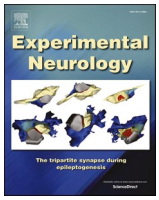
Experimental Neurology 382 (2024) 114978

Contents lists available at ScienceDirect

Experimental Neurology

journal homepage: www.elsevier.com/locate/yexnr



Research paper

The GENESIS database and tools: A decade of discovery in Mendelian genomics

Matt C. Danzi [a,*], Eric Powell [a,*], Adriana P. Rebelo [a], Maike F. Dohrn [a,b], Danique Beijer [a], Sarah Fazal [a], Isaac R.L. Xu [a], Jessica Medina [a], Sitong Chen [a], Yeisha Arcia de Jesus [a], Jacquelyn Schatzman [a], Ray E. Hershberger [c], Mario Saporta [a,aa], Jonathan Baets [d], Marni Falk [e,f], David N. Herrmann [g], Steven S. Scherer [h], Mary M. Reilly [i], Andrea Cortese [i,j], Wilson Marques [k], Mario R. Carnejo-Olivas [l], Oranee Sanmaneechai [m], Marina L. Kennerson [n], Albena Jordanova [o,p], Thiago Y.T. Silva [q], Jose Luiz Pedroso [q], Luca Schierbaum [r], Darius Ebrahimi-Fakhari [r], Stojan Peric [s], Yi-Chung Lee [t], Matthis Synofzik [u,v], Mustafa Tekin [a], Gianina Ravenscroft [w], Mike Shy [x], Nazli Basak [y], Rebecca Schule [z,ab], Stephan Zuchner [a,*]

[a] *Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA*
[b] *Department of Neurology, Medical Faculty of the RWTH Aachen University, Aachen, Germany*
[c] *Divisions of Human Genetics and Cardiovascular Medicine, Department of Internal Medicine, and the Davis Heart and Lung Research Institute, The Ohio State University, Columbus, OH, USA*
[d] *Translational Neurosciences, Faculty of Medicine and Health Sciences and Born-Bunge Institute, University of Antwerp, Antwerp, Belgium; Neuromuscular Reference Center, Department of Neurology, Antwerp University Hospital, Antwerp, Belgium*
[e] *Mitochondrial Medicine Frontier Program, Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA*
[f] *Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA*
[g] *Department of Neurology, University of Rochester Medical Center, Rochester, New York, USA*
[h] *Department of Neurology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA*
[i] *Centre for Neuromuscular Diseases, Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK*
[j] *Department of Brain and Behaviour Sciences, University of Pavia, Pavia, Italy*
[k] *Department of Neurology, School of Medicine of Ribeirão Preto, University of São Paulo, 2650 Ribeirão Preto, Brazil*
[l] *Neurogenetics Research Center, Instituto Nacional de Ciencias Neurologicas, Lima 15003, Peru*
[m] *Department of Pediatrics, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*
[n] *ANZAC Research Institute, Sydney Local Health District, Concord, NSW 2139 and School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Sydney, NSW 2050, Australia*
[o] *Molecular Neurogenomics Group, VIB-UAntwerp Center for Molecular Neurology and Department of Biomedical Sciences, University of Antwerp, Antwerpen 2610, Belgium*
[p] *Molecular Medicine Center Department of Medical Chemistry and Biochemistry, Medical University-Sofia, Sofia 1431, Bulgaria*
[q] *Department of Neurology, Ataxia Unit, Universidade Federal de São Paulo, São Paulo, Brazil*
[r] *Movement Disorders Program, Department of Neurology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA*
[s] *Faculty of Medicine, University of Belgrade, Dr Subotica 6, Belgrade, Serbia*
[t] *Department of Neurology, National Yang Ming Chiao Tung University, Taipei, Taiwan*
[u] *Division of Translational Genomics of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Tübingen, Germany*
[v] *German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany*
[w] *Centre for Medical Research, University of Western Australia and Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia*
[x] *Department of Neurology, University of Iowa Carver College of Medicine, Iowa City, IA, USA*
[y] *Koç University, School of Medicine, Suna and İnan Kıraç Foundation, Neurodegeneration Research Laboratory (NDAL), Research Center for Translational Medicine, 34010 Istanbul, Turkey*
[z] *Center for Neurology and Hertie Institute for Clinical Brain Research (HIH), University of Tübingen, Tübingen, Germany*
[aa] *Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA*
[ab] *Division of Neurodegenerative Diseases, Department of Neurology, Heidelberg University Hospital and Faculty of Medicine, Heidelberg, Germany*

* Corresponding authors at: John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Biomedical Research Building (BRB), 1501 NW 10th Avenue, Miami, FL 33136, USA.
E-mail address: szuchner@med.miami.edu (S. Zuchner).

https://doi.org/10.1016/j.expneurol.2024.114978
Received 25 May 2024; Received in revised form 25 September 2024; Accepted 27 September 2024
Available online 30 September 2024

ARTICLE INFO

Keywords:
Data sharing
Monogenic diseases
Data aggregation
Genome sequencing
Neuromuscular diseases

ABSTRACT

In the past decade, human genetics research saw an acceleration of disease gene discovery and further dissection of the genetic architectures of many disorders. Much of this progress was enabled via data aggregation projects, collaborative data sharing among researchers, and the adoption of sophisticated and standardized bioinformatics analyses pipelines. In 2012, we launched the GENESIS platform, formerly known as GEM.app, with the aims to 1) empower clinical and basic researchers without bioinformatics expertise to analyze and explore genome level data and 2) facilitate the detection of novel pathogenic variation and novel disease genes by leveraging data aggregation and genetic matchmaking. The GENESIS database has grown to over 20,000 datasets from rare disease patients, which were provided by multiple academic research consortia and many individual investigators. Some of the largest global collections of genome-level data are available for Charcot-Marie-Tooth disease, hereditary spastic paraplegia, and cerebellar ataxia. A number of rare disease consortia and networks are archiving their data in this database. Over the past decade, more than 1500 scientists have registered and used this resource and published over 200 papers on gene and variant identifications, which garnered >6000 citations. GENESIS has supported >100 gene discoveries and contributed to approximately half of all gene identifications in the fields of inherited peripheral neuropathies and spastic paraplegia in this time frame. Many diagnostic odysseys of rare disease patients have been resolved. The concept of genomes-to-therapy has borne out for a number of such discoveries that let to rapid clinical trials and expedited natural history studies. This marks GENESIS as one of the most impactful data aggregation initiatives in rare monogenic diseases.

1. Introduction

An estimated 300 million individuals worldwide suffer from rare inherited diseases (Nguengang Wakap et al., 2020). Genetic testing plays an important role in the diagnosis of inherited rare conditions. Such genetic diagnoses are essential for patients to participate in the many gene-specific treatment developments (Adams et al., 2018; Balwani et al., 2020; Finkel et al., 2017; van der Ploeg and Reuser, 2008). Yet, a significant number of patients with rare diseases have not yet received a genetic diagnosis, exposing a large diagnostic gap. A recent study on UK biobank data reported a diagnostic success rate of 16 % on 7065 rare disease patients (Turro et al., 2020). What medical genetics professionals view as a diagnostic gap, patients often experience as a diagnostic odyssey with an average of 19 years from symptom onset to obtaining a genetic diagnosis in some studies (Schuermans et al., 2022). One of the most impactful facilitators in solving rare disease cases is sequencing multiple family members and comparing their genotypes with those of many other patients afflicted by the same symptoms. This is particularly important for dominant diseases, adult-onset diseases and for conditions where there is reduced penetrance, as these factors make variant curation more difficult.

Declining sequencing costs, targeted gene panels, exomes, and both short- and long-read whole genomes are now fully implemented into routine clinical genetic work-up strategies. Ideally, once the clinical genetic workup has concluded without a definite diagnosis, data should be made available to scientists for secondary use in genetic research studies. However, the further mining of such research data requires bioinformatic expertise and many scientists have limited skills and resources to explore large genomic data collections. In order to lower such barriers to analyzing unsolved cases, we developed the GENESIS database and tools (Gonzalez et al., 2015, Gonzalez et al., 2013). GENESIS is a web-based platform that allows geneticists and scientists of all skill levels to analyze genomic sequencing data from individual research patients to large cohorts of thousands of participants, applying state-of-the-art tools and annotations. This has facilitated the discovery of disease genes, clarified variants of uncertain significance (VUS), allowed for genetic matchmaking, and catalogued allelic series for specific genes and phenotypes.

2. Results

2.1. The impact of GENESIS on rare disease research

2.1.1. A comprehensive analysis tool for non-bioinformatics scientists

The GENESIS database and tools were designed to simplify the complex task of analyzing human genome-scale variant data, where each patient typically carries hundreds of variants of interest. Given the large cohorts of unsolved patients with related phenotypes, statistically meaningful cohort comparisons and enrichment of rare alleles facilitates disease gene discovery. GENESIS allows for direct analysis of patient genomes using a web-based point-and-click graphical user interface (Supplementary Fig. 1). It facilitates analyses of genomic single nucleotide variations (SNVs), small insertions or deletions (InDels), structural variants (SVs), tandem repeats (TRs), and mitochondrial DNA changes (Supplementary Figs. 1 and 2) across targeted panels, exomes, short-read genomes, and long-read genomes. GENESIS contains tools to conduct patient or single-family analysis as well as unlimited cohort queries for allele-level and gene-level matchmaking (Fig. 1A,B; Supplementary Fig. 1F). Allele-level matchmaking is the identification of specific variants that are only present in patients with the same disease phenotype (Fig. 1A). Gene-level matchmaking is the pursuit of multiple different variants (or variant pairs in the compound heterozygous recessive case) within a gene that occur in multiple different patients with the same disease phenotype and are not observed in healthy individuals or individuals with other phenotypes (Fig. 1B). Such queries can also be computed as allele or gene enrichment analyses to detect incomplete penetrance or a spectrum of phenotypes. As new data is added on a weekly basis, the outcome of queries continuously changes since all queries are performed in real time rather than being precomputed.

2.1.2. Benefits of a shared, large international dataset

Allele-level and gene-level cohort studies both benefit from access to large numbers of patients and healthy controls. Currently, GENESIS has approximately 20,000 datasets available from individuals with dozens of disease phenotypes, where the ten most common phenotypes account for 92 % of patient datasets (Fig. 1C). Scientists from all continents have contributed deidentified/pseudonymized genomic data resulting in deep ancestral diversity of the database. Unique example are genomes from endogenous people from the Amazon rain forest, a large set of Taiwanese and South Korean data, as well as samples from Sudan, Mali and the Middle East. This highly diverse dataset serves as its own cross-reference for rare variants in any query. As much as gnomAD provides access to rare alleles in healthy individuals, GENESIS shares variants in genomes from solved and unsolved rare disease patients. The data in GENESIS are openly available when determined by the data owner. It is also increasingly expected that any data will become available, further anonymized, after a grace period of 12 months. To protect the research aspect of this database, access is only provided to scientists verifiably working at academic institutions.

2.1.3. Supporting consortia and large networks of rare diseases

The world-wide largest data collections for Charcot-Marie-Tooth (CMT) patients, hereditary spastic paraplegia (HSP), and rare ataxia cases are stored in the GENESIS database. Other large collections include hearing loss, amyotrophic lateral sclerosis (ALS), and cardiomyopathies. Major consortia in these fields of research have contributed thousands of patients to GENESIS and granted access to their team members to expedite secondary analysis of cases. This collaboration has resulted in the discovery of new CMT, HSP, and cerebellar ataxia disease genes, including *ATP1A1*, *UBAP1*, and *PRDX3* (Fig. 2) (Farazi Fard et al., 2019; Lassuthova et al., 2018; Rebelo et al., 2021b). Since GENESIS also functions as a long-term genetic data archive, patient organizations such as Charcot-Marie-Tooth Association (CMTA), the Hereditary Neuropathy Foundation (HNF), and the Muscular Dystrophy Association (MDA) have supported this platform as an important resource for their fields. Further, it is now established that a number of rare neurodegenerative disorders form a continuous phenotypic and genotypic spectrum and can exhibit overlapping phenotypes. The GENESIS database contains several such phenotypic continuums with unmatched sample sizes. Examples include the motoneuron disorders from ALS to HSP, SMA and motor dominant CMT (dHMN); as well as the cerebellar ataxia - sensory neuropathy axis.

