

RESEARCH

Open Access



Meta-analysis of 46,000 germline *de novo* mutations linked to human inherited disease

Mónica Lopes-Marques¹, Matthew Mort², João Carneiro¹, António Azevedo^{3,4,5}, Andreia P. Amaro^{4,5}, David N. Cooper² and Luísa Azevedo^{4,5*}

Abstract

Background *De novo* mutations (DNMs) are variants that occur anew in the offspring of noncarrier parents. They are not inherited from either parent but rather result from endogenous mutational processes involving errors of DNA repair/replication. These spontaneous errors play a significant role in the causation of genetic disorders, and their importance in the context of molecular diagnostic medicine has become steadily more apparent as more DNMs have been reported in the literature. In this study, we examined 46,489 disease-associated DNMs annotated by the Human Gene Mutation Database (HGMD) to ascertain their distribution across gene and disease categories.

Results Most disease-associated DNMs reported to date are found to be associated with developmental and psychiatric disorders, a reflection of the focus of sequencing efforts over the last decade. Of the 13,277 human genes in which DNMs have so far been found, the top-10 genes with the highest proportions of DNM relative to gene size were *H3-3 A*, *DDX3X*, *CSNK2B*, *PURA*, *ZC4H2*, *STXPB1*, *SCN1A*, *SATB2*, *H3-3B* and *TUBA1A*. The distribution of CADD and REVEL scores for both disease-associated DNMs and those mutations not reported to be *de novo* revealed a trend towards higher deleteriousness for DNMs, consistent with the likely lower selection pressure impacting them. This contrasts with the non-DNMs, which are presumed to have been subject to continuous negative selection over multiple generations.

Conclusion This meta-analysis provides important information on the occurrence and distribution of disease-associated DNMs in association with heritable disease and should make a significant contribution to our understanding of this major type of mutation.

Keywords *De novo* mutations (DNMs), Genetic disease, Development, Neurodevelopmental disorders, Psychiatric disorders, Autism spectrum disorder

*Correspondence:

Luísa Azevedo

lazevedo@icbas.up.pt

¹CIIMAR-Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Porto, Portugal

²Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK

³CHUdSA-Centro Hospitalar Universitário de Santo António, Porto, Portugal

⁴UMIB-Unit for Multidisciplinary Research in Biomedicine, ICBAS - School of Medicine and Biomedical Sciences, University of Porto, Porto, Portugal

⁵ITR - Laboratory for Integrative and Translational Research in Population Health, Porto, Portugal



© The Author(s) 2024. **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.

Background

De novo mutations (DNMs) challenge traditional notions of Mendelian inheritance because the parents of affected offspring bearing DNMs are not themselves carriers [1–6]. In recent years, increasing numbers of DNMs have been identified as a consequence of the widespread adoption of whole exome/genome sequencing to screen patient cohorts.

In principle, there are two junctures at which such mutations can arise: (1) during gametogenesis in one of the parents, or (2) during the early divisions of embryogenesis. In the former instance, the mutation occurs in the germline of one of the parents and there is a tendency for the germline mutation rate to increase with age in both males and females [7–11], although DNMs originate more frequently in the paternal germline due to the comparatively high number of cell divisions occurring during spermatogenesis [6]. In the latter instance, by dint of their occurrence post-fertilization, the mutations are termed postzygotic DNMs [12]. The precise timepoint at which a mutation occurs during embryonic development is important for the establishment of the somatic mutational distribution pattern. Thus, if the mutation arises prior to primordial germline cell specification, it can be transmitted through the germline, resulting in recurrence of the disease in the next generation [13]. By contrast, if it arises after primordial germline cell specification, it will give rise to either mosaicism in the germline (which has the potential to result in disease recurrence) or mosaicism in the somatic tissues [8]. In contradistinction to germline mutations where paternal age has a considerable influence on the mutation rate [8, 14–16], currently available data are consistent with the absence of any parent-of-origin bias in relation to postzygotic mutations [17].

DNMs arise mainly through the action of endogenous processes mediated by the specific features and intrinsic properties of the genomic DNA sequence (e.g. methylation-mediated deamination of 5-methylcytosine, DNA sequence repetitivity, GC content, non-B DNA structures, recombination hotspots), chromosomal architecture (e.g. chromatin structure and interactions) and replication/repair errors [3, 17–19].

Our study, based on a large collection of germline DNMs, has explored the impact of these lesions on human inherited disease, with the specific aim of understanding their distribution and their key role in increasing the incidence of such disorders.

Methods

DNM dataset

A total of 443,508 germline disease-associated mutations (annotated as DM, DM?, DP and DFP [20]) were sourced from the Human Gene Mutation Database (HGMD

Professional v.2023.2), which includes a set of 46,489 putatively disease-causing DNMs from 13,277 genes. This constitutes a highly reliable source of germline DNMs due to the manual curation of the scientific literature related to human inherited disease [21]. Mutations were included in this DNM set if they were classified as “disease-causing mutations” (DM) or “probable/possible pathogenic mutations” (DM?) and had been annotated as DNMs by HGMD (reflecting the claims made by the authors in the original articles reporting them). The only exception was the prediction of the deleteriousness (described below) in which only DM were included.

Mapping of disease terms onto the Unified Medical Language System (UMLS)

Categorization of the disease-associated DNM set into high level disease concepts (e.g. developmental disorders or immune system disorders) was based on the Unified Medical Language System (UMLS) annotations [22] using a simple word permutation-based method. The disease names were mapped to UMLS concept identifiers (CUI) using the open source UMLS-Query module [23]. UMLS-Query provides a function called `mapToId`, which accepts a phrase and maps it to a CUI. A total of 39,125 (approx. 84% of the total) disease terms relating to DNMs were mapped to the UMLS with high confidence. The hierarchy of disease terms from the UMLS ontology was used to explore the relationships between the disease classes and DNMs. Using graph traversal in the UMLS Metathesaurus, a DNM could possibly (if appropriate) be associated with multiple high level disease classes (e.g. Primary sclerosing cholangitis is classed both as an ‘immune’ disorder and as a ‘digestive system’ disorder).

DNM enrichment analysis and Gene Ontology (GO) enrichment analysis

To identify disease genes enriched for DNMs, a relative DNM enrichment rate was calculated. The relative DNM enrichment rate allows for intergenic differences in coding sequence length and DNM frequency between specific genes to be taken into account and is defined as the fraction of the observed number of DNMs normalised with respect to the coding sequence length calculated on a gene wise basis:

$$\text{Relative mutability of DNMs} = \frac{\text{Number of DNMs for gene}}{\text{Coding sequence length of gene (bp)}}$$

Of the 13,249 genes (out of 13,277) from the DNM mutation set for which transcript information was available, we excluded genes with fewer than 5 DNMs (arbitrary cut-off; this excluded 11,105). For the remaining genes, the mean+1SD values of relative mutability were calculated (0.424+0.615=1.039), so that only genes with

a DNM enrichment rate greater than the mean+1SD were included in the analyses ($N=187$). For these genes, we also normalized the frequency of disease-associated *de novo* mutations by the estimates of the per gene mutations rates previously reported by Bethune and collaborators [24]. For this, the “expected_genovo_missense_corrected” values were used as missense variants represent the vast majority of DNM among the 187 enriched genes. The subset of 187 genes was then used for the analysis of biological processes using the DAVID Gene Ontology (GO) tool (<https://david.ncifcrf.gov/>).

