, Spain
- ⁵⁶ Institutió Catalana de Recerca i estudis Avançats (ICREA), Barcelona, Spain
- ⁵⁷ Division of Human Genetics, National Health Laboratory Service, and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
- ⁵⁸ Centre for Population Genomics, Garvan Institute of Medical Research and Murdoch Children's Research Institute, Australia
- ⁵⁹ European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome, Hinxton, UK
- ⁶⁰ R&D Oncology, AstraZeneca UK
- ⁶¹ Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC USA
- ⁶² Department of Biology, University of Toronto, Toronto, Canada
- ⁶³ Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA USA
- ⁶⁴ Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA USA
- ⁶⁵ Oregon Health & Science University, Portland, OR USA
- ⁶⁶ Stanford Center for Inherited Cardiovascular Disease, Stanford School of Medicine, CA USA
- ⁶⁷ Ambry Genetics, Aliso Viejo, CA USA
- ⁶⁸ Centre for Genomics Research, Discovery Sciences, R&D, AstraZeneca UK
- ⁶⁹ QIMR Berghofer Medical Research Institute, Brisbane, Australia
- ⁷⁰ Proteogenomics, BioLegend USA
- ⁷¹ Department Human Genetics, Radboud University, Nijmegen, NL
- ⁷² The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH 43215, USA
- ⁷³ Departments of Pediatrics and Biomedical Informatics, The Ohio State University College of Medicine, Columbus, OH 43210, USA
- ⁷⁴ The Walter and Eliza Hall Institute, Parkville, Vic, Australia & Department of Obstetrics and Gynaecology, The University of Melbourne, Australia
- ⁷⁵ Guardant Health, Palo Alto USA
- ⁷⁶ Center for Genomic Medicine Massachusetts General Hospital, Harvard Medical School USA

Authors' contributions

D.M.F., A.L.G. and M.E.H. contributed to the conceptualization and writing of the original draft. The remaining co-authors contributed to the writing of the original draft. Contributing authors listed in the acknowledgements reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

F.P.R. acknowledges support from the NIH/NHGRI Impact of Genomic Variation on Function (IGVF) Initiative (HG011989) and from an NIH/NHLBI R01 grant (HL164675) and from a Canadian Institutes of Health Research Foundation Grant. L.M.S., L.A.M., D.M.F. and A.F.R. acknowledge support from NIH/NHGRI Impact of Genomic Variation on Function (IGVF) Initiative (HG011969). L.M.S., L.A.M., D.M.F., A.F.R. and F.P.R. all receive support from the NIH/NHGRI Center of Excellence in Genomic Science (HG010461). DMF also receive support from R01HL152066. L.M.S. is also supported by the Brotman Baty Institute. A.L.G. is a Wellcome Trust Senior Fellow (200837/Z/16/Z) and is also supported by NIDDK (UM-1DK126185). W.C.H. acknowledges support from NIH/NCI U01CA176058. J.T.N. acknowledges support from the Novo Nordisk Foundation (NNF21SA0072102) and from an NIH DP2 grant (1DP2GM146252). D.J.A. is supported by Cancer Research UK (CG-MAVE: EDDPGM-Nov22/100004) and the Wellcome Trust. A.F.R. received grant funding from the Australian Government. D.S.M. acknowledges support from Chan Zuckerberg Initiative CZI2018- 191853. and NIH TR01 grant (1R01CA260415).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

A.L.G. declares that her spouse is an employee of Genentech and holds stock options in Roche. D.J.A. is a consultant for Microbionics and Astra Zeneca. D.S.M. is a consultant for Insitro, Dyno and Octant. J.T.N. receives research support from Bristol Myers Squibb. F.P.R. holds shares in Ranomics, Inc., and is an investor and advisor for SeqWell, Inc. and Constantiam Biosciences, Inc. L.M.S. is a consultant for Nostos Genomics. W.C.H. is a consultant for Thermo Fisher, Solasta Ventures, MPM Capital, Tyra Biosciences, Frontier Medicines, Jubilant Therapeutics, KSQ Therapeutics, RAPPTA Therapeutics, Serinus Biosciences, Hexagon Bio, Function Oncolog, Riva Therapeutics, and Calyx. M.E.H. is a consultant for AstraZeneca and co-founder, director, shareholder of Congenica Ltd.

Author details

¹Department of Genome Sciences, University of Washington, Seattle, WA, USA. ²Department of Bioengineering, University of Washington, Seattle, WA, USA. ³Brotman Baty Institute for Precision Medicine, Seattle, WA, USA. ⁴Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK. ⁵Department of Pediatrics & Department of Genetics, Division of Endocrinology, Stanford School of Medicine, Stanford University, Stanford, CA, USA. ⁶Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ⁷Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁸Department of Systems Biology, Harvard Medical School, Cambridge, USA. ⁹Novo Nordisk Foundation Center for Genomic Mechanisms of Disease

at Broad Institute, Cambridge, MA, USA. ¹⁰Donnelly Centre and Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON, Canada. ¹¹Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada. ¹²Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia. ¹³Department of Medical Biology, University of Melbourne, Melbourne, VIC, Australia.