2.1.4. Fostering a high pace of discovery to narrow the diagnostic gap

Since its inception, GENESIS has contributed to or has been cited by more than 200 peer-reviewed publications (Fig. 1D). Many of these studies represent major findings that impacted their respective fields, as evidenced by over 6000 citations of these papers (Fig. 1D). This imputes an h-index for GENESIS of 44 and a 5-year h-index of 35. These manuscripts which have referenced and utilized GENESIS, describe discoveries or major phenotypic expansions of 101 Mendelian genes (Fig. 2A,B, Supplementary Table 1). Over half of all genes discovered in the fields of inherited peripheral neuropathies and spastic paraplegia in the past ten years had a contribution from GENESIS. GENESIS has also facilitated

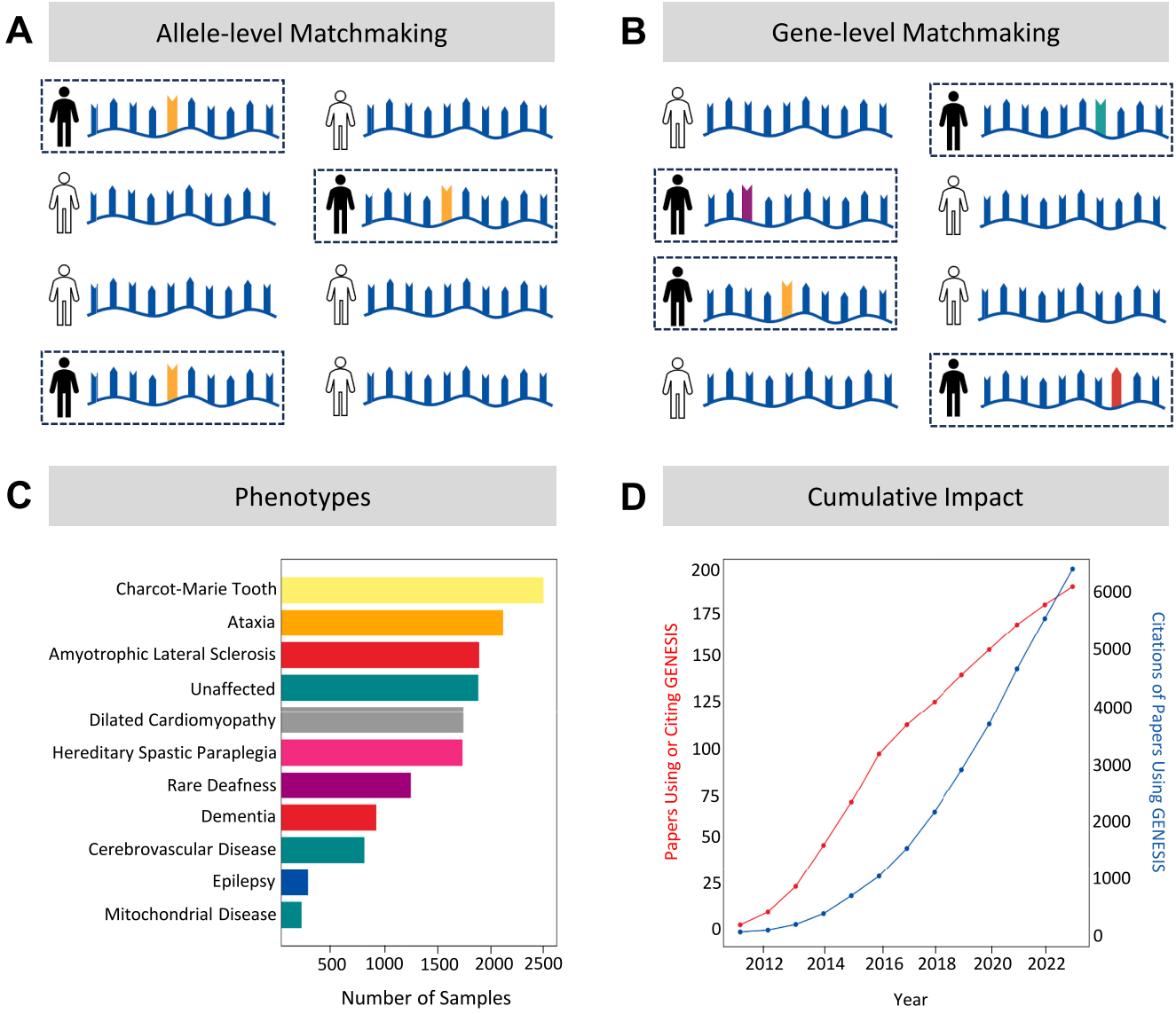


Fig. 1. GENESIS enables discovery through genetic matchmaking. A) Conceptual diagram of allele-level matchmaking. The causal variant shared by each of the affected individuals is shown in yellow. Individuals shown in black are affected and share the disease allele, while those shown in white do not. B) Conceptual diagram of gene-level matchmaking. The causal variant in each individual is shown as a distinct non-blue color, indicating that these are distinct variants within the same gene for each individual. Individuals shown in black are affected and have causal variants in this gene, while individuals shown in white are unaffected. C) Sample count for each of the top-10 disease phenotypes in GENESIS, plus unaffected individuals. D) Scatterplot of the cumulative impact of GENESIS since its inception, as measured by the cumulative number of papers using or citing GENESIS (red), and the cumulative citations of those papers (green).

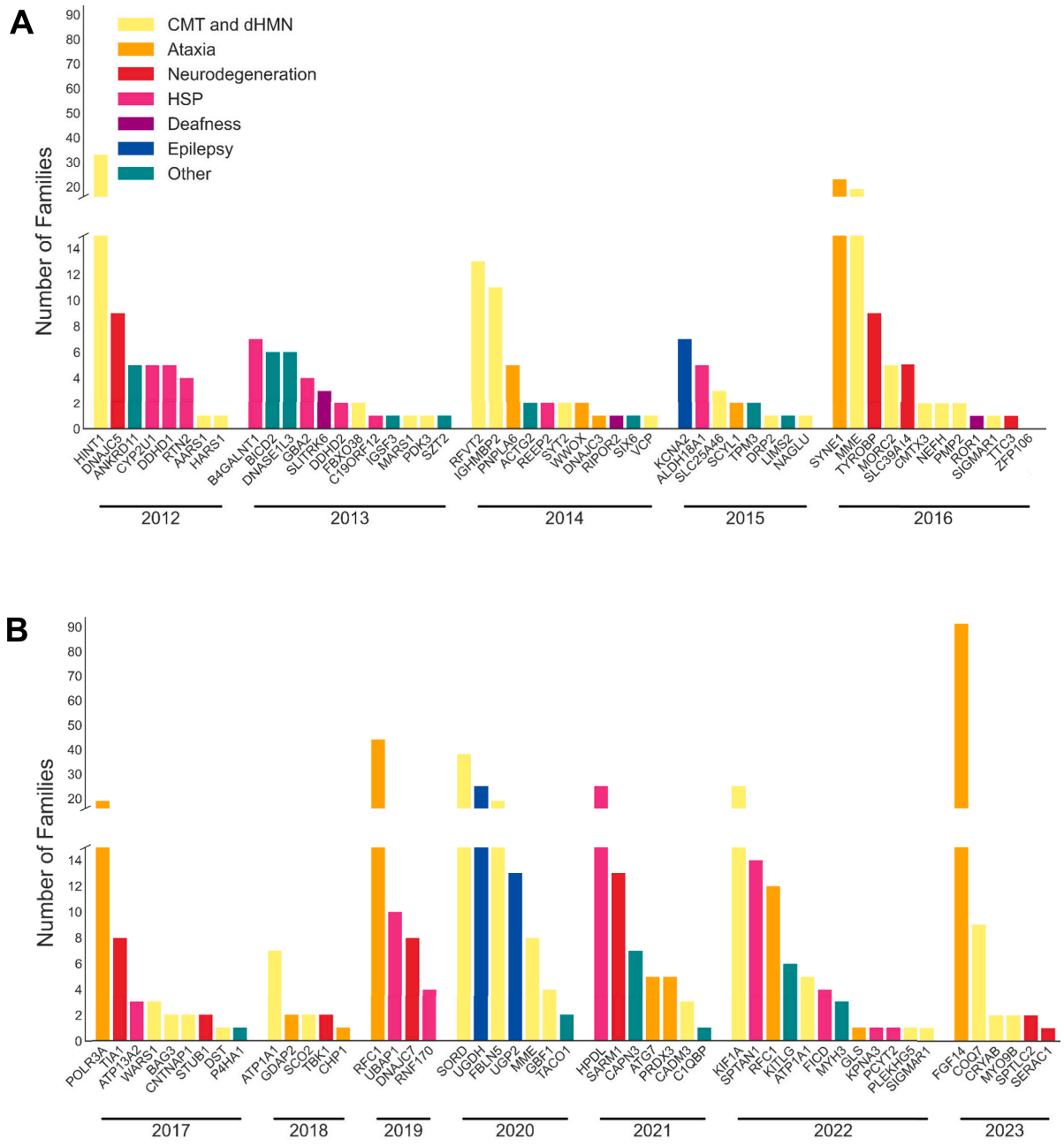


Fig. 2. GENESIS facilitates novel gene-disease relationship discoveries. A-B) Bar plots of the novel gene-disease relationships discovered with the help of GENESIS, ordered by year published. The y-axis plots the number of families carrying the novel variant(s) in that initial manuscript. A) presents years 2012 to 2016. B) presents years 2017 to 2023. Acronyms used: Charcot-Marie-Tooth disease (CMT); distal Hereditary Motor Neuropathy (dHMN); Hereditary Spastic Paraplegia (HSP).

numerous basic science discoveries in model organisms. We regularly sequence genomes of *drosophila* selected from forward genetic screens to identify protein-altering variants. Furthermore, researchers working with worms, zebrafish, or mice have used our species-specific pipelines on the GENESIS compute-backend for similar questions. This has opened opportunities for comparisons of gene effects in humans and in animal models.

2.2. *Different approaches to gene discoveries*

2.2.1. *Allele-level matchmaking identifies CADM3*

By screening the entire GENESIS catalog, three unrelated patients were identified sharing the same private variant in *CADM3*, p.Tyr172-Cys, including one de novo patient (Rebelo et al., 2021a). This variant is absent in 230,000 control chromosomes from gnomAD. All CADM3 patients shared a phenotype of dominant CMT2 with marked upper limb involvement. High resolution mass spectrometry analysis detected a newly created disulphide bond in the mutant CADM3 potentially modifying the native protein conformation. The CADM family of proteins consists of four neuronal specific adhesion molecules (CADM1, CADM2, CADM3 and CADM4) that mediate the direct contact and interaction between axons and glia. In the peripheral nerve, axon-Schwann cell interaction is essential for the structural organization of myelinated fibers and is primarily mediated by the binding of CADM3, expressed in axons, to CADM4, expressed by myelinating Schwann cells. This seemingly simple yet powerful genetic matching approach is only viable in large datasets of similar phenotypes. It demonstrates the value of data aggregation and variant matching. Subsequently, another

variant, p.Gly368Cys, has been identified in a large family in the country of Mali and observed de novo in a Caucasian child. Patients exhibited a similar, rather specific phenotype of axonal CMT and upper limb involvement (Yalcouyé et al., 2023).