Prediction of the functional impact of missense mutations

To predict the functional impact of mutations, the tool CADD (Combined Annotation Dependent Depletion) was employed [25]. The datasets were both normalised with respect to mutation type by selecting only missense mutations from each dataset. CADD predictions were calculated on two sets of HGMD missense disease-causing mutations (only DM mutations were included), viz. 5,307 mutations from the DNM set and 32,605 disease-causing mutations from HGMD (non-DNMs). This functional impact analysis was then repeated by using REVEL (Rare Exome Variant Ensemble Learner) prediction scores [26]. REVEL prediction scores were available for 5,506 mutations in the HGMD missense disease-causing DNM and 33,191 disease-causing non-DNMs from HGMD.

The above-mentioned strategy is graphically shown in Additional file 1.

Results

Frequency and distribution of DNM types among disease-associated mutations

A total of 443,508 germline disease-associated mutations were obtained from HGMD and subsequently analysed. Of these, 46,489 were identified as DNMs (from author-provided information), representing 10.5% of the total number of mutations in the sample (Fig. 1A). Missense replacements were found to be the most common type

of mutation among both DNMs and disease-associated DNMs not reported by *de novo* (non-DNMs), accounting for 56% and 46% of the listed mutations, respectively (Fig. 1B). One potentially interesting finding was the higher proportion of synonymous replacements noted among DNMs (13%) compared to just 1% for non-DNMs. Although this difference was statistically significant ($\chi^2(1, N=443,508)=3469, p<0.01$), it is likely to be artefactual, simply reflecting the criteria used for identifying and including DNMs rather than the underlying mechanisms driving these replacements. In the absence of mRNA phenotyping data, synonymous substitutions would normally be excluded from HGMD because there would be no direct and cogent evidence for their pathogenicity. By contrast, synonymous substitutions that occurred *de novo* would probably have been prioritized by the reporting authors because of the focus on DNMs being of pathological significance in the context of the various neurodevelopmental disorders under study. At the same time one cannot exclude the possibility that pathogenic synonymous substitutions would tend to be under ascertained in the context of non-DNMs as they often tend to go unreported in the context of molecular diagnostic testing.

Distribution of DNMs between disease concepts

Figure 2 presents the findings when UMLS disease concepts [22] were utilized to categorize the 46,489 DNMs annotated by HGMD. The majority of DNMs occurred in genes belonging to two predominant classificatory categories: “Developmental” disorders, accounting for 47% of DNMs, and “Psychiatric” disorders, comprising 32% of DNMs (Fig. 2A). It is important to note that owing to the nature of the inclusion criteria (by mapping DNMs to multiple high level classes for each disease concept), a single disease may be classified under multiple categories, resulting in overlaps between concepts. Nevertheless, the high prevalence of DNMs among developmental and psychiatric diseases is clear. In agreement with this assertion, the enrichment analysis (Fig. 2B) revealed log2-fold

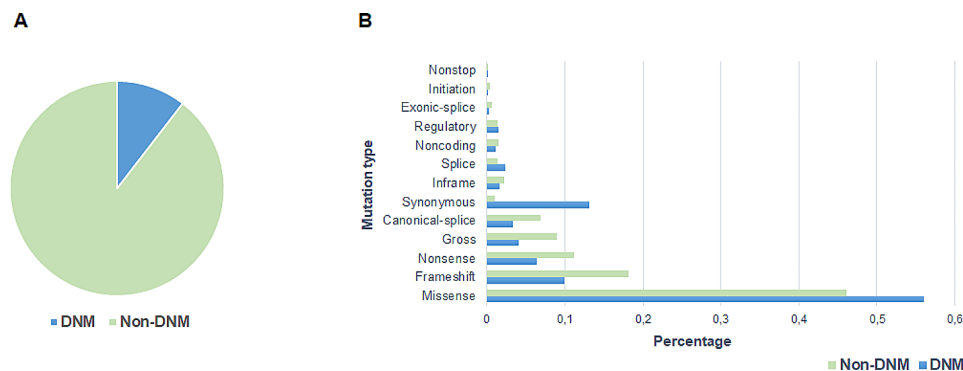


Fig. 1 Proportion of DNMs and non-DNMs in HGMD (A) and distribution of mutation types for both DNMs and non-DNMs (B)

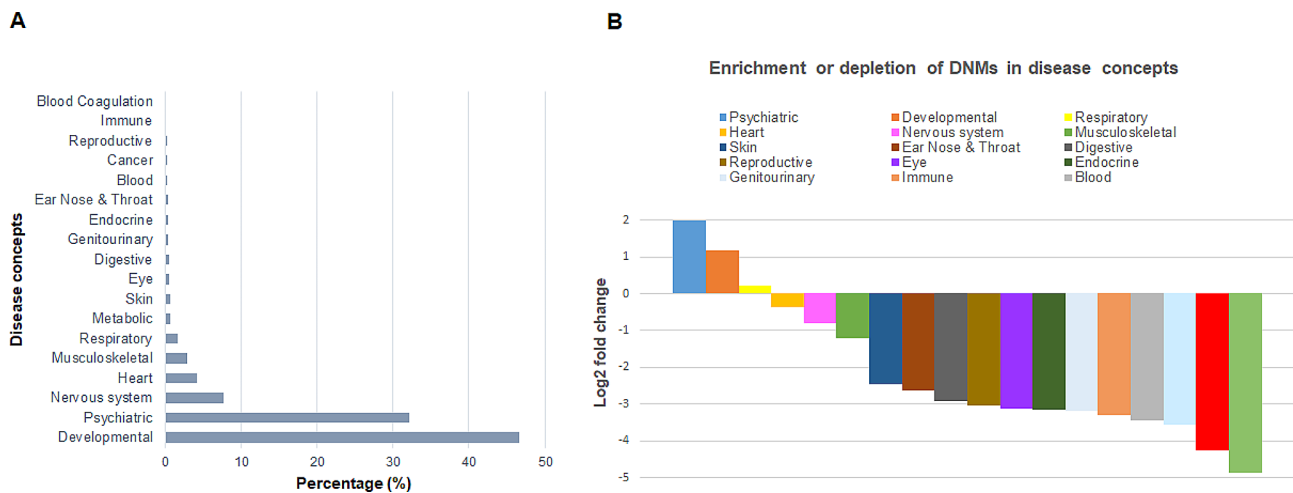


Fig. 2 Distribution of DNMs by disease concepts (A). Enrichment or depletion of DNMs in disease concepts (B)

