Operative list of genes associated with autism and neurodevelopmental disorders based on database review

Claire S. Leblond, Thuy-Linh Le, Simon Malesys, Freddy Cliquet, Anne-Claude Tabet, Richard Delorme, Thomas Rolland, Thomas Bourgeron

Operative list of genes associated with autism and neurodevelopmental disorders based on database review

Claire S Leblond[1], Thuy-Linh Le[1], Simon Malesys[1], Freddy Cliquet[1], Anne-Claude Tabet[1,2], Richard Delorme[1,3], Thomas Rolland[1]and Thomas Bourgeron[1]

[1] Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, Université de Paris, F-75015, Paris, France

[2] Genetics Department, Cytogenetic Unit, Robert Debré Hospital, APHP, F-75019, Paris, France.

[3] Child and Adolescent Psychiatry Department, Robert Debré Hospital, APHP, F-75019, Paris, France

Corresponding author : Claire S Leblond, claire.leblond@pasteur.fr, Thomas Bourgeron, thomas.bourgeron@pasteur.fr, Institut Pasteur, 25 rue du docteur Roux, 75015 Paris, France,

Keywords: autism, neurodevelopmental disorders, databases, high-confidence neurodevelopmental disorder genes, genetic diagnostic, fetal brain expression & synaptic function.

Abstract

The genetics of neurodevelopmental disorders (NDD) has made tremendous progress during the last few decades with the identification of more than 1,500 genes associated with conditions such as intellectual disability and autism. The functional roles of these genes are currently studied to uncover the biological mechanisms influencing the clinical outcome of the mutation carriers. To integrate the data, several databases and curated gene lists have been generated. Here, we provide an overview of the main databases focusing on the genetics of NDD, that are widely used by the medical and scientific communities, and extract a list of high confidence NDD genes (HC-NDD). This gene set can be used as a first filter for interpreting large scale omics dataset or for diagnostic purposes. Overall HC-NDD genes (N= 1,586) are expressed at very early stages of fetal brain development and enriched in several biological pathways such as chromosome organization, cell cycle, metabolism and synaptic function. Among those HC-NDD genes, 204 (12,9%) are listed in the synaptic gene ontology SynGO are enriched in genes expressed after birth in the cerebellum and the cortex of the human brain. Finally, we point at several limitations regarding the relatively poor standardized information available, especially on the carriers of the mutations. Progress on the phenotypic characterization and genetic profiling of the carriers will be crucial to improve our knowledge on the biological mechanisms and on risk and protective factors for NDD.

Introduction

The broad category of neurodevelopmental disorders (NDD) is characterized by impairment of brain functions leading to behavioral and cognitive abnormalities that impact on the intellectual, motor, language, and/or social abilities of the individuals (ICD-11 and DSM-5). NDD gather an heterogenous group of mental health conditions that occur during early period of development and include mainly intellectual disability (ID), learning disorders, motor coordination disorders, speech and language disorders, attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD). Overall, the prevalence of NDD ranges from 9 to 18% (Shen et al., 2018; Zablotsky et al., 2019), with usually males being more frequently affected than females.

Among NDD, ASD is characterized by social communication impairment and restricted/repetitive behaviors and interests (Lai et al., 2014). ASD accounts for 1-2% of the population worldwide (Baron-Cohen et al., 2009; Idring et al., 2012; Maenner et al., 2020; Saemundsen et al., 2013), and affects 3 males for 1 female, with a sex ratio bias for males more pronounced among individuals without ID. More than 70% of individuals with ASD display at least one co-occurring NDD and/or epilepsy (Gillberg, 2010; Saito et al., 2020).

ID is characterized by both impaired intellectual and adaptive functioning. Measured mostly with intellectual quotient tests using the Weschler scales, the cognitive abilities of individuals with ID are by definition below 70. The diagnosis of ID is made only when the patient also has deficits in two or more adaptive behaviors in their daily life. ID concerns 1-2% of the population with a biased sex-ratio towards males, which is negatively correlated with the cognitive impairment (Zablotsky et al., 2019, 2017).