2.2.2. Gene-level matchmaking enables discovery of *COQ7*

Enrichment of bi-allelic, rare, protein-altering variants in *COQ7* were identified by Rebelo and colleagues in nine families diagnosed with distal hereditary motor neuropathy (dHMN) via GENESIS and collaborative matchmaking (Rebelo et al., 2023). A total of five variants were observed in several combinations of homozygosity and compound heterozygosity across the families, with four of these being missense variants and one being a start-loss (Fig. 3A). This discovery highlights the power of collaborative efforts and advanced genomic tools in solving rare diseases, which would have been difficult to resolve as isolated cases. Since then, a number of papers have confirmed and replicated this finding (Jacquier et al., 2023b, Jacquier et al., 2023a; Liu et al., 2023; Qiu et al., 2024; Sadr et al., 2023; Wang et al., 2022). Coenzyme Q10 supplementation represents a potential treatment for this disease.

2.2.3. Tandem repeats analysis reveals a common ataxia, SCA27B

GENESIS leverages ExpansionHunter Denovo (EHDn), a bioinformatic tool which identifies large repetitive regions in short-read genomes in a catalog-free manner, to identify tandem repeat (TR) expansions in both known and novel disease genes (Dolzhenko et al., 2020). Six patients from three families all diagnosed with autosomal dominant, late-onset episodic, cerebellar ataxia were found by Pellerin and colleagues to all have evidence of a large repetitive region deep in intron 1 of *FGF14*, a known cerebellar ataxia gene, by EHDn during their processing into GENESIS (Pellerin et al., 2023). GENESIS additionally performed follow-up assessment with ExpansionHunter (Dolzhenko et al., 2019) and visualization tool REViewer (Dolzhenko et al., 2022), which revealed that the individuals indeed had one allele with a much larger TR than the other (representative REViewer image from one of

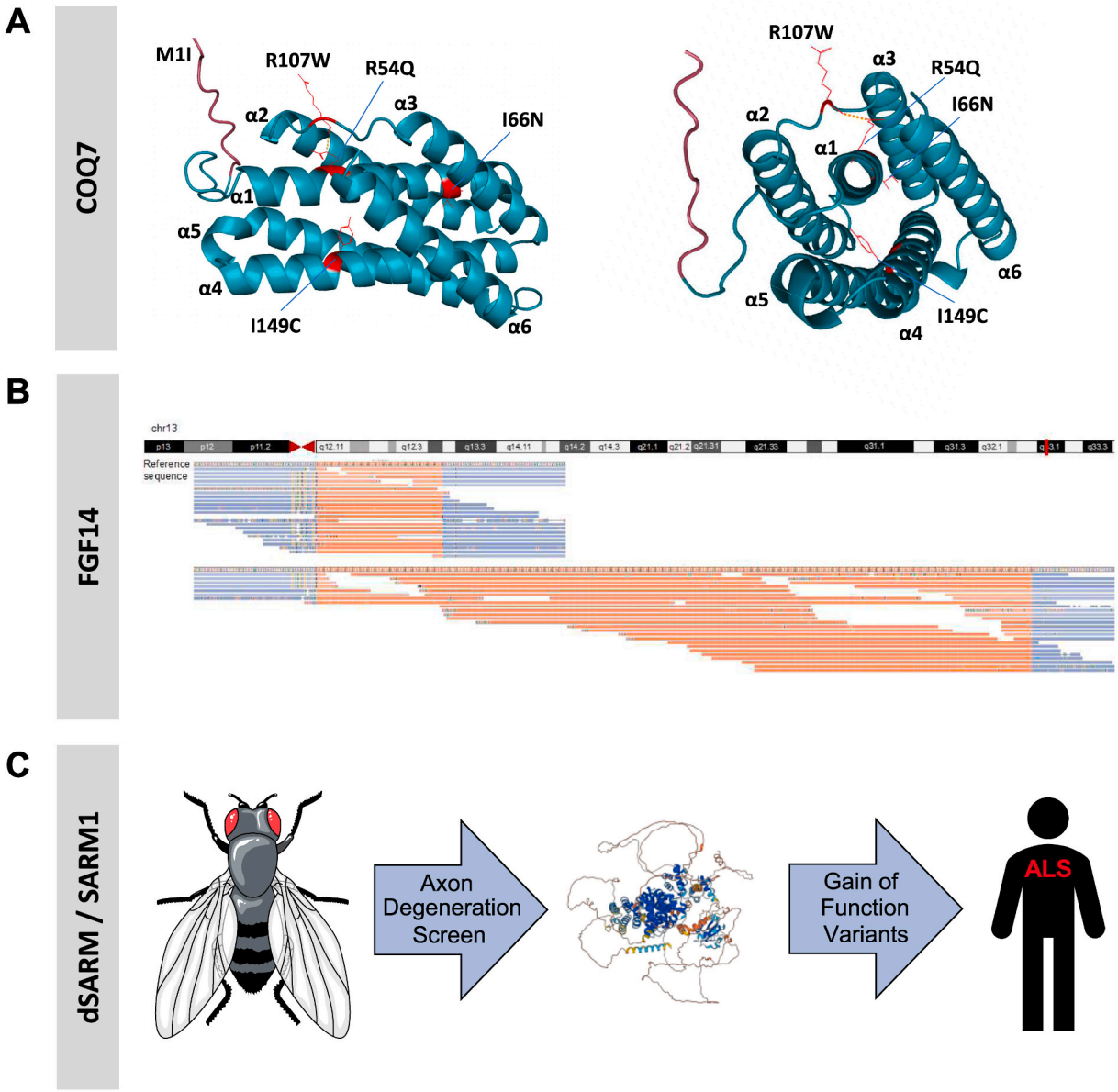


Fig. 3. GENESIS has aided discoveries involving diverse classes of variants. A) Structure of *COQ7* protein as predicted by AlphaFold, with the five variants identified by Rebelo and colleagues indicated (Figure adapted from Rebelo et al., 2023). B) REViewer visualization of Illumina reads supporting the heterozygous expansion of a GAA polymer in *FGF14* (Figure adapted from Pellerin et al., 2023). C) Diagram of the fly screen leading to the identification of dSARM as necessary for Wallerian degeneration of axons. This in turn led to the identification of gain-of-function variants in human SARM1 that cause ALS.

these individuals in Fig. 3B). This variant-level matchmaking success ultimately led to screening of additional patients and segregation studies within families which identified 128 individuals with this novel GAA expansion in *FGF14*, which was further confirmed through long-range PCR, repeat-primed PCR, and nanopore sequencing. This repeat expansion causes SCA27B and is now recognized as one of the most common genetic causes of adult-onset ataxia with nearly 1000 cases diagnosed globally within the first year after its publication.

2.2.4. Beyond Mendelian genetics and classic variants

Oligogenic and modifier studies are difficult to perform in rare diseases, as sample sizes are generally small. Aggregated genomic datasets from the GENESIS database have been useful for conducting studies of genetic modifiers of Mendelian phenotypes. Tao and colleagues performed such modifier gene study using patients with CMT1A and identified SNVs significantly associated with four clinical outcomes (Tao et al., 2019). A follow-up modifier study on a larger cohort of CMT1A patients is currently under way (Xu et al., 2024). A second study that leveraged GENESIS to identify genetic modifiers was conducted by Bis-Brewer and colleagues (Bis-Brewer et al., 2020). They performed a rare variant burden analysis using nearly a thousand CMT and HSP cases with a similar number of non-neurological controls. They were able to identify a heterozygous variant in the *EXOC4* gene which made carriers 9 times more likely to develop CMT than non-carriers. This study also uncovered evidence of increased mutational burden across known CMT and HSP disease genes and found evidence for oligogenic inheritance patterns in HSP cases. The ordered availability of large, rare disease datasets of raw data also facilitates efficient offline specialized statistical analyses. More such studies are expected in the future as modifier loci are potentially valuable targets for drug development (Chiò et al., 2014).

2.3. Genomics-to-therapy

2.3.1. Genetic therapies require detailed genotypic data and interpretation

One of the great promises of the coming decade is the advent of genetic therapies for inherited conditions. However, for such therapies to become a possibility for any given patient, a genetic diagnosis is required. The determination of a pathogenic variant to be causative requires a high level of confidence and data-driven support. The GENESIS database contains a wealth of rare alleles, provides diagnostic codes for each dataset, is highly enriched for rare diseases, and contains tools that go beyond what most other research resources offer. This includes deep annotation, artificial intelligence tools, detection of regions of homozygosity, relatedness and biological sex analyses, and the identification of phased compounded changes of different classes of genetic variation.

2.3.2. Unique success with homology regions: CMT-SORD

In 2020, Cortese and colleagues identified a recessive frameshift mutation in the *SORD* gene which causes CMT (Cortese et al., 2020). SORD-deficiency neuropathy is now recognized as one of the most common heritable forms of peripheral neuropathy. The frameshift variant was initially dismissed as a sequencing artifact caused by misalignment, because the frameshift variant commonly occurs in the homologous region of the *SORD2P* pseudogene. However, the GENESIS pipeline in 2020, correctly aligned sequencing reads in enough patients to make allelic matchmaking possible. At the time, most other GATK-based pipelines, including ExAc did not fully recognize the frequencies of this allele. Follow-up studies in fruit flies and iPSC-derived neurons supported sorbitol accumulation and treatment benefits of aldose reductase inhibitors. In less than 24 months, a multi-site treatment trial was fully enrolled (NCT05397665).

2.3.3. dSARM/SARM1 discovery and identification of highly penetrant ALS variants

Osterloh and colleagues performed forward genetic screening in *Drosophila* for mutants that exhibited long-term survival of severed axons (Osterloh et al., 2012). Their search yielded three lines which avoided Wallerian degeneration for weeks after axotomy. These three mutants were all recessive and fell into a single lethal complementation group, indicating that the same gene was responsible for the phenotype in all of them. Whole genome sequencing of these three lines and processing with an early version of GENESIS revealed dSARM/SARM1 to be the only gene with loss-of-function variants in all three *Drosophila* lines. SARM1 inhibitors are now being tested in clinical trials by several companies in order to protect peripheral nerves from degeneration (Disarm Therapeutics, 2024; Nura Bio Initiates Phase I Clinical Trial for its Oral, Brain-Penetrant SARM1 Inhibitor, NB-4746, 2023; Hughes et al., 2021). In 2021, Gilley and colleagues identified gain-of-function variants in the ARM domain of SARM1 in 13 individuals with ALS from GENESIS and several other ALS cohorts. Those changes were never observed in controls or individuals with any other phenotype, suggesting that they may be fully penetrant variants for ALS (Fig. 3C) (Gilley et al., 2021).