Received: 2 June 2023 Accepted: 13 June 2023

Published online: 03 July 2023

References

- Shirts BH, Pritchard CC, Walsh T. Family-specific variants and the limits of human genetics. *Trends Mol Med*. 2016;22:925–34.
- Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet*. 2001;27:234–6.
- Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilieny M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
- ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699–710.
- Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15:565–74.
- Schiabor Barrett KM, Masnick M, Hatchell KE, Savatt JM, Banet N, Buchanan A, et al. Clinical validation of genomic functional screen data: analysis of observed BRCA1 variants in an unselected population cohort. *HGG Adv*. 2022;3:100086.
- Dorling L, Carvalho S, Allen J, Parsons MT, Fortuno C, González-Neira A, et al. Breast cancer risks associated with missense variants in breast cancer susceptibility genes. *Genome Med*. 2022;14:51.
- Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–60.
- Kinney JB, McCandlish DM. Massively parallel assays and quantitative sequence-function relationships. *Annu Rev Genomics Hum Genet*. 2019;20:99–127.
- Gray VE, Hause RJ, Luebeck J, Shendure J, Fowler DM. Quantitative missense variant effect prediction using large-scale mutagenesis data. *Cell Syst*. 2018;6:116–24.e3.
- Riesselman AJ, Ingraham JB, Marks DS. Deep generative models of genetic variation capture the effects of mutations. *Nat Methods*. 2018;15:816–22.
- Rollins NJ, Brock KP, Poelwijk FJ, Stiffler MA, Gauthier NP, Sander C, et al. Inferring protein 3D structure from deep mutation scans. *Nat Genet*. 2019;51:1170–6.
- Schmiedel JM, Lehner B. Determining protein structures using deep mutagenesis. *Nat Genet*. 2019;51:1177–86.
- Kinney JB, Murugan A, Callan CG Jr, Cox EC. Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci U S A*. 2010;107:9158–63.
- Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol*. 2009;27:1173–5.
- Faure AJ, Domingo J, Schmiedel JM, Hidalgo-Carcedo C, Diss G, Lehner B. Mapping the energetic and allosteric landscapes of protein binding domains. *Nature*. 2022;604:175–83.
- Chiasson MA, Rollins NJ, Stephany JJ, Sitko KA, Matreyek KA, Verby M, et al. Multiplexed measurement of variant abundance and activity reveals VKOR topology, active site and human variant impact. *Elife*. 2020;9:<https://doi.org/10.7554/eLife.58026>.
- Livesey BJ, Marsh JA. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Mol Syst Biol*. 2020;16:e9380.
- Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature*. 2021;599:91–5.
- Starr TN, Picton LK, Thornton JW. Alternative evolutionary histories in the sequence space of an ancient protein. *Nature*. 2017;549:409–13.
- Klein JC, Keith A, Agarwal V, Durham T, Shendure J. Functional characterization of enhancer evolution in the primate lineage. *Genome Biol*. 2018;19:99.
- Bloom JD. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol*. 2014;31:1956–78.
- Gallego Romero I, Lea AJ. Leveraging massively parallel reporter assays for evolutionary questions. *Genome Biol*. 2023;24:26.
- Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, et al. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc Natl Acad Sci U S A*. 2018;115:E8276–85.
- Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 β -lactamase. *Cell*. 2015;160:882–92.
- Wright CF, Campbell P, Eberhardt RY, Aitken S, Perrett D, Brent S, et al. Optimising diagnostic yield in highly penetrant genomic disease. *bioRxiv*. 2022. Available from: <https://www.medrxiv.org/content/10.1101/2022.07.25.22278008v1>.
- Tabet D, Parikh V, Mali P, Roth FP, Claussnitzer M. Scalable functional assays for the interpretation of human genetic variation. *Annu Rev Genet*. 2022;56:441–65.
- Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant interpretation: functional assays to the rescue. *Am J Hum Genet*. 2017;101:315–25.
- Gasperini M, Starita L, Shendure J. The power of multiplexed functional analysis of genetic variants. *Nat Protoc*. 2016;11:1782–7.

30. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. *Nat Methods*. 2010;7:741–6.
31. Zhang H, Torkamani A, Jones TM, Ruiz DI, Pons J, Lerner RA. Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems. *Proc Natl Acad Sci U S A*. 2011;108:13456–61.
32. Ernst A, Gfeller D, Kan Z, Seshagiri S, Kim PM, Bader GD, et al. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst*. 2010;6:1782–90.
33. Kim I, Miller CR, Young DL, Fields S. High-throughput analysis of in vivo protein stability. *Mol Cell Proteomics*. 2013;12:3370–8.
34. Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc Natl Acad Sci U S A*. 2012;109:16858–63.
35. Hietpas RT, Jensen JD, Bolon DNA. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci Natl Acad Sci*. 2011;108:7896–901.
36. Weile J, Roth FP. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. *Hum Genet*. 2018;137:665–78.
37. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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. *Genet Med*. 2015;17:405–24.
38. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med*. 2019;12:3.
39. Fayer S, Horton C, Dines JN, Rubin AF, Richardson ME, McGoldrick K, et al. Closing the gap: systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. *Am J Hum Genet*. 2021;108:2248–58.
40. Radford EJ, Tan HK, Andersson MHL, Stephenson JD, Gardner EJ, Ironfield H, et al. Saturation genome editing of DDX3X clarifies pathogenicity of germline and somatic variation. *medRxiv* [Internet]. Cold Spring Harbor Laboratory Press; 2022; Available from: <https://www.medrxiv.org/content/10.1101/2022.06.10.22276179v1>.
41. Scott A, Hernandez F, Chamberlin A, Smith C, Karam R, Kitzman JO. Saturation-scale functional evidence supports clinical variant interpretation in Lynch syndrome. *Genome Biol*. 2022;23:266.
42. Adkar BV, Tripathi A, Sahoo A, Bajaj K, Goswami D, Chakrabarti P, et al. Protein model discrimination using mutational sensitivity derived from deep sequencing. *Structure*. 2012;20:371–81.
43. Kachroo AH, Laurent JM, Yellman CM, Meyer AG, Wilke CO, Marcotte EM. Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science*. 2015;348:921–5.
44. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res*. 2019;47:D520–8.
45. Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, et al. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol*. 2019;20:223.
46. Rubin AF, Min JK, Rollins NJ, Da EY, Esposito D, Harrington M, et al. MaveDB v2: a curated community database with over three million variant effects from multiplexed functional assays. *bioRxiv*. 2022. p. 2021.11.29.470445. Available from: <https://www.biorxiv.org/content/10.1101/2021.11.29.470445v2>. [Cited 2022 Dec 5].
47. Gelman H, Dines JN, Berg J, Berger AH, Brnich S, Hisama FM, et al. Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. *Genome Med*. 2019;11:85.
48. Kuang D, Weile J, Kishore N, Rubin AF, Fields S, Fowler DM, et al. MaveRegistry: a collaboration platform for multiplexed assays of variant effect. *Bioinformatics*. 2021;37:3382–3.
49. Mighell TL, Thacker S, Fombonne E, Eng C, O’Roak BJ. An integrated deep-mutational-scanning approach provides clinical insights on PTEN genotype-phenotype relationships. *Am J Hum Genet*. 2020;106:818–29.
50. Suiter CC, Moriyama T, Matreyek KA, Yang W, Scaletti ER, Nishii R, et al. Massively parallel variant characterization identifies *NUDT15* alleles associated with thiopurine toxicity. *Proc Natl Acad Sci U S A*. 2020;117:5394–401.
51. Jepsen MM, Fowler DM, Hartmann-Petersen R, Stein A, Lindorff-Larsen K. Chapter 5 - Classifying disease-associated variants using measures of protein activity and stability. In: Pey AL, editor. *Protein Homeostasis Diseases*. Academic Press; 2020. p. 91–107. <https://doi.org/10.1101/688234>, <https://www.biorxiv.org/content/10.1101/688234v2.full.pdf>.
52. Cagiada M, Johansson KE, Valanciute A, Nielsen SV, Hartmann-Petersen R, Yang JJ, et al. Understanding the origins of loss of protein function by analyzing the effects of thousands of variants on activity and abundance. *Mol Biol Evol*. 2021;38:3235–46.
53. Wilkinson MD, Dumontier M, Jan Aalbersberg I, Appleton G, Axton M, Baak A, et al. Addendum: the FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2019;6:6.
54. DECIPHER v1.1.16: Mapping the clinical genome. Available from: <http://www.deciphergenomics.org>. [Cited 2022 Dec 3].
55. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980–5.
56. Atlas of variant effects alliance. Atlas of Variant Effects Alliance. Available from: <http://www.varianteffect.org>. [Cited 2022 Dec 3].
57. AVE Alliance Founding Members. The Atlas of Variant Effects (AVE) Alliance: understanding genetic variation at nucleotide resolution. Zenodo; 2021. Available from: <https://zenodo.org/record/4989960>.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.