Epilepsy is a group of neurological disorders that are characterized by two or more unprovoked seizures separated by more than 24 hours. Worldwide prevalence of epilepsy

ranges from 0.4 to 1% and there is gender differences in specific epilepsy syndromes, but overall the male / female ratio is very close to 1 (McHugh and Delanty, 2008). It is still unclear if specific subtypes of epilepsy can be considered as NDD, but early-onset epilepsies often co-occur with NDD. For example, the risk of epilepsy in ASD is increased seven-fold compared to the general population (8.6 vs. 1.2 %) (Thomas et al., 2017).

The heritability of NDD and epilepsy is high (e.g. ASD: >0.8 (Sandin et al., 2017, 2014; Taylor et al., 2020; Tick et al., 2016) ; Epilepsy: 0.25-0.7 (Kjeldsen et al., 2001; Miller et al., 1998; Speed et al., 2014)). The genetic architecture of NDD can vary from one patient to another. In some cases, one single *de novo* mutation can be causative whereas in other cases, a complex interplay between rare and frequent genetic variants increases the risk of developing a NDD (Bourgeron, 2015). Recent studies indicate that distinct categories of NDD share some genetic risk factors with distinct phenotypic expression (Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu and Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019; The Brainstorm Consortium et al., 2018; Zarrei et al., 2019). The overlap at the genetic level was expected since patients diagnosed with different NDD also share comorbid disorders (Hansen et al., 2018; Sokolova et al., 2017). Altogether it is now accepted that NDD is a wide continuum of genetic and phenotypic dimensions instead of restricted categories.

Molecular diagnosis of NDD was empowered by the relatively recent high-throughput genotyping/sequencing approaches. Using these technologies allows to replicate findings initially obtained from medical case-reports and to detect new genes associated with NDD via the identification of rare and penetrant mutations including structural variants (SV) such as copy-number variants (CNV), tandem repeats, indels and point mutations. Zarrei and colleagues reported a recent large data resource of CNVs across NDD. Causative CNVs were identified in 10.5 % of subjects with NDD including 11.4% among individuals with ASD and

9.4% with ADHD (Zarrei et al., 2019). Interestingly, 4.4% out of the 10.5% of the individuals with NDD were carriers of new relevant CNVs while the remaining individuals were carriers of aneuploidies or known genomic disorder variants. The contribution of whole-exome sequencing in the diagnostic of NDD has dramatically increased the yields of positive genetic diagnosis for NDD. A recent meta-analysis of 30 articles estimated that the yield of WES positive finding was 36% (95%CI: 30-43%) overall for NDD, 31% (95%CI: 25-38%) for isolated NDD, and 53% (95%CI: 41-64%) for the NDD plus associated conditions (such as Rett-like features) (Srivastava et al., 2019). More precisely, a meta-analysis focused on 103 studies (epilepsy, N = 72; ASD, N = 14; ID, N = 21) across 32,331 individuals provided estimates of diagnostic yield for ASD (17.1%; 95%CI: 11-25%), epilepsy (24%; 22-27%), and ID (28.2%; 22-35%) (Stefanski et al., 2020). For some genes recurrently mutated in patients with NDD, the link between the mutations, the mechanisms and the clinical profiles are better understood, but for the large majority of the genes, little clinical information on the carriers of mutations is available and more multidisciplinary collaborative work is warranted.

Several lists of NDD genes and databases have been designed to inform the scientific and medical communities about gene- or variation-disease pairs and the biological functions of these genes. They also help clinicians to provide the most accurate NDD diagnostic and genetic counselling for patients and their parents (Savatt and Myers, 2021) (Table 1). This chapter provides an overview of all the current databases, the genes associated with NDD as well as their main biological functions. Notably, we show that high-confidence NDD genes (described below) are expressed at very early stages of fetal brain development and are enriched in several biological pathways such as chromosome organization, cell cycle, metabolism and synaptic functions.

Curated databases for genetic studies on NDD

Genes associated with NDD were first identified through targeted studies, e.g. medical case reports and region-based or candidate gene approaches, and more recently from large-scale genome wide studies, including whole genome SNP genotyping array and whole exome/genome sequencing. Some databases are dedicated specifically to autism research (e. g. SFARI) or ID (e.g. SysID) but others cover the broad category of NDD (e.g. DBD) or more generally genetic diseases (e. g. ClinVar, OMIM).