2.4. The technical basis of GENESIS

2.4.1. Bioinformatics pipeline overview

The GENESIS pipeline covers variant calls for genomic SNVs, InDels, structural variants (SVs), tandem repeats (TRs), as well as mitochondrial SNVs and InDels (Fig. 4A-B). The following sentences describe which tools are used to genotype each of those variant types in either short-read or long-read samples. GENESIS supports short-read and long-read data ranging from targeted panels to whole genomes (Fig. 4A-B, Supplementary Fig. 3). Genomic and mitochondrial SNVs and InDels for short-read data are called using a pipeline based around GATK4 best practices (Poplin et al., 2018), while in long-read data they are called using DeepVariant (Yun et al., 2020). SVs are called with Manta (Chen et al., 2016) in short-read genomes. In long-read PacBio samples, SVs are called with PBSV, while in long-read Nanopore samples, SVs are called with Sniffles2 (Smolka et al., 2024). TRs in short-read data are identified applying an in-house developed pipeline based around ExpansionHunter and ExpansionHunter Denovo (Dolzhenko et al., 2020, 2019). In long-read PacBio samples, TRs are determined with TRGT (Dolzhenko et al., 2024), while in long-read Nanopore samples, TRs are called with LongTR (Ziaei Jam et al., 2024) (Fig. 4C). Each of these pipelines have been made open source (https://gitlab.com/genesis-genomics/sample-processing). More details on these tools can be found in Supplementary Table 2.

2.4.2. Variant annotation

For each of these classes of genomic variation, GENESIS provides comprehensive annotations (Fig. 5). This includes gnomAD for allele frequencies for SNVs, InDels, SVs, and a select set of known pathogenic TRs (Karczewski et al., 2020). For broader allele frequencies of TRs, we developed our own resource and are looking towards integrating AOURP data in the future (Fazal et al., 2020). Major functional artificial intelligence-based prediction tools have been implemented in GENESIS, including SpliceAI for splicing variants (Jaganathan et al., 2019), RExPRT AI for TRs (Fazal et al., 2024), and MAVERICK AI, which can score InDels in addition to protein altering SNVs (Danzi et al., 2023). Further annotations apply at the gene level, such as OMIM data and constraint scores from gnomAD. An important component of this annotation framework is that it is regularly updated for actively changing resources such as ClinVar and OMIM. A total of 212 annotations are applied to variants in GENESIS. More details on select major annotations can be found in Supplementary Table 2.

2.4.3. Quality assessment

The processing pipeline performs quality checks for each dataset with 15 parameters, such as average depth, SNV count, and reference coverage. Every dataset is also compared to the distribution of all data

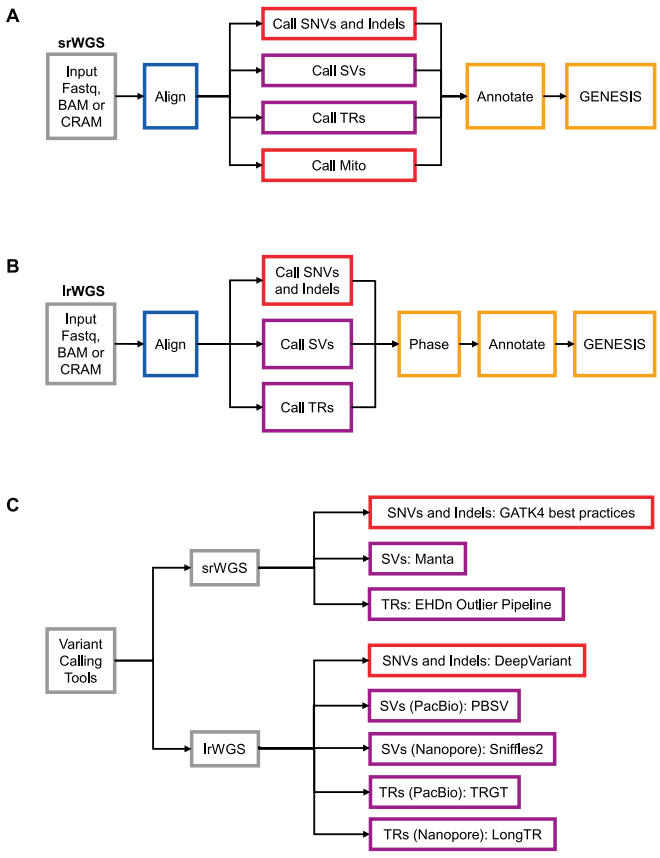


Fig. 4. GENESIS handles variant calling and annotation across a wide range of variant types. A) Schematic of short-read WGS data processing in GENESIS, which enables calling of genomic SNVs, InDels, structural variants (SVs), tandem repeats (TRs), and mitochondrial SNVs and InDels. B) Schematic of long-read WGS data processing in GENESIS, which enables calling of genomic SNVs, InDels, SVs, and TRs, and phases them all together. C) Schematic of the different variant calling tools used by GENESIS. Acronyms used: Binary Alignment Map (BAM); Compressed Reference-oriented Alignment Map (CRAM); Single Nucleotide Variation (SNV); small Insertion or Deletion (InDel); Structural Variation (SV); Tandem Repeat (TR); Mitochondrial variation (Mito); short-read Whole Genome Sequencing (srWGS); long-read Whole Genome Sequencing (lrWGS); Genomes Analysis ToolKit (GATK); ExpansionHunter Denovo (EHDn); Tandem Repeat Genotyping Tool (TRGT). See Supplementary Table 2 for descriptions of the utilized tools.

for each parameter calculating a Z score. Data are further checked for relatedness to all other genomes in the database and for sex, allowing for detection of sample duplications as well as family mis-assignments. Variant-level quality assessment of short-read SNVs and InDels is performed using GATK's CNNScoreVariants utility, which provides valuable confidence estimations of the quality of each called variant.

2.4.4. Security and regulatory considerations

Data security is of highest priority for any genetic data collection. Multiple layers of protection have been implemented, such as true dual factor authentication, end-to-end encryption of data, and no storage and processing of personal health information (PHI and PPI). GENESIS is following guiding principles of other public data collections, such as dbGaP security best practices, requirements for GDPR, and expectations of academic computing centers. Each user of GENESIS is assigned a random universally unique identifier (UUID) upon registration. This random UUID is used for all client (e.g. web browser) transactions with GENESIS servers. All requests must include a JSON Web Token (JWT) encrypted with a secret key storing a secret base64-encoded value. Any requests without a valid JWT token are denied and the event is recorded in the server logs. The use of the AWS platform has the further benefit of standardized and ISO certified security protocols accepted in major legal systems, such as the USA and the European Union.

3. Discussion

Data aggregation and sharing at scale, broadly usable tools, and standardized analysis pipelines have become indispensable assets for rare disease research. GENESIS was one of the very first such projects and by dissecting its impact on the field we are providing evidence for the value of such efforts. The specific challenges that come with impressive locus heterogeneity and low incidence rate of many rare neurodegenerative disease can be solved by creating common data pools that are appropriately de-identified and secured by technical means. By creating a relatively simple yet powerful graphical user interface, the platform is useful for researchers, clinicians, domain experts, and students alike. The fact that queries are computed in real time, rather than precomputed, is a major advantage and a clear distinction from simple dashboard functionalities. In the decade since its inception, GENESIS has been able to assist in the efforts of hundreds of researchers in discovering over 100 novel gene-disease relationships (Supplementary Table 1). The database supports major rare disease initiatives, which have plans to grow further, as typically 30–50 % of patients still do not receive a genetic diagnosis. This hints at additional rare causes of disease, but increasingly suggests an unrealized discovery opportunity in the non-coding space. Structural changes, such as repeat expansions, in particular, have the potential to close the diagnostic gap significantly. Other bioinformatically challenging conditions include genomic regions of homology, that may include protein coding transcripts or important regulatory elements. GENESIS already provides expert tools for structural variant detection and can return compounded combinations of SNVs, SVs, and more. We expect that multi-modal 'omics' datasets will become increasingly common and vital for solving particularly challenging cases. We plan to develop pipelines that enable simultaneous analysis of multiple datasets from the same individual, such as RNA-seq, methylation, and WGS. GENESIS is one of few easy-to-use research tools with long-read genomics technologies integrated. We are expecting this area of the database to grow rapidly over the next few years. Further, it is expected that machine learning techniques, especially neural network-based artificial intelligence (AI) approaches, will be able to pinpoint alleles and genes of interest for rare disease research. GENESIS already uses artificial intelligence elements, such as Splice AI, CNN for quality assessment, RExPRT for repeat expansion classification, and MAVERICK for coding variant pathogenicity ranking. The available rich genomic dataset is an ideal testing ground to develop and further improve such tools. In summary, the next decade for GENESIS will be as groundbreaking and impactful, driven and enabled by talented and diverse teams of scientists committed to data sharing.

4. Methods

4.1. GENESIS data structure

GENESIS is implemented as a ClickHouse database hosted on AWS. There are separate groups of tables for SNV/InDel, SV, and TR variants. Within those groups are tables for annotations, tables for the genotype calls in short-read samples, and tables for the genotype calls in long-read samples. On top of this database, we have constructed the Django-based website hosted at tgp-foundation.org, which facilitates querying the database for users, maintaining divisions in which samples may be queried by each user. Queries for protein-altering variants segregating among members of a family with 2–5 members typically take ~10 s to complete and return results to users. Queries across many hundreds of families or that include non-coding regions will take longer.

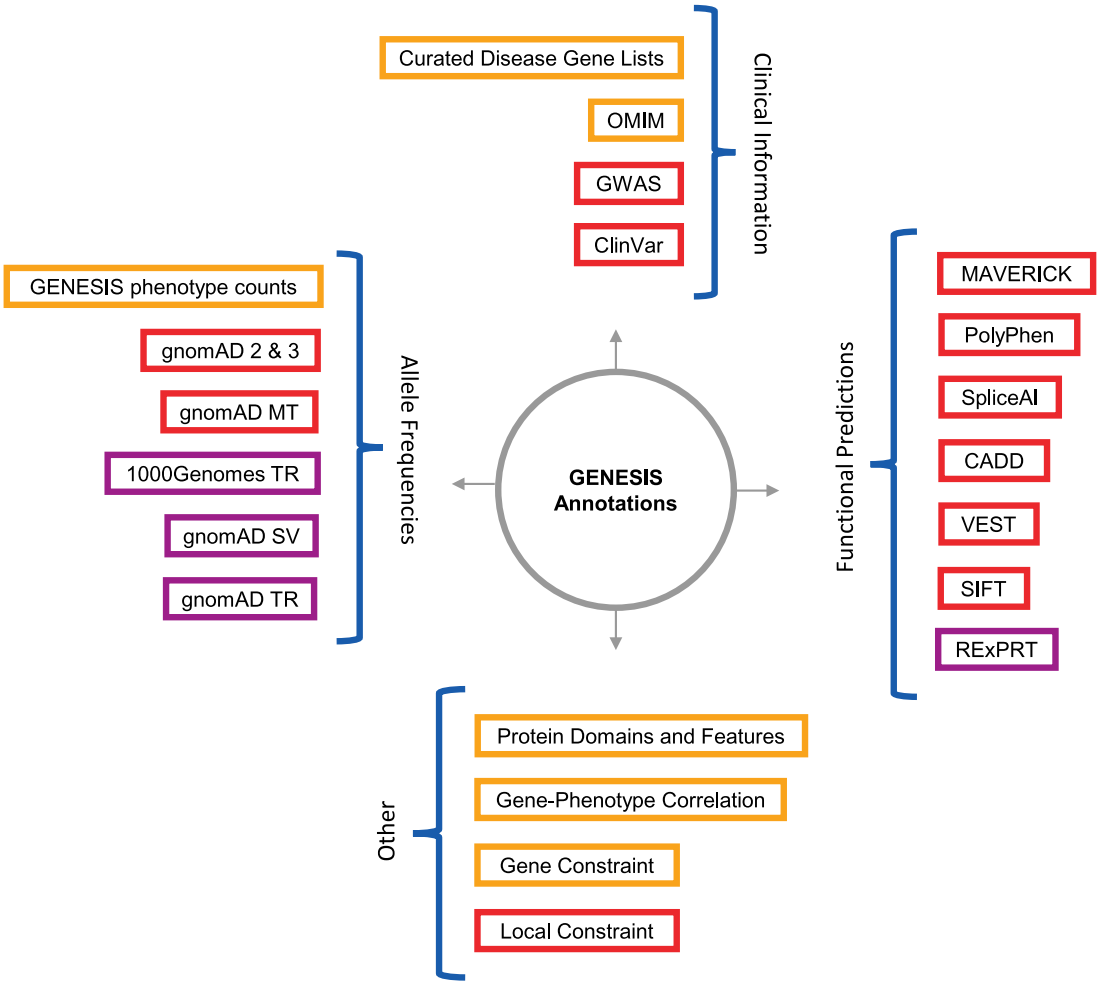


Fig. 5. GENESIS annotates variants to facilitate prioritization. Schematic of the major groups of annotations in GENESIS and the sources of data for each. Acronyms used: Genome Aggregation Database (gnomAD); Online encyclopedia of Mendelian Inheritance in Man (OMIM); Genome-Wide Association Study (GWAS). See Supplementary Table 2 for descriptions of the utilized annotations.