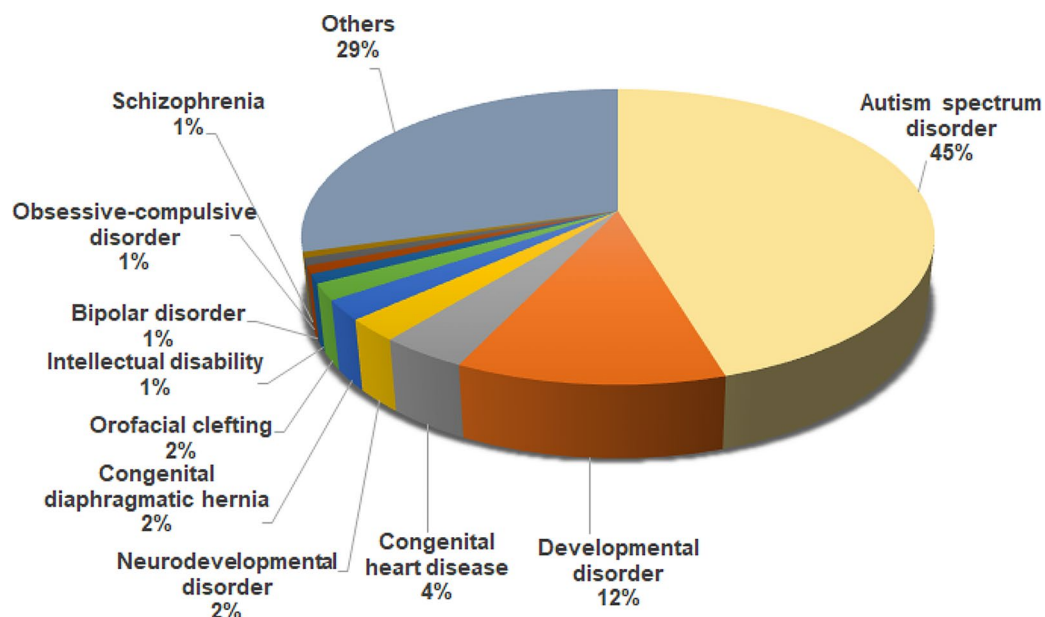


Fig. 3 Genetic diseases with the highest proportion of DNMs

changes of 2 and 1.2 for psychiatric and developmental concepts, respectively, highlighting a clear association between DNMs and these conditions that may have resulted, at least in part, from the considerable efforts that have been undertaken in recent years to unravel their genetic basis by whole genome sequencing or whole exome sequencing methodologies [3, 17, 27–35].

Next, the DNMs dataset was interrogated by disease term (Fig. 3). The most frequent term obtained was ‘autism’ reaching 45% of all DNMs. Autism spectrum disorder (ASD), the most frequent neurodevelopmental disorder in Western populations, is characterized by impaired social communication and interactions, and repetitive behavior [36]. The incidence of ASD has been estimated to be 60.38×10^4 according to the Global

Burden of Disease Study 2019 [37]. In terms of the molecular basis of autism spectrum disorders, and according to previous estimates, DNMs account for approximately one third of all cases ascertained [38]. This high proportion is probably due to a high proportion of DNMs being anticipated in ASD cohorts and because identifying a DNM in an individual with autism is generally held to be supportive of pathological authenticity (although by the very nature of this approach, there will probably also be a considerable number of false positives).

Congenital heart disease is another multi-gene phenotype that exhibits a high proportion of DNMs, with approximately 4% of all DNMs in our dataset associated with this condition, Other congenital phenotypes, such as orofacial clefting and congenital diaphragmatic

hernia, are also represented at a relatively high level in our dataset, each accounting for 2% of DNMs. These figures might reflect the fact that these birth defects are not only frequent in human populations but also that they have come under close molecular scrutiny by whole exome/genome sequencing in recent years [39–43]. About 29% of DNMs tagged in our analyses as belonging to the “Others” category, the disease terms with the highest number of DNMs were: developmental and epileptic encephalopathy, hydrocephalus, epilepsy, neurofibromatosis type 1, Dravet syndrome, Tourette syndrome, Coffin-Siris syndrome, Tetralogy of Fallot, periventricular nodular heterotopia and KBG syndrome.

Distribution of DNMs between and among disease-associated genes

Next, we examined the genes that harbored the highest numbers of disease-associated DNMs. The 20 genes with the highest number of DNMs accounted for only 5.8% of all the DNMs in our dataset (Additional file 2). The gene with the highest reported number of DNMs was *SCN1A* which encodes the sodium voltage-gated channel alpha subunit 1 involved in severe myoclonic epilepsy of infancy or Dravet syndrome [44, 45]. DNMs in the

SCN1A gene have been reported as a major cause of this disease [46, 47]. The second most common occurrence was observed for *ARID1B*, one of the genes underlying Coffin-Siris syndrome [48, 49]. It encodes a component of the SWI/SNF (BAF) chromatin remodeling complex which is essential for gene expression during development [50]. The *NFI* gene, known for some time to have a high mutation rate [51, 52], has one of the highest numbers of DNMs. This gene is responsible for neurofibromatosis type 1, a common autosomal dominant tumor predisposition syndrome [53–55], in which approximately half of the cases are caused by DNMs [56].

Two highly penetrant autism spectrum disorder genes [35], *SCN2A* and *SHANK3*, are represented among the top 20 genes with the highest number of DNMs. In addition, many of the genes shown in Additional file 2 (e.g. *SCN1A*, *ANKRD11*, *KMT2A*, *SYNGAP1*, *SATB2*, *CHD7*, *STXBPI*, *SHANK3*) have been shown to be associated with autism and other neurodevelopmental phenotypes (e.g. [57–63]). Because neurodevelopmental disorders share genetic risk genes and variants (inherited and *de novo*), they have been postulated to represent a continuum of etiological and genetic factors [64–66]. In fact, Ghiania and Faudez have proposed that impairments of specific windows of vulnerability during brain development may result in distinct disease entities with overlapping clinical symptoms [67].

Because gene complexity can contribute to the high number of mutations in any given gene, we investigated it by normalizing the number of DNMs by the coding length of the 187 genes enriched in DNMs (Table 1). To further contextualize our findings, we used estimates of per-gene mutation rates from Bethune and collaborators [24]. Although the coverage among the 187 genes was incomplete, we nevertheless observed a strong correlation between the two datasets (Additional file 2). Among the genes presented in Table 1, five (*DDX3X*, *STXBPI*, *SCN1A*, *SATB2*, *CTNNA1*) overlap with the top 20 genes with the highest number of DNMs (Additional file 2). This finding is consistent with previous research that has established a correlation between longer transcripts and genes that play a functional role at early developmental stages [68]. It is also important to note that genes associated with other phenotypes, such as the *SLC35A2* gene, associated with an inborn error of metabolism [2], are among the genes with the high proportions of DNMs.