4.2. Genomic data processing in GENESIS

Genomic data processing for input into the GENESIS database is performed using AWS Batch and conducted by Nextflow pipelines. When short-read genomic samples are uploaded, they are aligned to the GRCh37 reference genome by BWA version 0.7.17 (Li, 2013), sorted by Sambamba (Tarasov et al., 2024), and have their optical duplicate reads marked by GATK's MarkDuplicates utility(Poplin et al., 2018). That bam file is then used for processing SNVs/InDels, mitochondrial SNVs/InDels, SVs, and TRs by parallel processes. For nuclear SNVs and InDels, GATK v4.2.6.1 is used to perform base quality score recalibration and then haplotype calling (Poplin et al., 2018). This creates a gVCF for each sample which is stored in perpetuity and later used to joint-call the sample with any additional family members in GENESIS (as well as additional family members added to GENESIS in the future). For mitochondrial variants, Mutect is used on reads aligned to the mitochondrial reference genome as well as a version of the reference shifted by 7000 bp in order to alleviate issues with read alignment near the computational edges of the circular chromosome (Laricchia et al., 2022). Variant calls are then combined together and filtered. Structural variants are called on whole genomes with Manta version 1.6.0 as implemented in Parliament2 (Chen et al., 2016; Zarate et al., 2018). No SV calling is performed on exomes or targeted panels. Candidate tandem repeats are called using ExpansionHunter Denovo (EHDn) on whole genomes (Dolzhenko et al., 2020). These candidate TRs are then compared to a reference panel of EHDn signal across the 1000Genomes control sample set to identify outliers. These outliers are then groomed and reformatted for ExpansionHunter (Dolzhenko et al., 2019). In whole genomes, ExpansionHunter genotypes these candidate loci along with all currently known-pathogenic loci (list updated regularly). In exomes, ExpansionHunter only genotypes currently known pathogenic loci in protein-coding regions. In both cases, REViewer is then used to visualize the genotypes (Dolzhenko et al., 2022).

When long-read genomic samples are uploaded, they are processed in a technology-dependent manner. Tools used for Oxford Nanopore Technologies (ONT) samples versus Pacific Biosciences (PacBio) samples will be denoted. Samples are aligned to the GRCh37 reference genome by minimap2 v2.26 (ONT) or pbmm2 v1.13.1 (PacBio) (Li, 2018). SNVs and InDels are called with DeepVariant v1.6.1 (Yun et al., 2020). SVs are called with Sniffles2 v2.0.7 (Smolka et al., 2024). TRs are called with Medaka (ONT) or TRGT (PacBio) (Dolzhenko et al., 2024). For ONT, SNVs and InDels variants are phased with WhatsHap and phasing is currently not performed on SVs or TRs (Martin et al., 2016). For PacBio, SNVs, InDels, SVs, and TRs are jointly phased with HiPhase (Holt et al., 2024). As done for short-read samples, the SNVs and InDels are joint-called with other members of their family, though only within technology type. Families with samples spanning multiple technology types will not have fully jointly called variant sets. For long-read samples, joint calling is performed with GLNexus (Yun et al., 2020a,b).

4.3. GENESIS quality assessment

Each sample is benchmarked for quality across a wide range of values and compared to all other samples of the same sequencing type (WES, srWGS, PacBio lrWGS, or ONT lrWGS). We report the raw value of each quality measure along with a Z score showing how it compares with other samples of that same sequencing type in GENESIS. For WES and srWGS, the quality checking values we employ are: the number of sequencing reads; the average read PHRED quality for reads 1 and 2 separately and the difference between them; the alignment rate; the read pairing rate; the optical duplication rate; reference coverage; exon enrichment; TSTV ratio; count of SNVs; count of InDels; average read depth for loci with SNVs or InDels; average QUAL score for SNVs and InDels; average GQ score for SNVs and InDels; and the heterozygous-to-homozygous ratio for SNVs and InDels. For long-read samples, the quality checking values we employ are: the number of sequencing reads, the average PHRED quality of the reads; the alignment rate; reference coverage; TSTV ratio; count of SNVs; count of InDels; count of SVs; average read depth for loci with SNVs or InDels; average QUAL score for SNVs and InDels; average GQ score for SNVs and InDels; the heterozygous-to-homozygous ratio for SNVs and InDels; and the NG50 of phase block size. In addition to the quality measures described above, we also utilize Peddy to aid in the detection of sample swaps (Pedersen and Quinlan, 2017). First, we predict the sex of each sample and check if it matches the sex indicated in the metadata. Second, Peddy is further used to calculate relatedness coefficients between all pairs of samples.

4.4. Cumulative impact calculation

In order to calculate the direct and indirect cumulative impact of GENESIS, we collected the set of papers citing either of the two original GENESIS/GEM.app manuscripts (Gonzalez et al., 2015, Gonzalez et al., 2013) as well as any other manuscripts we were aware of that made their central genetic discovery using GENESIS. These formed the 'direct impact' set of 'papers using or citing GENESIS'. We then used the PubMed API to collect all manuscripts that cited each of those works. These manuscripts formed the 'indirect impact' set of 'citations of papers using GENESIS'.

CRediT authorship contribution statement

Matt C. Danzi: Writing – review & editing, Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Eric Powell:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Conceptualization. **Adriana P. Rebelo:** Writing – review & editing, Investigation. **Maike F. Dohrn:** Writing – review & editing, Investigation. **Danique Beijer:** Writing – review & editing, Investigation. **Sarah Fazal:** Writing – review & editing, Investigation. **Isaac R.L. Xu:** Writing – review & editing, Investigation. **Jessica Medina:** Writing – review & editing, Investigation. **Sitong Chen:** Writing – review & editing, Investigation. **Yeisha Arcia de Jesus:** Writing – review & editing, Investigation. **Jacquelyn Schatzman:** Writing – review & editing, Investigation. **Ray E. Hershberger:** Writing – review & editing, Supervision. **Mario Saporta:** Writing – review & editing, Supervision. **Jonathan Baets:** Writing – review & editing, Supervision. **Marni Falk:** Writing – review & editing, Supervision. **David N. Herrmann:** Writing – review & editing, Supervision. **Steven S. Scherer:** Writing – review & editing, Supervision. **Mary M. Reilly:** Writing – review & editing, Supervision. **Andrea Cortese:** Writing – review & editing, Visualization. **Wilson Marques:** Writing – review & editing, Supervision. **Mario R. Carnejo-Olivas:** Writing – review & editing, Supervision. **Oranee Sanmaneechai:** Writing – review & editing, Supervision. **Marina L. Kennerson:** Writing – review & editing, Supervision. **Albena Jordanova:** Writing – review & editing, Supervision. **Thiago Y.T. Silva:** Writing – review & editing, Investigation. **Jose Luiz Pedroso:** Writing – review & editing, Supervision. **Luca Schierbaum:** Writing – review & editing, Investigation. **Darius Ebrahimi-Fakhari:** Writing – review & editing, Supervision. **Stojan Peric:** Writing – review & editing, Supervision. **Yi-Chung Lee:** Writing – review & editing, Supervision. **Matthis Synofzik:** Writing – review & editing, Supervision. **Mustafa Tekin:** Writing – review & editing, Supervision. **Gianina Ravenscroft:** Writing – review & editing, Supervision. **Mike Shy:** Writing – review & editing, Supervision. **Nazli Basak:** Writing – review & editing, Supervision. **Rebecca Schule:** Writing – review & editing, Supervision. **Stephan Zuchner:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Data availability

Data will be made available on request.

Acknowledgments

This study and data analysis was supported by the CMT Association, Hereditary Neuropathy Foundation, and Muscular Dystrophy Association (to SZ). GR is supported by an EL2 fellowship from the Australian NHMRC (APP2007769). JB is supported by a Senior Clinical Researcher mandate of the Research Fund - Flanders (FWO) under grant agreement number 1805021 N. AJ is supported by the Fund for Scientific Research (FWO-Flanders) (#G048220N, #G0A2122N), the Association Belge contre les Maladies Neuromusculaires' (ABMM-Telethon), the French Muscular Dystrophy Association (AFM-Telethon, #23708.), the Bulgarian National Science Fund and the Bulgarian National Plan for Recovery and Resilience (#BG-RRP-2.004-0004-C01). Several authors of this publication are members of the European Reference Network for Rare Neuromuscular Diseases (ERN EURO-NMD) and of the European Reference Network for Rare Neurological Diseases (ERN-RND). JB is a member of the μNEURO Research Centre of Excellence of the University of Antwerp.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.expneurol.2024.114978.

References

Adams, D., Gonzalez-Duarte, A., O'Riordan, W.D., Yang, C.-C., Ueda, M., Kristen, A.V., Tournev, I., Schmidt, H.H., Coelho, T., Berk, J.L., Lin, K.-P., Vita, G., Attarian, S., Planté-Bordeneuve, V., Mezei, M.M., Campistol, J.M., Buades, J., Brannagan, T.H., Kim, B.J., Oh, J., Parman, Y., Sekijima, Y., Hawkins, P.N., Solomon, S.D., Polydefkis, M., Dyck, P.J., Gandhi, P.J., Goyal, S., Chen, J., Strahs, A.L., Nochur, S. V., Sweetser, M.T., Garg, P.P., Vaishnaw, A.K., Gollob, J.A., Suhr, O.B., 2018. Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. New England J. Med. 379, 11–21. https://doi.org/10.1056/nejmoa1716153.

Balwani, M., Sardh, E., Ventura, P., Peiró, P.A., Rees, D.C., Stölzel, U., Bissell, D.M., Bonkovsky, H.L., Windyga, J., Anderson, K.E., Parker, C., Silver, S.M., Keel, S.B., Wang, J.-D., Stein, P.E., Harper, P., Vassiliou, D., Wang, B., Phillips, J., Ivanova, A., Langendonk, J.G., Kauppinen, R., Minder, E., Horie, Y., Penz, C., Chen, J., Liu, S., Ko, J.J., Sweetser, M.T., Garg, P., Vaishnaw, A., Kim, J.B., Simon, A.R., Gouya, L., 2020. Phase 3 trial of RNAi therapeutic Givosiran for acute intermittent Porphyria. New England J. Med. 382, 2289–2301. https://doi.org/10.1056/nejmoa1913147.