GO enrichment analysis

We performed a GO analysis on biological processes for 187 disease genes enriched for DNMs (Additional file 2). This GO analysis identified DNM-enriched disease genes as being significantly enriched in 190 different types of biological process (e.g. system development or transcription related processes) (Additional file 3).

Table 1 Top 20 genes with the highest proportion of DNMs

Gene Symbol	Relative mutability (DNM frequency/coding length)*100	Most frequent clinical phenotype (HGMD)
<i>H3-3A</i>	7.056	Neurodegenerative disease
<i>DDX3X</i>	6.747	DDX3X syndrome
<i>CSNK2B</i>	6.636	Poirier-Bienvenu neurodevelopmental syndrome
<i>PURA</i>	6.295	PURA syndrome
<i>ZC4H2</i>	5.926	Arthrogyrosis multiplex congenita
<i>STXBPI</i>	5.243	Epileptic encephalopathy, early infantile
<i>SCN1A</i>	5.158	Dravet syndrome
<i>SATB2</i>	4.723	SATB2-associated syndrome
<i>H3-3B</i>	4.623	Neurodegenerative disease
<i>TUBA1A</i>	4.425	Tubulinopathy
<i>SLC35A2</i>	4.146	Congenital disorder of glycosylation
<i>CTNNA1</i>	4.007	Neurodevelopmental disorder
<i>VAMP2</i>	3.989	Axial hypotonia, intellectual disability and autism
<i>GNAO1</i>	3.944	Encephalopathy
<i>MECP2</i>	3.901	Rett syndrome
<i>DYRK1A</i>	3.665	Intellectual disability/Developmental disorder
<i>GATAD2B</i>	3.591	Neurodevelopmental disorder
<i>WDR45</i>	3.591	Neurodegeneration with brain iron accumulation
<i>SLC6A1</i>	3.556	Neurodevelopmental disorder
<i>H1-4</i>	3.333	Developmental disorder/Intellectual disability and distinct facial gestalt

The top 10 enriched clusters are shown in Table 2. The term GO:0048731 refers to system development which is the category that embraces a multitude of processes that together contribute to the formation and growth of an individual. It comprises not only nervous system development, (GO:0007399 with an enrichment of 3.7), but all other physiological systems. Other enriched GO terms are related to the regulation of transcription (GO:0045893, GO:1,903,508, GO:1,902,680, GO:0044767, GO:0010628).

Is there a tendency for pathogenic DNMs to be more deleterious than pathogenic mutations not reported to be *de novo*?

Because disease-associated DNMs are genetic changes that occur in the children of apparently healthy parents, they have not previously experienced negative selection, or at least only during the developmental time window from gametogenesis to adulthood in one generation. As a result, we speculate that DNMs might exert more detrimental effects than disease-associated mutations not reported to be *de novo* which are likely to have been exposed to negative selection for multiple/many generations since their inception [3, 69]. To investigate this postulate further, we first used the extensive collection of DNMs and disease-associated missense mutations available through HGMD not reported to be *de novo*, although we are aware that we cannot exclude the possibility that some non-DNMs have also occurred *de novo*, to ascertain the deleteriousness as measured by CADD scores. In line with our expectation, the CADD scores were found to be significantly higher for missense DNMs than for missense non-DNMs (t-test; $P < 2.2e^{-16}$) (Fig. 4A and

B). To further validate these findings, we also calculated the REVEL scores given their high performance with rare variants [26] (Additional file 4). As was observed for CADD scores, there was a statistically significant difference between the two sets (t-test; $P = 0.0374$), indicating that the DNMs set is enriched in missense mutations with greater impact on their protein products. This is consistent with the view that disease-associated variants not reported to be *de novo* have undergone multiple generations of negative selection thereby ensuring that those mutations with the greatest deleterious impact will have been lost from the population and hence would be less likely to contribute to future generations.

Discussion

Genome and exome sequencing efforts have revealed a high number of DNMs in genes related to human heritable disease. Germline disease-associated DNMs occur in parental germ cells and can be inherited by the offspring leading to a spectrum of health issues ranging from rare Mendelian diseases to complex traits. By using a large dataset of 46,489 DNMs reported in the literature and collected by HGMD, we observed that the most common disease category associated with DNMs is 'developmental disorder', possibly a consequence of efforts to sequence large cohorts of patients with these prevalent disorders. Neurodevelopmental disorders are associated with impairments of brain function [70–73] including intellectual disability, autism spectrum disorder, attention-deficit hyperactivity disorder, etc., Although recognized as discrete entities, they represent an interconnected genetic system [74], sharing etiological and genetic risk variants [64–66] that impair the functional

Table 2 Biological processes for 187 genes enriched in DNMs

Gene Ontology Term	Number of genes	Fold enrichment	P value	FDR
GO:0048731 system development	124	2.548578	3.01E-31	9.52E-28
GO:0007275 multicellular organism development	128	2.449113	4.52E-31	9.52E-28
GO:0007399 nervous system development	89	3.659853	1.31E-30	1.84E-27
GO:0044707 single-multicellular organism process	136	2.152079	3.94E-28	4.15E-25
GO:0048856 anatomical structure development	130	2.234834	1.25E-27	1.01E-24
GO:0045893 positive regulation of transcription, DNA-templated	71	4.31298	1.68E-27	1.01E-24
GO:1,903,508 positive regulation of nucleic acid-templated transcription	71	4.31298	1.68E-27	1.01E-24
GO:1,902,680 positive regulation of RNA biosynthetic process	71	4.290988	2.28E-27	1.19E-24
GO:0044767 single-organism developmental process	134	2.150094	2.53E-27	1.19E-24
GO:0010628 positive regulation of gene expression	79	3.696033	1.05E-26	4.43E-24

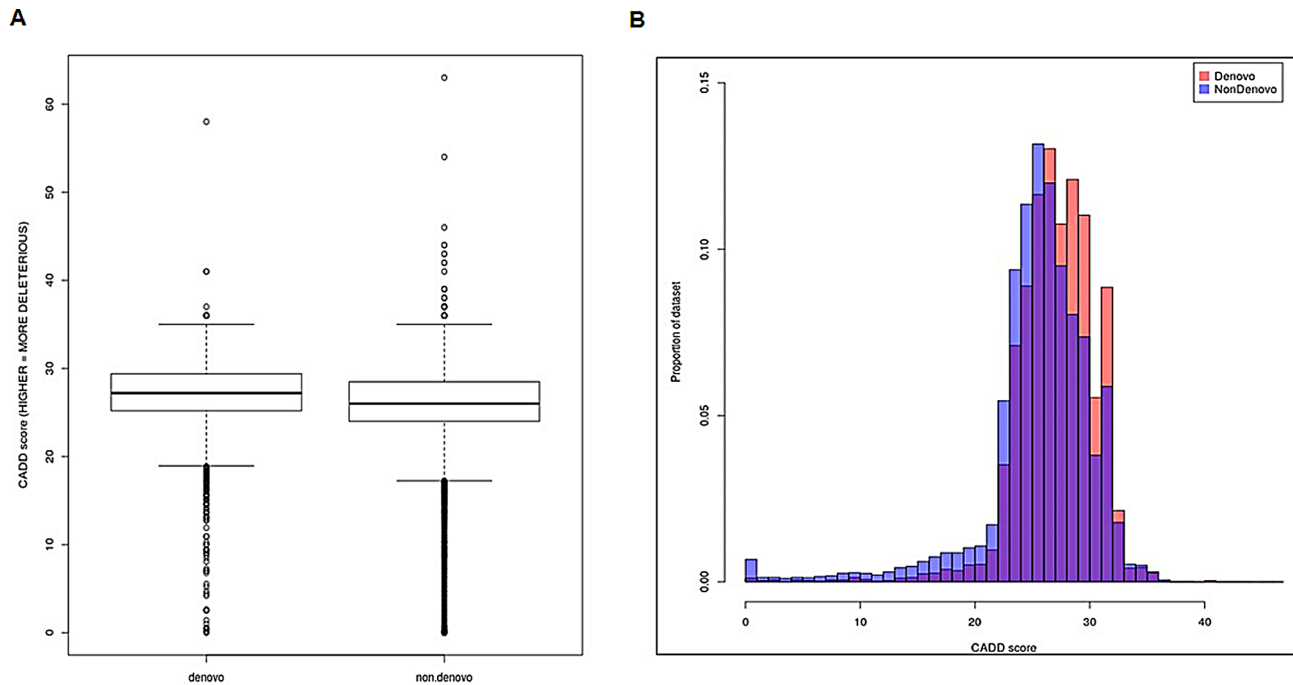


Fig. 4 Comparison of predicted deleteriousness of two groups of HGMD disease-causing missense mutations (5,307 DNM versus 32,605 non-DNM) (A). A purple bar indicates the proportion of dataset overlap between DNM and non-DNM for a specific range of CADD scores. The colour of the remaining stacked bar indicates an enrichment of a specific dataset at a specific CADD score; thus, red indicates enrichment of the DNM set whereas blue denotes enrichment of the non-DNM set (B)

integrity of brain-expressed genes related to molecular pathways such as protein synthesis, chromatin remodeling, transcriptional or epigenetic regulation and synaptic signaling [71, 75]. Disease-associated DNMs are intrinsically linked to developmental disorders [3, 17, 27–35], contributing to an estimated prevalence of 400,000 affected children born each year [27]. The most highly represented entity in the disease-associated DNM dataset analysed here was clearly autism. Although definitive evidence is lacking to confirm that locus heterogeneity is significantly higher for autism compared to other neurodevelopmental disorders, a plausible explanation could be the high prevalence of ASD in the general population and the efforts undertaken to sequence the exomes/genomes of affected individuals and their relatives. It is, however, important to note that the set of DNMs analyzed in this work include mutations classified as “DM”, which are clearly linked to the corresponding phenotype as inferred by the original publication, as well as mutations classified as “DM?”. However, these “DM?” variants represent an important source of information because this category of lesion poses a challenge for the interpretation of pathogenicity, which is important for distinguishing the genes that are causal from those that are coincidental. Interpreting the impact of these DNMs can be even more challenging when they occur in genes that have not previously been implicated in any disease [76]. An interesting example was recently reported by Jia and

collaborators in the *UBAP2L*, a gene that is involved in regulating stress granule formation during cortical development [77]. This neurodevelopmental disorder involves speech-language impairment, intellectual disability and behavioral problems.

With an average germline *de novo* mutation rate of 1.20×10^{-8} [78] (see also [79–81]), it is expected that an individual’s coding sequence will contain 1–2 DNMs [3, 82]. This low rate of spontaneous occurrence of novel mutations in an individual can be leveraged as a source of information in support of both gene and variant disease candidacy [83]. Whilst many DNMs are still waiting for the confirmation of causal genotype-phenotype linkage, the recurrence of DNMs in different cohorts, plus their absence from control datasets, provides good evidence for pathological authenticity.

A variety of strategies can be employed for the effective evaluation of the impact of individual DNMs prior to functional *in vitro* testing or analysis in cellular and animal models. For example, scanning protein sequence conservation scores is an important source of information, as it is widely accepted that proteins associated with human disease have been preferentially conserved through evolution [84–87]. In line with this notion, our analysis has shown that amino acid residues affected by DNMs tend to be associated with higher CADD and REVEL scores (Fig. 4 and Additional file 4). In principle, DNMs might also be screened using protein molecular

modelling tools and virtual screening [88] and the evaluation of each variant could be performed by free energy binding calculations and chemical descriptors [89]. Additional information might be obtained from other metrics such as the gene damage index [90]. This type of workflow is now possible and could be applied to a large number of DNMs. Very recently, a novel machine learning tool known as AlphaMissense [91] was introduced. This tool utilizes structural information predicted by AlphaFold2 to infer the pathogenicity of human variants, including DNMs, and could help in ranking these variants. Such screening techniques promise to be particularly important in the case of DNMs because, by their very nature, this type of genetic lesions lacks potentially supporting information provided by co-inheritance of the mutation and the clinical phenotype through multi-generational family pedigrees.

One limitation of our study relates to the fact that we used HGMD as a source of disease-associated DNMs. Although the HGMD data are the best available and most accurate source of deleterious DNMs, they do not allow one to consider recurrent DNMs at mutational hotspots. This could in principle impact the interpretation of our findings, although the extent of the impact is unpredictable. Future studies may add this new layer of information that while challenging in terms of data processing, would justify the effort expended in terms of the robustness of the results obtained.

Conclusions

DNMs appear anew at every generation and are clinically significant in the context of rare and common diseases alike. As the pace of genome sequencing increases, we anticipate a steady increase in the number of DNMs reported, and with it our understanding of the potential contribution of each newly arisen DNM to heritable disease, which is of the utmost importance to the medical genetics field. To the best of our knowledge, the meta-analyses we present here are the largest ever performed on disease-associated DNMs, and we expect that they can represent a gateway for further our understanding of this important category of gene lesions.

Abbreviations

DNM	<i>De novo</i> mutation
HGMD	Human Gene Mutation Database
DM	Disease-causing mutation
DM?	Probable/possible pathogenic mutation
UMLS	Unified Medical Language System
CADD	Combined Annotation Dependent Depletion
GO	Gene Ontology
ASD	Autism Spectrum Disorder
REVEL	Rare Exome Variant Ensemble Learner

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-024-00587-8>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4

Acknowledgements

JC acknowledges the FCT funding for his research contract established under the transitional rule of Decree Law 57/2016, amended by Law 57/2017.