Bis-Brewer, D.M., Gan-Or, Z., Sleiman, P., Rodriguez, A., Bacha, A., Kosikowski, A., Wood, B., McCray, B., Blume, B., Siskind, C., Sumner, C., Calabrese, D., Walk, D., Vujovic, D., Park, E., Muntoni, F., Donlevy, G., Acsadi, G., Day, J., Burns, J., Li, J., Krajewski, K., Eichinger, K., Cornett, K., Mullen, K., Perez, L., Gutmann, L., Barrett, M., Saporta, M., Skorupinska, M., Grant, N., Bray, P., Seyedsadjadi, R., Zuccarino, R., Finkel, R., Lewis, R., Yum, S., Hilbert, S., Thomas, S., Behrens-Spraggins, S., Jones, T., Grider, T., Estilow, T., Fridman, V., Reilly, M.M., Shy, M.E., Bacon, C.J., Feely, S.M.E., Rossor, A.M., Herrmann, D.N., Hakonarson, H., Fazal, S.,

Courel, S., Cintra, V., Tao, F., Estiar, M.A., Tarnopolsky, M., Boycott, K.M., Yoon, G., Suchowersky, O., Dupré, N., Cheng, A., Lloyd, T.E., Rouleau, G., Schüle, R., Züchner, S., 2020. Assessing non-Mendelian inheritance in inherited axonopathies. Genet. Med. 22, 2114–2119. https://doi.org/10.1038/s41436-020-0924-0.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., Saunders, C.T., 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32, 1220–1222. https://doi.org/10.1093/bioinformatics/btv710.

Chiò, A., Calvo, A., Moglia, C., Canosa, A., Brunetti, M., Barberis, M., Restagno, G., Conte, A., Bisogni, G., Marangi, G., Moncada, A., Lattante, S., Zollino, M., Sabatelli, M., Bagarotti, A., Corrado, L., Mora, G., Bersano, E., Mazzini, L., Alfonso, S. D., 2014. ATXN2 polyQ Intermediate Repeats are a Modifier of ALS Survival.

Cortese, A., Zhu, Y., Rebelo, A.P., Negri, S., Courel, S., Abreu, L., Bacon, C.J., Bai, Y., Bis-Brewer, D.M., Bugiardini, E., Buglo, E., Danzi, M.C., Feely, S.M.E., Athanasiou-Fragkouli, A., Haridy, N.A., Rodriguez, A., Bacha, A., Kosikowski, A., Wood, B., McCray, B., Blume, B., Siskind, C., Sumner, C., Calabrese, D., Walk, D., Vujovic, D., Park, E., Muntoni, F., Donlevy, G., Acsadi, G., Day, J., Burns, J., Li, J., Krajewski, K., Eichinger, K., Cornett, K., Mullen, K., Laura, P.Q., Gutmann, L., Barrett, M., Saporta, M., Skorupinska, M., Grant, N., Bray, P., Seyedsadjadi, R., Zuccarino, R., Finkel, R., Lewis, R., Yum, S., Hilbert, S., Thomas, S., Behrens-Spraggins, S., Jones, T., Lloyd, T., Grider, T., Estilow, T., Fridman, V., Isasi, R., Khan, A., Laurà, M., Magri, S., Pipis, M., Pisciotta, C., Powell, E., Rossor, A.M., Saveri, P., Sowden, J.E., Tozza, S., Vandrovcova, J., Dallman, J., Grignani, E., Marchioni, E., Scherer, S.S., Tang, B., Lin, Z., Al-Ajmi, A., Schüle, R., Synofzik, M., Maisonobe, T., Stojkovic, T., Auer-Grumbach, M., Abdelhamed, M.A., Hamed, S.A., Zhang, R., Manganelli, F., Santoro, L., Taroni, F., Pareyson, D., Houlden, H., Herrmann, D.N., Reilly, M.M., Shy, M.E., Zhai, R.G., Zuchner, S., 2020. Biallelic mutations in SORD cause a common and potentially treatable hereditary neuropathy with implications for diabetes. Nat. Genet. 52, 473–481. https://doi.org/10.1038/s41588-020-0615-4.

Danzi, M.C., Dohrn, M.F., Fazal, S., Beijer, D., Rebelo, A.P., Cintra, V., Züchner, S., 2023. Deep structured learning for variant prioritization in Mendelian diseases. Nat. Commun. 14. https://doi.org/10.1038/s41467-023-39306-7.

Disarm Therapeutics [WWW Document], 2024. https://www.lilly.com/discovery/research-and-scientific-discovery/disarm-therapeutics.

Dolzhenko, E., Deshpande, V., Schlesinger, F., Krusche, P., Petrovski, R., Chen, S., Emig-Agius, D., Gross, A., Narzisi, G., Bowman, B., Scheffler, K., Van Vugt, J.J.F.A., French, C., Sanchis-Juan, A., Ibáñez, K., Tucci, A., Lajoie, B.R., Veldink, J.H., Raymond, F.L., Taft, R.J., Bentley, D.R., Eberle, M.A., Birol, I., 2019. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics 35, 4754–4756. https://doi.org/10.1093/bioinformatics/btz431.

Dolzhenko, E., Bennett, M.F., Richmond, P.A., Trost, B., Chen, S., Van Vugt, J.J.F.A., Nguyen, C., Narzisi, G., Gainullin, V.G., Gross, A.M., Lajoie, B.R., Taft, R.J., Wasserman, W.W., Scherer, S.W., Veldink, J.H., Bentley, D.R., Yuen, R.K.C., Bahlo, M., Eberle, M.A., 2020. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol. 21, 1–14. https://doi.org/10.1186/s13059-020-02017-z.

Dolzhenko, E., Weisburd, B., Ibañez, K., Rajan-Babu, I.S., Anyansi, C., Bennett, M.F., Billingsley, K., Carroll, A., Clamons, S., Danzi, M.C., Deshpande, V., Ding, J., Fazal, S., Halman, A., Jadhav, B., Qiu, Y., Richmond, P.A., Saunders, C.T., Scheffler, K., van Vugt, J.J.F.A., Zwamborn, R.R.A.J., Chong, S.S., Friedman, J.M., Tucci, A., Rehm, H.L., Eberle, M.A., 2022. REViewer: haplotype-resolved visualization of read alignments in and around tandem repeats. Genome Med. 14. https://doi.org/10.1186/s13073-022-01085-z.

Dolzhenko, E., English, A., Dashnow, H., De Sena Brandine, G., Mokveld, T., Rowell, W. J., Karniski, C., Kronenberg, Z., Danzi, M.C., Cheung, W.A., Bi, C., Farrow, E., Wenger, A., Chua, K.P., Martínez-Cerdeño, V., Bartley, T.D., Jin, P., Nelson, D.L., Zuchner, S., Pastinen, T., Quinlan, A.R., Sedlazeck, F.J., Eberle, M.A., 2024. Characterization and visualization of tandem repeats at genome scale. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-02057-3.

Farazi Fard, M.A., Rebelo, A.P., Buglo, E., Nemati, H., Dastsooz, H., Gehweiler, I., Reich, S., Reichbauer, J., Quintáns, B., Ordóñez-Ugalde, A., Cortese, A., Courel, S., Abreu, L., Powell, E., Danzi, M., Martuscelli, N.B., Bis-Brewer, D.M., Tao, F., Zarei, F., Habibzadeh, P., Yavarian, M., Modarresi, F., Silawi, M., Tabatabaei, Z., Yousefi, M., Farpour, H.R., Kessler, C., Mangold, E., Kobeleva, X., Mueller, A.J., Haack, T.B., Tarnopolsky, M., Gan-Or, Z., Rouleau, G.A., Synofzik, M., Sobrido, M.-J., Jordanova, A., Schüle, R., Zuchner, S., Faghihi, M.A., 2019. Truncating mutations in UBAP1 cause hereditary spastic paraplegia. Am. J. Hum. Genet. 104. https://doi.org/10.1016/j.ajhg.2019.03.001.

Fazal, S., Danzi, M.C., Cintra, V.P., Bis-Brewer, D.M., Dolzhenko, E., Eberle, M.A., Zuchner, S., 2020. Large scale in silico characterization of repeat expansion variation in human genomes. Sci Data 7, 1–14. https://doi.org/10.1038/s41597-020-00633-9.

Fazal, S., Danzi, M.C., Xu, I., Kobren, S.N., Sunyaev, S., Reuter, C., Marwaha, S., Wheeler, M., Dolzhenko, E., Lucas, F., Wuchty, S., Tekin, M., Züchner, S., Aguiar-Pulido, V., 2024. RExPRT: a machine learning tool to predict pathogenicity of tandem repeat loci. Genome Biol. 25, 39. https://doi.org/10.1186/s13059-024-03171-4.

Finkel, R.S., Mercuri, E., Darras, B.T., Connolly, A.M., Kuntz, N.L., Kirschner, J., Chiriboga, C.A., Saito, K., Servais, L., Tizzano, E., Topaloglu, H., Tulinius, M., Montes, J., Glanzman, A.M., Bishop, K., Zhong, Z.J., Gheuens, S., Bennett, C.F., Schneider, E., Farwell, W., De Vivo, D.C., 2017. Nusinersen versus sham control in infantile-onset spinal muscular atrophy. New England J. Med. 377, 1723–1732. https://doi.org/10.1056/nejmoa1702752.

Gilley, J., Jackson, O., Pipis, M., Estiar, M.A., Al-Chalabi, A., Danzi, M.C., van Eijk, K.R., Goutman, S.A., Harms, M.B., Houlden, H., Iacoangeli, A., Kaye, J., Lima, L., Genomics, Q.S., Ravits, J., Rouleau, G.A., Schüle, R., Xu, J., Züchner, S., Cooper-Knock, J., Gan-Or, Z., Reilly, M.M., Coleman, M.P., 2021. Enrichment of sarm1 alleles encoding variants with constitutively hyperactive nadase in patients with als and other motor nerve disorders. Elife 10. https://doi.org/10.7554/eLife.70905.

Gonzalez, M., Falk, M.J., Gai, X., Postrel, R., Schüle, R., Zuchner, S., 2015. Innovative genomic collaboration using the GENESIS (GEM.App) platform. Hum. Mutat. 36, 950–956. https://doi.org/10.1002/humu.22836.

Gonzalez, M.A., Lebrigio, R.F.A., Van Booven, D., Ulloa, R.H., Powell, E., Speziani, F., Tekin, M., Schüle, R., Züchner, S., 2013. GEnomes management application (GEM. App): a new software tool for large-scale collaborative genome analysis. Hum. Mutat. 34, 842–846. https://doi.org/10.1002/humu.22305.

Holt, J.M., Saunders, C.T., Rowell, W.J., Kronenberg, Z., Wenger, A.M., Eberle, M., 2024. HiPhase: jointly phasing small, structural, and tandem repeat variants from HiFi sequencing. Bioinformatics 40. https://doi.org/10.1093/bioinformatics/btae042.

Hughes, R.O., Bosanac, T., Mao, X., Engber, T.M., DiAntonio, A., Milbrandt, J., Devraj, R., Krauss, R., 2021. Small molecule SARM1 inhibitors recapitulate the SARM1−/− phenotype and allow recovery of a metastable Pool of axons fated to degenerate. Cell Rep. 34. https://doi.org/10.1016/j.celrep.2020.108588.

Jacquier, A., Theuriet, J., Fontaine, F., Mosbach, V., Lacoste, N., Ribault, S., Risson, V., Carras, J., Coudert, L., Simonet, T., Latour, P., Stojkovic, T., Piard, J., Cosson, A., Lesca, G., Bouhour, F., Allouche, S., Puccio, H., Pegat, A., Schaeffer, L., 2023a. Homozygous COQ7 mutation: a new cause of potentially treatable distal hereditary motor neuropathy. Brain 146, 3470–3483. https://doi.org/10.1093/brain/awac453.

Jacquier, A., Theuriet, J., Ribault, S., Lacoste, N., Pegat, A., Latour, P., Schaeffer, L., 2023b. Reply: Biallelic variants in the COQ7 gene cause distal hereditary motor neuropathy in two Chinese families. Brain. https://doi.org/10.1093/brain/awad041.