Author contributions

LA conceived the idea of the study. MLM, MM, DNC and LA performed the main analyses and interpreted the data generated. LA and MM generated the Figures included in the manuscript. JC, AA and APA contributed to the analysis of the data. LA and DNC drafted the initial version of the manuscript. All authors participated in critical revisions and contributed to the final version. All authors reviewed the manuscript.

Funding

This research was supported by Fundação para a Ciência e a Tecnologia (FCT) to UMIB (UIDB/00215/2020 and UIDP/00215/2020), to ITR (LA/P/0064/2020) and to CIIMAR (UIDB/04423/2020 and UIDP/04423/2020), as well as by INCD funded by FCT and FEDER under the project 2021.09782.CPCA. LA and MLM are supported by FCT under the Scientific Employment Stimulus (CEECINST/00007/2021/CP2775/CT0002) and 2022.00397.CEECIND/CP1728/CT0006, respectively).

Data availability

All data analysed in this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 November 2023 / Accepted: 15 February 2024

Published online: 23 February 2024

References

1. Ku CS, Polychronakos C, Tan EK, Naidoo N, Pawitan Y, Roukos DH, Mort M, Cooper DN. A new paradigm emerges from the study of *de novo* mutations in the context of neurodevelopmental disease. *Mol Psychiatry*. 2013;18(2):141–53.
2. Quelhas D, Correia J, Jaeken J, Azevedo L, Lopes-Marques M, Bandeira A, Keldermans L, Matthijs G, Sturiale L, Martins E. SLC35A2-CDG: novel variant and review. *Mol Genet Metabolism Rep*. 2021;26:100717.
3. Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of *de novo* mutations in health and disease. *Genome Biol*. 2016;17(1):241.
4. Sevim Bayrak C, Zhang P, Tristani-Firouzi M, Gelb BD, Itan Y. *De novo* variants in exomes of congenital heart disease patients identify risk genes and pathways. *Genome Med*. 2020;12(1):9.
5. Azevedo L, Soares PA, Quental R, Vilarinho L, Teles EL, Martins E, Diogo L, Garcia P, Cenni B, Wermuth B, et al. Mutational spectrum and linkage

- disequilibrium patterns at the ornithine transcarbamylase gene (OTC). *Ann Hum Genet.* 2006;70(Pt 6):797–801.
6. Ohno M. Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences. *Genes Genet Syst.* 2019;94(1):13–22.
 7. Lindsay SJ, Rahbari R, Kaplanis J, Keane T, Hurles ME. Similarities and differences in patterns of germline mutation between mice and humans. *Nat Commun.* 2019;10(1):4053.
 8. Goldmann JM, Veltman JA, Gilissen C. De novo mutations reflect development and aging of the human germline. *Trends Genet.* 2019;35(11):828–39.
 9. Sasani TA, Pedersen BS, Gao Z, Baird L, Przeworski M, Jorde LB, Quinlan AR. Large, three-generation human families reveal post-zygotic mosaicism and variability in germline mutation accumulation. *Elife.* 2019;8:e46922.
 10. Goldmann JM, Wong WSW, Pinelli M, Farrah T, Bodian D, Stittrich AB, Glusman G, Vissers LELM, Hoischen A, Roach JC, et al. Parent-of-origin-specific signatures of de novo mutations. *Nat Genet.* 2016;48(8):935–9.
 11. Goldmann JM, Seplyarskiy VB, Wong WSW, Vilboux T, Neerincx PB, Bodian DL, Solomon BD, Veltman JA, Deeken JF, Gilissen C, et al. Germline de novo mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. *Nat Genet.* 2018;50(4):487–92.
 12. D’Gama AM, Walsh CA. Somatic mosaicism and neurodevelopmental disease. *Nat Neurosci.* 2018;21(11):1504–14.
 13. Jónsson H, Sulem P, Arnadóttir GA, Pálsson G, Eggertsson HP, Kristmundsdóttir S, Zink F, Kehr B, Hjorleifsson KE, Jónsson B, et al. Multiple transmissions of de novo mutations in families. *Nat Genet.* 2018;50(12):1674–80.
 14. Costa CIS, da Silva Campos G, da Silva Montenegro EM, Wang JYT, Scliar M, Monfardini F, Zachi EC, Lourenço NCV, Chan AJS, Pereira SL, et al. Three generation families: analysis of de novo variants in autism. *Eur J Hum Genet.* 2023;31(9):1017–22.
 15. Dubov T, Toledano-Alhadeh H, Bokstein F, Constantini S, Ben-Shachar S. The effect of parental age on the presence of de novo mutations - lessons from neurofibromatosis type I. *Mol Genet Genom Med.* 2016;4(4):480–6.
 16. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, van Duijn CM, Swertz M, Wijmenga C, van Ommen G, et al. Genome-wide patterns and properties of de novo mutations in humans. *Nat Genet.* 2015;47(7):822–6.
 17. Noyes MD, Harvey WT, Porubsky D, Sulovari A, Li R, Rose NR, Audano PA, Munson KM, Lewis AP, Hoekzema K, et al. Familial long-read sequencing increases yield of de novo mutations. *Am J Hum Genet.* 2022;109(4):631–46.
 18. Guiblet WM, Cremona MA, Harris RS, Chen D, Eckert KA, Chiaromonte F, Huang Y-F, Makova KD. Non-B DNA: a major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Res.* 2021;49(3):1497–516.
 19. Cooper DN, Bacolla A, Férec C, Vasquez KM, Kehrer-Sawatzki H, Chen JM. On the sequence-directed nature of human gene mutation: the role of genomic architecture and the local DNA sequence environment in mediating gene mutations underlying human inherited disease. *Hum Mutat.* 2011;32(10):1075–99.
 20. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet.* 2017;136(6):665–77.
 21. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, et al. (R): optimizing its use in a clinical diagnostic or research setting. *Hum Genet.* 2020;139(10):1197–207. The Human Gene Mutation Database (HGMD).
 22. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004;32(Database issue):D267–270.
 23. Shah NH, Muse MA. UMLS-Query: a perl module for querying the UMLS. *AMIA Annual Symposium Proceedings AMIA Symposium 2008, 2008:652–656.*
 24. Bethune J, Kleppe A, Besenbacher S. A method to build extended sequence context models of point mutations and indels. *Nat Commun.* 2022;13(1):7884.
 25. Rentsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886–94.
 26. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet.* 2016;99(4):877–85.
 27. Study DDD. Prevalence and architecture of de novo mutations in developmental disorders. *Nature.* 2017;542(7642):433–8.
 28. Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature.* 2014;515(7526):216–21.
 29. Howrigan DP, Rose SA, Samocha KE, Fromer M, Cerrato F, Chen WJ, Churchhouse C, Chambert K, Chandler SD, Daly MJ, et al. Exome sequencing in schizophrenia-affected parent-offspring trios reveals risk conferred by protein-coding de novo mutations. *Nat Neurosci.* 2020;23(2):185–93.
 30. Chen WX, Liu B, Zhou L, Xiong X, Fu J, Huang ZF, Tan T, Tang M, Wang J, Tang YP. De novo mutations within metabolism networks of amino acid/protein/energy in Chinese autistic children with intellectual disability. *Hum Genomics.* 2022;16(1):52.
 31. Järvelä I, Määttä T, Acharya A, Leppälä J, Jhangiani SN, Arvio M, Siren A, Kankuri-Tammilehto M, Kokkonen H, Palomäki M, et al. Exome sequencing reveals predominantly de novo variants in disorders with intellectual disability (ID) in the founder population of Finland. *Hum Genet.* 2021;140(7):1011–29.
 32. Brunet T, Jech R, Brugger M, Kovacs R, Alhaddad B, Leszinski G, Riedhammer KM, Westphal DS, Mahle I, Mayerhanser K, et al. De novo variants in neurodevelopmental disorders—experiences from a tertiary care center. *Clin Genet.* 2021;100(1):14–28.
 33. Wang W, Corominas R, Lin GN. De novo mutations from whole exome sequencing in neurodevelopmental and psychiatric disorders: from discovery to application. *Front Genet.* 2019;10:258.
 34. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, Peng M, Collins R, Grove J, Klei L, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell.* 2020;180(3):568–584e523.
 35. Zhou X, Feliciano P, Shu C, Wang T, Astrovskaya I, Hall JB, Obiajulu JU, Wright JR, Murali SC, Xu SX, et al. Integrating de novo and inherited variants in 42,607 autism cases identifies mutations in new moderate-risk genes. *Nat Genet.* 2022;54(9):1305–19.
 36. Rylaarsdam L, Guemez-Gamboa A. Genetic causes and modifiers of autism spectrum disorder. *Front Cell Neurosci.* 2019;13:385.
 37. Li Z, Yang L, Chen H, Fang Y, Zhang T, Yin X, Man J, Yang X, Lu M. Global, regional and national burden of autism spectrum disorder from 1990 to 2019: results from the global burden of Disease Study 2019. *Epidemiol Psychiatr Sci.* 2022;31:e33.
 38. Yoon S, Munoz A, Yamrom B, Lee Y-h, Andrews P, Marks S, Wang Z, Reeves C, Winterkorn L, Krieger AM, et al. Rates of contributory de novo mutation in high and low-risk autism families. *Commun Biology.* 2021;4(1):1026.
 39. Morton SU, Quiat D, Seidman JG, Seidman CE. Genomic frontiers in congenital heart disease. *Nat Reviews Cardiol.* 2022;19(1):26–42.
 40. Bishop MR, Diaz Perez KK, Sun M, Ho S, Chopra P, Mukhopadhyay N, Hetmanski JB, Taub MA, Moreno-Urbe LM, Valencia-Ramirez LC, et al. Genome-wide enrichment of de novo coding mutations in orofacial cleft trios. *Am J Hum Genet.* 2020;107(1):124–36.
 41. Bendixen C, Reutter H. The role of de novo variants in patients with congenital diaphragmatic hernia. *Genes.* 2021;12(9):1405.
 42. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, DePalma SR, McKean D, Wakimoto H, Gorham J, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science.* 2015;350(6265):1262–6.
 43. Jin SC, Homsy J, Zaidi S, Lu Q, Morton S, DePalma SR, Zeng X, Qi H, Chang W, Sierant MC, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet.* 2017;49(11):1593–601.
 44. Mulley JC, Scheffer IE, Petrou S, Dibbens LM, Berkovic SF, Harkin LA. SCN1A mutations and epilepsy. *Hum Mutat.* 2005;25(6):535–42.
 45. Cornejo-Sanchez DM, Acharya A, Bharadwaj T, Marin-Gomez L, Pereira-Gomez P, Nouel-Saied LM, University of Washington Center for Mendelian, Nickerson G, Bamshad DA, Mefford MJ et al. HC : SCN1A variants as the underlying cause of genetic epilepsy with febrile seizures plus in two multi-generational colombian families. *Genes* 13:2022.
 46. Claes L, Ceulemans B, Audenaert D, Smets K, Löfgren A, Del-Favero J, Ala-Mello S, Basel-Vanagaite L, Plecko B, Raskin S, et al. De novo SCN1A mutations are a major cause of severe myoclonic epilepsy of infancy. *Hum Mutat.* 2003;21(6):615–21.
 47. Sun H, Zhang Y, Liu X, Ma X, Yang Z, Qin J, Jiang Y, Qi Y, Wu X. Analysis of SCN1A mutation and parental origin in patients with Dravet syndrome. *J Hum Genet.* 2010;55(7):421–7.
 48. Tan Y, Chen J, Li Y, Liu Y, Wang Y, Xia S, Chen L, Wei W, Chen Z. Three novel ARID1B variations in coffin-Siris syndrome patients. *Neurol India.* 2022;70(5):2174–9.