Jaganathan, K., Panagiotopoulou, S.K., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., Chow, E.D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S.J., Farh, K.K.H., 2019. Predicting splicing from primary sequence with deep learning. Cell 176, 535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015.

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., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E. G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Aguilar Salinas, C.A., Ahmad, T., Albert, C.M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E.J., Boehnke, M., Bonnycastle, L. L., Bottinger, E.P., Bowden, D.W., Bown, M.J., Chambers, J.C., Chan, J.C., Chasman, D., Cho, J., Chung, M.K., Cohen, B., Correa, A., Dabelea, D., Daly, M.J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P.T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S.J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M.M., Hultman, C.M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.J., Kim, Y.J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H.M., Kugathasan, S., Kwak, S.H., Laakso, M., Lehtimäki, T., Loos, R.J.F., Lubitz, S.A., Ma, R.C.W., MacArthur, D.G., Marrugat, J., Mattila, K.M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Meigs, J. B., Melander, O., Metspalu, A., Neale, B.M., Nilsson, P.M., O'Donovan, M.C., Ongur, D., Orozco, L., Owen, M.J., Palmer, C.N.A., Palotie, A., Park, K.S., Pato, C., Pulver, A.E., Rahman, N., Remes, A.M., Rioux, J.D., Ripatti, S., Roden, D.M., Saleheen, D., Salomaa, V., Samani, N.J., Scharf, J., Schunkert, H., Shoemaker, M.B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P.F., Suvisaari, J., Tai, E.S., Teo, Y.Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Watkins, H., Weersma, R.K., Wessman, M., Wilson, J.G., Xavier, R.J., Neale, B.M., Daly, M.J., MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434–443. https://doi.org/10.1038/s41586-020-2308-7.

Laricchia, K.M., Lake, N.J., Watts, N.A., Shand, M., Haessly, A., Gauthier, L., Benjamin, D., Banks, E., Soto, J., Garimella, K., Emery, J., Rehm, H.L., MacArthur, D. G., Tiao, G., Lek, M., Mootha, V.K., Calvo, S.E., 2022. Mitochondrial DNA variation across 56,434 individuals in gnomAD. Genome Res. 32, 569–582. https://doi.org/10.1101/gr.276013.121.

Lassuthova, P., Rebelo, A.P., Ravenscroft, G., Lamont, P.J., Davis, M.R., Manganelli, F., Feely, S.M., Bacon, C., Brožková, D.Š., Haberlova, J., Mazanec, R., Tao, F., Saghira, C., Abreu, L., Courel, S., Powell, E., Buglo, E., Bis, D.M., Baxter, M.F., Ong, R.W., Marns, L., Lee, Y.C., Bai, Y., Isom, D.G., Barro-Soria, R., Chung, K.W., Scherer, S.S., Larsson, H.P., Laing, N.G., Choi, B.O., Seeman, P., Shy, M.E., Santoro, L., Zuchner, S., 2018. Mutations in ATP1A1 cause dominant Charcot-Marie-tooth type 2. Am. J. Hum. Genet. 102, 505–514. https://doi.org/10.1016/j.ajhg.2018.01.023.

Li, H., 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Liu, X.X., Wang, N., Chen, Y.K., Lv, W.Q., Hong, J.M., Xu, G.R., Zhou, L.Y., Chen, W.J., Fan, D.S., He, J., 2023. Biallelic variants in the COQ7 gene cause distal hereditary motor neuropathy in two Chinese families. Brain. https://doi.org/10.1093/brain/awad040.

Martin, M., Patterson, M., Garg, S., Fischer, S.O., Pisanti, N., Klau, G.W., Schönhuth, A., Marschall, T., 2016. WhatsHap: fast and accurate read-based phasing. BioRxiv. https://doi.org/10.1101/085050.

Nguengang Wakap, S., Lambert, D.M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., Rath, A., 2020. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur. J. Hum. Genet. 28, 165–173. https://doi.org/10.1038/s41431-019-0508-0.

Nura Bio Initiates Phase I Clinical Trial for its Oral, Brain-Penetrant SARM1 Inhibitor, NB-4746 [WWW Document], 2023. https://nurabio.com/news/nura-bio-initiates-phase-1-clinical-trial-for-its-oral-brain-penetrant-sarm1-inhibitor-nb-4746/.

Osterloh, J.M., Yang, J., Rooney, T.M., Fox, A.N., Adalbert, R., Powell, E.H., Sheehan, A. E., Avery, M.A., Hackett, R., Logan, M.A., MacDonald, J.M., Ziegenfuss, J.S., Milde, S., Hou, Y.J., Nathan, C., Ding, A., Brown, R.H., Conforti, L., Coleman, M., Tessier-Lavigne, M., Züchner, S., Freeman, M.R., 2012. dSarm/Sarm1 is required for activation of an injury-induced axon death pathway. Science 1979 (337), 481–484. https://doi.org/10.1126/science.1223899.

Pedersen, B.S., Quinlan, A.R., 2017. Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. Am. J. Hum. Genet. 100, 406–413. https://doi.org/10.1016/j.ajhg.2017.01.017.

Pellerin, D., Danzi, M.C., Wilke, C., Renaud, M., Fazal, S., Dicaire, M.-J., Scriba, C.K., Ashton, C., Yanick, C., Beijer, D., Rebelo, A., Rocca, C., Jaunmuktane, Z., Sonnen, J. A., Larivière, R., Genís, D., Molina Porcel, L., Choquet, K., Sakalla, R., Provost, S., Robertson, R., Allard-Chamard, X., Tétreault, M., Reiling, S.J., Nagy, S., Nishadham, V., Purushottam, M., Vengalil, S., Bardhan, M., Nalini, A., Chen, Z., Mathieu, J., Massie, R., Chalk, C.H., Lafontaine, A.-L., Evoy, F., Rioux, M.-F., Ragoussis, J., Boycott, K.M., Dubé, M.-P., Duquette, A., Houlden, H., Ravenscroft, G., Laing, N.G., Lamont, P.J., Saporta, M.A., Schüle, R., Schöls, L., La Piana, R., Synofzik, M., Zuchner, S., Brais, B., 2023. Deep Intronic FGF14 GAA repeat expansion in late-onset cerebellar Ataxia. N. Engl. J. Med. 388, 128–141. https://doi.org/10.1056/nejmoa2207406.

Poplin, R., Ruano-Rubio, V., Depristo, M.A., Fennell, T.J., Carneiro, M.O., Van Der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M.J., Neale, B., Macarthur, D.G., Banks, E., 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. Biorxiv. https://doi.org/10.1101/201178.

Qiu, Y., Xiong, Y., Wang, L., Zhu, M., Tan, D., Hong, D., 2024. Homozygous variant in COQ7 causes autosomal recessive hereditary spastic paraplegia. Ann. Clin. Transl. Neurol. 11, 1067–1074. https://doi.org/10.1002/acn3.52037.

Rebelo, A., Tomaselli, P., Medina, J., Wang, Y., Dohrn, M., Nyvltova, E., Danzi, M., Garrett, M., Smith, S., Pestronk, A., Li, C., Ruiz, A., Jacobs, E., Feely, S., França, M., Gomes, M., Santos, D., Kumar, S., Lombard, D., Saporta, M., Hekimi, S., Barrientos, A., Weihl, C., Shy, M., Marques, W., Zuchner, S., 2023. Biallelic variants in COQ7 cause distal hereditary motor neuropathy with upper motor neuron signs. Brain 146, 4191–4199. https://doi.org/10.1093/brain/awad158.

Rebelo, A.P., Cortese, A., Abraham, A., Eshed-Eisenbach, Y., Shner, G., Vainshtein, A., Buglo, E., Camarena, V., Gaidosh, G., Shiekhattar, R., Abreu, L., Courel, S., Burns, D. K., Bai, Y., Bacon, C., Feely, S.M.E., Castro, Di, Peles, E., Reilly, M.M., Shy, M.E., Zuchner, S., 2021a. A CADM3 variant causes Charcot-Marie-tooth disease with marked upper limb involvement. Brain 144, 1197–1213. https://doi.org/10.1093/brain/awab019.

Rebelo, A.P., Eidhof, I., Cintra, V.P., Guillot-Noel, L., Pereira, C.V., Timmann, D., Traschütz, A., Schöls, L., Coarelli, G., Durr, A., Anheim, M., Tranchant, C., Van De Warrenburg, B., Guissart, C., Koenig, M., Howell, J., Moraes, C.T., Schenck, A., Stevanin, G., Züchner, S., Synofzik, M., 2021b. Biallelic loss-of-function variations in PRDX3 cause cerebellar ataxia. Brain 144, 1467–1481. https://doi.org/10.1093/brain/awab071.

Sadr, Z., Zare-Abdollahi, D., Rohani, M., Alavi, A., 2023. A founder mutation in COQ7, p. (Leu111Pro), causes pure hereditary spastic paraplegia (HSP) in the Iranian population. Neurol. Sci. https://doi.org/10.1007/s10072-023-06707-x.

Schuermans, N., Hemelsoet, D., Terryn, W., Steyaert, S., Van Coster, R., Coucke, P.J., Steyaert, W., Callewaert, B., Bogaert, E., Verloo, P., Vanlander, A.V., Debackere, E., Ghijsels, J., LeBlanc, P., Verdin, H., Naesens, L., Haerynck, F., Callens, S., Dermaut, B., Poppe, B., De Bleecker, J., Santens, P., Boon, P., Laureys, G., Kerre, T., 2022. Shortcutting the diagnostic odyssey: the multidisciplinary program for undiagnosed rare diseases in adults (UD-PrOZA). Orphanet J. Rare Dis. 17. https://doi.org/10.1186/s13023-022-02365-y.

Smolka, M., Paulin, L.F., Grochowski, C.M., Horner, D.W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S.W., Carvalho, C.M.B., Proukakis, C., Sedlazeck, F.J., 2024. Detection of mosaic and population-level structural variants with Sniffles2. Nat. Biotechnol. https://doi.org/10.1038/s41587-023-02024-y.

Tao, F., Beecham, G.W., Rebelo, A.P., Svaren, J., Blanton, S.H., Moran, J.J., Lopez-Anido, C., Morrow, J.M., Abreu, L., Rizzo, D., Kirk, C.A., Wu, X., Feely, S., Verhamme, C., Saporta, M.A., Herrmann, D.N., Day, J.W., Sumner, C.J., Lloyd, T.E., Li, J., Yum, S.W., Taroni, F., Baas, F., Choi, B.O., Pareyson, D., Scherer, S.S., Reilly, M.M., Shy, M.E., Züchner, S., 2019. Variation in SIPA1L2 is correlated with phenotype modification in Charcot– Marie– tooth disease type 1A. Ann. Neurol. 85, 316–330. https://doi.org/10.1002/ana.25426.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2024. Sambamba: Fast Processing of NGS Alignment Formats. https://doi.org/10.5281/zenodo.13200.

Turro, E., Astle, W.J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H.L., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O.S., Downes, K., Haimel, M., Tuna, S., Deevi, S.V.V., Aitman, T.J., Bennett, D.L., Calleja, P., Carss, K., Caulfield, M.J., Chinnery, P.F., Dixon, P.H., Gale, D.P., James, R., Koziell, A., Laffan, M.A., Levine, A.P., Maher, E.R., Markus, H.S., Morales, J., Morrell, N.W., Mumford, A.D., Ormondroyd, E., Rankin, S., Rendon, A., Richardson, S., Roberts, I., Roy, N.B.A., Saleem, M.A., Smith, K.G.C., Stark, H., Tan, R.Y.Y., Themistocleous, A. C., Thrasher, A.J., Watkins, H., Webster, A.R., Wilkins, M.R., Williamson, C., Whitworth, J., Humphray, S., Bentley, D.R., Abbs, S., Abulhoul, L., Adlard, J., Ahmed, M., Alachkar, H., Allsup, D.J., Almeida-King, J., Ancliff, P., Antrobus, R., Armstrong, R., Arno, G., Ashford, S., Attwood, A., Aurora, P., Babbs, C., Bacchelli, C., Bakchoul, T., Banka, S., Bariana, T., Barwell, J., Batista, J., Baxendale, H.E., Beales, P.L., Bentley, D.R., Bierzynska, A., Biss, T., Bitner-Glindzicz, M.A.K., Black, G.C., Bleda, M., Blesneac, I., Bockenhauer, D., Bogaard, H., Bourne, C.J., Boyce, S., Bradley, J.R., Bragin, E., Breen, G., Brennan, P., Brewer, C., Brown, M., Browning, A.C., Browning, M.J., Buchan, R.J., Buckland, M.S., Bueser, T., Diz, C.B., Burn, J., Burns, S.O., Burren, O.S., Burrows, N., Campbell, C., Carr-White, G., Carss, K., Casey, R., Chambers, Jenny, Chambers, John, Chan, M.M.Y., Cheah, C., Cheng, F., Chinnery, P.F., Chitre, M., Christian, M.T., Church, C., Clayton-Smith, J., Cleary, M., Brod, N.C., Coghlan, G., Colby, E., Cole, T.R.P., Collins, J., Collins, P.W., Colombo, C., Compton, C.J., Condliffe, R., Cook, S., Cook, H.T., Cooper, N., Corris, P. A.A., Furnell, A., Cunningham, F., Curry, N.S., Cutler, A.J., Daniels, M.J., Dattani, M., Daugherty, L.C., Davis, J., De Soyza, A., Deevi, S.V.V., Dent, T., Deshpande, C., Dewhurst, E.F., Dixon, P.H., Douzgou, S., Downes, K., Drazyk, A.M., Drewe, E., Duarte, D., Dutt, T., Edgar, J.D.M., Edwards, K., Egner, W., Ekani, M.N., Elliott, P., Erber, W.N., Erwood, M., Estiu, M.C., Evans, D.G., Evans, G., Everington, T., Eyries, M., Fassihi, H., Favier, R., Findhammer, J., Fletcher, D., Flinter, F.A., Floto, R.A., Fowler, T., Fox, J., Frary, A.J., French, C.E., Freson, K., Frontini, M., Gale, D.P., Gall, H., Ganesan, V., Gattens, M., Geoghegan, C., Gerighty, T.S.A., Gharavi, A.G., Ghio, S., Ghofrani, H.A., Gibbs, J.S.R., Gibson, K., Gilmour, K.C., Girerd, B., Gleadall, N.S., Goddard, S., Goldstein, D.B., Gomez, K., Gordins, P., Gosal, D., Gräf, S., Graham, J., Grassi, L., Greene, D., Greenhalgh, L., Greinacher, A., Gresele, P., Griffiths, P., Grigoriadou, S., Grocock, R.J., Grozeva, D., Gurnell, M., Hackett, S., Hadinnapola, C., Hague, W.M., Hague, R., Haimel, M., Hall, M., Hanson, H.L., Haque, E., Harkness, K., Harper, A.R., Harris, C.L.L., Hart, D., Hassan, A., Hayman, G., Henderson, A., Herwadkar, A., Hoffman, J., Holden, S., Horvath, R., Houlden, H., Houweling, A.C.C., Howard, L.S., Hu, F., Hudson, G., Hughes, J., Huissoon, A.P., Humbert, M., Humphray, S., Hunter, S., Hurles, M., Irving, M., Izatt, L., Johnson, S.A., Jolles, S., Jolley, J., Josifova, D., Jurkute, N., Karten, T., Karten, J., Kasanicki, M.A., Kazkaz, H., Kazmi, R., Kelleher, P., Kelly, A. M., Kelsall, W., Kempster, C., Kiely, D.G., Kingston, N., Klima, R., Koelling, N., Kostadima, M., Kovacs, G., Koziell, A., Kreuzhuber, R., Kuijpers, T.W., Kumar, A., Kumararatne, D., Kurian, M.A., Laffan, M.A., Lalloo, F., Lambert, M., Lawrie, A., Layton, D.M., Lench, N., Lentaigne, C., Lester, T., Levine, A.P., Linger, R., Longhurst, H., Lorenzo, L.E., Louka, E., Lyons, P.A., Machado, R.D., MacKenzie Ross, R.V., Madan, B., Maher, E.R., Maimaris, J., Malka, S., Mangles, S., Mapeta, R., Marchbank, K.J., Marks, S., Markus, H.S., Marschall, H.U., Marshall, A., Martin, J., Mathias, M., Matthews, E., Maxwell, H., McAlinden, P., McCarthy, M.I., McKinney, H., McMahon, A., Meacham, S., Mead, A.J., Castello, I.M., Megy, K., Mehta, S.G.G., Michaelides, M., Millar, C., Mohammed, S.N., Moledina, S., Montani, D., Moore, A.T., Morales, J., Morrell, N.W., Mozere, M., Muir, K.W., Mumford, A.D., Nemeth, A.H., Newman, W.G., Newnham, M., Noorani, S., Nurden, P., O'Sullivan, J., Obaji, S., Odhams, C., Okoli, S., Olschewski, A., Olschewski, H., Ong, K.R., Oram, S.H., Ormondroyd, E., Ouwehand, W.H., Palles, C., Papadia, S., Park, S.M., Parry, D., Patel, S., Paterson, J., Peacock, A., Pearce, S.H.H., Peden, J., Peerlinck, K., Penkett, C.J., Pepke-Zaba, J., Petersen, R., Pilkington, C., Poole, K.E.S., Prathalingam, R., Psaila, B., Pyle, A., Quinton, R., Rahman, S., Rankin, S., Rao, A., Raymond, F.L., Rayner-Matthews, P.J., Rees, C., Rendon, A., Renton, T., Rhodes, C.J., Rice, A.S.C., Richardson, S., Richter, A., Robert, L., Roberts, I., Rogers, A., Rose, S.J., Ross-Russell, R., Roughley, C., Roy, N.B.A., Ruddy, D.M., Sadeghi-Alavijeh, O., Saleem, M.A., Samani, N., Samarghitean, C., Sanchis-Juan, A., Sargur, R.B., Sarkany, R.N., Satchell, S., Savic, S., Sayer, J.A., Sayer, G., Scelsi, L., Schaefer, A.M., Schulman, S., Scott, R., Scully, M., Searle, C., Seeger, W., Sen, A., Sewell, W.A.C., Seyres, D., Shah, N., Shamardina, O., Shapiro, S. E., Shaw, A.C., Short, P.J., Sibson, K., Side, L., Simeoni, I., Simpson, M.A.A., Sims, M. C., Sivapalaratnam, S., Smedley, D., Smith, K.R., Smith, K.G.C., Snape, K., Soranzo, N., Soubrier, F., Southgate, L., Spasic-Boskovic, O., Staines, S., Staples, E., Stark, H., Stephens, J., Steward, C., Stirrups, K.E., Stuckey, A., Suntharalingam, J., Swietlik, E.M., Syrris, P., Tait, R.C., Talks, K., Tan, R.Y.Y., Tate, K., Taylor, J.M., Taylor, J.C., Thaventhiran, J.E., Themistocleous, A.C., Thomas, E., Thomas, D., Thomas, M.J., Thomas, P., Thomson, K., Thrasher, A.J., Threadgold, G., Thys, C., Tilly, T., Tischkowitz, M., Titterton, C., Todd, J.A., Toh, C.H., Tolhuis, B., Tomlinson, I.P., Toshner, M., Traylor, M., Treacy, C., Treadaway, P., Trembath, R., Tuna, S., Turek, W., Turro, E., Twiss, P., Vale, T., van Geet, C., van Zuydam, N., Vandekuilen, M., Vandersteen, A.M., Vazquez-Lopez, M., von Ziegenweidt, J., Noordegraaf, A.V., Wagner, A., Waisfisz, Q., Walker, S.M., Walker, N., Walter, K., Ware, J.S., Watkins, H., Watt, C., Webster, A.R., Wedderburn, L., Wei, W., Welch, S. B., Wessels, J., Westbury, S.K., Westwood, J.P., Wharton, J., Whitehorn, D., Whitworth, J., Wilkie, A.O.M., Wilkins, M.R., Williamson, C., Wilson, B.T., Wong, E. K.S., Wood, N., Wood, Y., Woods, C.G., Woodward, E.R.R., Wort, S.J., Worth, A., Wright, M., Yates, K., Yong, P.F.K., Young, T., Yu, P., Yu-Wai-Man, P., Zlamalova, E., Kingston, N., Walker, N., Penkett, C.J., Freson, K., Stirrups, K.E., Raymond, F.L., 2020. Whole-genome sequencing of patients with rare diseases in a national health system. Nature 583, 96–102. https://doi.org/10.1038/s41586-020-2434-2.

van der Ploeg, A.T., Reuser, A.J., 2008. Pompe's disease. Lancet 372, 1342–1353. https://doi.org/10.1016/S0140-6736(08)61555-X.

Wang, Y., Gumus, E., Hekimi, S., 2022. A novel COQ7 mutation causing primarily neuromuscular pathology and its treatment options. Mol. Genet. Metab. Rep. 31. https://doi.org/10.1016/j.ymgmr.2022.100877.

Xu, I.R., Danzi, M.C., Ruiz, A., Raposo, J., Arcia De Jesus, Y., Reilly, M.M., Cortese, A., Shy, M.E., Scherer, S.S., Herrmann, D., Fridman, V., Baets, J., Saporta, M., Seyedsadjadi, R., Stojkovic, T., Claeys, K.G., Patel, P., Feely, S., Rebelo, A.P., Consortium, Inherited Neuropathy, Dohrn, M.F., Zuchner, S., 2024. A study concept of expeditious clinical enrollment for genetic modifier studies in charcot-marie-tooth 1A. J. Peripheral Nervous System 29, 202–212. https://doi.org/10.1111/jns.12621.

Yalcouyé, A., Rebelo, A.P., Cissé, L., Rives, L., Bamba, S., Cogan, J., Esoh, K., Diarra, S., Ezell, K.M., Taméga, A., Guinto, C.O., Dohrn, M.F., Hamid, R., Fischbeck, K.H., Zuchner, S., Landouré, G., 2023. Novel variant in CADM3 causes Charcot-Marie-tooth disease. Brain Commun. 5. https://doi.org/10.1093/braincomms/fcad227.

Yun, T., Li, H., Chang, P.C., Lin, M.F., Carroll, A., McLean, C.Y., 2020a. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics 36, 5582–5589. https://doi.org/10.1093/bioinformatics/btaa1081.

Yun, T., Li, H., Chang, P.-C., Lin, M.F., Carroll, A., Mclean, C.Y., 2020b. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Biorxiv. https://doi.org/10.1101/2020.02.10.942086.

Zarate, S., Carroll, A., Krashenina, O., Sedlazeck, F.J., Jun, G., Salerno, W., 2018. Parliament 2: Fast Structural Variant Calling Using Optimized Combinations of Callers.

Ziaei Jam, H., Zook, J.M., Javadzadeh, S., Park, J., Sehgal, A., Gymrek, M., 2024. LongTR: genome-wide profiling of genetic variation at tandem repeats from long reads. Genome Biol. 25. https://doi.org/10.1186/s13059-024-03319-2.