49. van der Sluijs PJ, Jansen S, Vergano SA, Adachi-Fukuda M, Alanay Y, AlKindy A, Baban A, Bayat A, Beck-Wödl S, Berry K, et al. The ARID1B spectrum in 143 patients: from nonsyndromic intellectual disability to coffin–Siris syndrome. *Genet Sci.* 2019;21(6):1295–307.
50. Alver BH, Kim KH, Lu P, Wang X, Manchester HE, Wang W, Haswell JR, Park PJ, Roberts CWM. The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. *Nat Commun.* 2017;8(1):14648.
51. Clementi M, Barbuji G, Turolla L, Tenconi R. Neurofibromatosis-1: a maximum likelihood estimation of mutation rate. *Hum Genet.* 1990;84(2):116–8.
52. Huson SM, Compston DA, Clark P, Harper PS. A genetic study of Von Recklinghausen neurofibromatosis in South East Wales. I. Prevalence, fitness, mutation rate, and effect of parental transmission on severity. *J Med Genet.* 1989;26(11):704–11.
53. Wang W, Wei C-J, Cui X-W, Li Y-H, Gu Y-H, Gu B, Li Q-F, Wang Z-C. Impacts of NF1 gene mutations and genetic modifiers in neurofibromatosis type 1. *Front Neurol.* 2021;12:704639.
54. Karacanjic T, Whist E, Jamieson RV, Flaherty MP, Grigg JRB. Neurofibromatosis type 1: review and update on emerging therapies. *Asia-Pacific J Ophthalmol.* 2019;8(1):62–72.
55. Kehrer-Sawatzki H, Cooper DN. Challenges in the diagnosis of neurofibromatosis type 1 (NF1) in young children facilitated by means of revised diagnostic criteria including genetic testing for pathogenic NF1 gene variants. *Hum Genet.* 2022;141(2):177–91.
56. Bata Bashar M, Hodge David O, Mohney Brian G. Neurofibromatosis type 1: a population-based study. *J Pediatr Ophthalmol Strabismus.* 2019;56(4):243–7.
57. Casanova EL, Sharp JL, Chakraborty H, Sumi NS, Casanova MF. Genes with high penetrance for syndromic and non-syndromic autism typically function within the nucleus and regulate gene expression. *Mol Autism.* 2016;7(1):18.
58. O’Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature.* 2012;485(7397):246–50.
59. Sahin M, Sur M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. *Science.* 2015;350(6263):aab3897.
60. Papp-Hertelendi R, Tényi T, Hadzsiev K, Hau L, Benyus Z, Csábi G. First report on the association of SCN1A mutation, childhood schizophrenia and autism spectrum disorder without epilepsy. *Psychiatry Res.* 2018;270:1175–6.
61. Bucher M, Niebling S, Han Y, Molodenskiy D, Hassani Nia F, Kreienkamp H-J, Svergun D, Kim E, Kostyukova AS, Kreuzt MR, et al. Autism-associated SHANK3 missense point mutations impact conformational fluctuations and protein turnover at synapses. *eLife.* 2021;10:e66165.
62. Berryer MH, Hamdan FF, Klitten LL, Møller RS, Carmant L, Schwartztruber J, Patry L, Dobrzniecka S, Rochefort D, Neugnot-Cerolioli M, et al. Mutations in SYNGAP1 cause intellectual disability, autism, and a specific form of epilepsy by inducing haploinsufficiency. *Hum Mutat.* 2013;34(2):385–94.
63. Zhang R, He H, Yuan B, Wu Z, Wang X, Du Y, Chen Y, Qiu Z. An intronic variant of CHD7 identified in autism patients interferes with neuronal differentiation and development. *Neurosci Bull.* 2021;37(8):1091–106.
64. Owen MJ, O’Donovan MC. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. *World Psychiatry.* 2017;16(3):227–35.
65. Morris-Rosendahl DJ, Crocq M-A. Neurodevelopmental disorders—the history and future of a diagnostic concept. *Dialog Clin Neurosci.* 2020;22(1):65–72.
66. Consortium C-DGPG. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 2013;381(9875):1371–9.
67. Ghiani CA, Faundez V. Cellular and molecular mechanisms of neurodevelopmental disorders. *J Neurosci Res.* 2017;95(5):1093–6.
68. Lopes I, Altab G, Raina P, de Magalhães JP. Gene size matters: an analysis of gene length in the human genome. *Front Genet.* 2021;12:559998.
69. Mohiuddin M, Kooy RF, Pearson CE. De novo mutations, genetic mosaicism and human disease. *Front Genet.* 2022;13:983668.
70. Cardoso AR, Lopes-Marques M, Silva RM, Serrano C, Amorim A, Prata MJ, Azevedo L. Essential genetic findings in neurodevelopmental disorders. *Hum Genomics.* 2019;13(1):31.
71. Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental disorders: from genetics to functional pathways. *Trends Neurosci.* 2020;43(8):608–21.
72. Rodin RE, Dou Y, Kwon M, Sherman MA, D’Gama AM, Doan RN, Rento LM, Girsakis KM, Bohron CL, Kim SN, et al. The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat Neurosci.* 2021;24(2):176–85.
73. Liu Z, Zhang N, Zhang Y, Du Y, Zhang T, Li Z, Wu J, Wang X. Prioritized high-confidence risk genes for intellectual disability reveal molecular convergence during brain development. *Front Genet.* 2018;9:349.
74. Cristiano AS, Williams SM, Hawi Z, An JY, Bellgrove MA, Schwartz CE, Costa LF, Claudianos C. Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Mol Psychiatry.* 2014;19(3):294–301.
75. Cardoso AR, Lopes-Marques M, Oliveira M, Amorim A, Prata MJ, Azevedo L. Genetic variability of the functional domains of chromodomains helicase DNA-binding (CHD) proteins. *Genes.* 2021;12(11):1827.
76. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet.* 2012;13(8):565–75.
77. Jia X, Zhang S, Tan S, Du B, He M, Qin H, Chen J, Duan X, Luo J, Chen F, et al. De novo variants in genes regulating stress granule assembly associate with neurodevelopmental disorders. *Sci Adv.* 2022;8(33):eabo7112.
78. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. Rate of de novo mutations and the importance of father’s age to disease risk. *Nature.* 2012;488(7412):471–5.
79. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, et al. Variation in genome-wide mutation rates within and between human families. *Nat Genet.* 2011;43(7):712–4.
80. Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, Kronenberg ZN, Hormozdiari F, Raja A, Pennacchio LA, et al. Genomic patterns of de novo mutation in simplex autism. *Cell.* 2017;171(3):710–722e712.
81. Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. *Hum Mutat.* 2003;21(1):12–27.
82. Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, Griffing AR, Côté M, Henrion E, Spiegelman D, Tarabeux J, et al. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am J Hum Genet.* 2010;87(3):316–24.
83. Cesana M, Vaccaro L, Larsen MJ, Kibaek M, Micale L, Riccardo S, Annunziata P, Colantuono C, Di Filippo L, De Brasi D, et al. Integrated exome and transcriptome analysis prioritizes MAP4K4 de novo frameshift variants in autism spectrum disorder as a novel disease-gene association. *Hum Genet.* 2023;142(3):343–50.
84. Casanova EL, Switala AE, Dandamudi S, Hickman AR, Vandenbrink J, Sharp JL, Feltus FA, Casanova MF. Autism risk genes are evolutionarily ancient and maintain a unique feature landscape that echoes their function. *Autism Res.* 2019;12(6):860–9.
85. López-Bigas N, Ouzounis CA. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res.* 2004;32(10):3108–14.
86. Reiter LT, Potocki L, Chien S, Gribskov M, Bier E. A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res.* 2001;11(6):1114–25.
87. Azevedo L, Mort M, Costa AC, Silva RM, Quelhas D, Amorim A, Cooper DN. Improving the in silico assessment of pathogenicity for compensated variants. *Eur J Hum Genet.* 2016;25(1):2–7.
88. Serrano C, Teixeira CSS, Cooper DN, Carneiro J, Lopes-Marques M, Stenson PD, Amorim A, Prata MJ, Sousa SF, Azevedo L. Compensatory epistasis explored by molecular dynamics simulations. *Hum Genet.* 2021;140(9):1329–42.
89. Vieira TF, Magalhães RP, Simões M, Sousa SF. Drug repurposing targeting *Pseudomonas aeruginosa* Mvfr using docking, virtual screening, molecular dynamics, and free-energy calculations. *Antibiotics.* 2022;11(2):185.
90. Itan Y, Shang L, Boisson B, Patin E, Bolze A, Moncada-Vélez M, Scott E, Ciancanelli MJ, Lafaille FG, Markle JG, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc Natl Acad Sci USA.* 2015;112(44):13615–20.
91. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, Pritzel A, Wong LH, Zielinski M, Sargeant T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science.* 2023;381(6664):eadg7492.

Publisher’s Note

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