Gene database from the SFARI

To date, the largest database that focuses on ASD was developed by the Simons Foundation Autism Research Initiative (SFARI). First launched in 2008, this database, called the SFARI Gene, has been systematically updated alongside the published scientific literature and became a reliable source of expertly curated information on ASD-associated genes. The database provides a scoring system, based on the presence of *de novo* likely-gene-disrupting mutations, replication studies, functional effects of the mutations and the inheritance patterns. Reported genes are thus classified based on the confidence levels for ASD risk. There are currently four ranking categories, including S - syndromic, 1 - high-confidence, 2 - strong candidate and 3 - suggestive evidence (Abrahams et al., 2013).

A large-scale project named SPARK — for Simons Foundation Powering Autism Research for Knowledge — aims to collect and analyze clinical and genetic data of up to 50,000 families with a member diagnosed with ASD. SPARK has developed a partnership with not only researchers but also with certified clinical biologists, that allows SPARK to return genetic findings to participants who consented to receive their results. The SPARK gene list, which comprises genomic regions and genes in which *de novo* pathogenic variants have been identified in at least three affected individuals and reported in at least two independent

publications, has been judged to be of sufficient scientific evidence for a clinical utility. This SPARK gene list is updated every 6 months.

The ClinVar database and its link with ClinGen project

Clinical Genome (ClinGen) is a US state-funded project that has recently taken a step forward especially in the curation of gene-disease associations. ClinGen seeks to centralize published genomic and phenotypic data related to NDD, among additional disorders, for both scientific and clinical use. After the concertation of a Gene Curation Expert Panel, an annotation is given for each gene of interest, ranging from definitive, strong to moderate or limited evidence in support of its relationship with diseases. Genes falling within the definitive or strong categories are finally determined as disease associated genes (Gene-Disease Validity).

ClinVar points at specific genetic variants previously detected in patients with human diseases and provides interpretations of their significance to diseases. To date, more than 500,000 variants (including germline and somatic variants) are listed in ClinVar (Landrum et al., 2020). Submitted variants and variant-condition pairs are aggregated in ClinVar, with indications as to whether there is consensus or disagreement among submitters on variant interpretations. For each submission, ClinVar provides a rating system from one to four stars, that corresponds to the "review status" approved by the ClinGen steering committee. The three/four stars rating is considered as high confidence and corresponds to review status for which the variant-condition pair interpretation is from an "expert panel" or a "practice guideline" assertion, respectively.

The SysID database

SysID, or Systems Biology Approaches to Intellectual Disability, is a database of genes implicated in ID, collected from the literature and the Online Mendelian Inheritance in Man

(OMIM) database. The SysID gene catalogue is maintained by the Institute of Human Genetics at Erlangen (Germany) for the diagnosis of isolated and syndromic ID, as well as multisystemic disorders, within which ID is one of the main clinical features. The SysID database includes a list of "primary" ID genes restricted to genes with high level of evidence (criteria are described by Kochinke and colleagues)(Kochinke et al., 2016). This list is further classified into 9 categories according to the clinical phenotype (syndromic or non-syndromic ID), its severity (from severe to moderated or very variable cognitive impairment), the penetrance (full or incomplete) and the additional clinical features (with or without malformations of organs, brain or limb). A second list of "candidate" ID genes is provided by the SysID database and gathers genes that are not confirmed by a sufficient number of patients carrying these mutations and/or by the robustness of the clinical information. Of interest, SysID database also provides additional data such as gene ontologies, mutation inheritance and disease subtypes.

The DDG2P database

Gene2Phenotype (G2P) is a database for many disorders including cancer, eye disorders, skin illness and NDD. This database is available from the DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER) website (Firth et al., 2009) and was first established by a systematic review of the literature, diagnostic tests by clinical laboratories in the UK and in the UniProt database (The UniProt Consortium, 2015). The Developmental Disorder gene panel of G2P (DDG2P) has been regularly updated and enriched using the results of the Deciphering Developmental Disorders (DDD) study. Genetic data of more than 13,400 exomes of individuals with severe NDD were analyzed using bioinformatic filtering and identified variants were reviewed by clinicians and molecular geneticists, all senior experts in the field. Once validated, the genetic result was returned to the patient using the regional genetic services, while the variant was submitted to the G2P database for further

queries. In addition to gene classification based on NDD risk, G2P also provides information about the relevant phenotypes, tissue specificity, molecular mutation consequences and mode of inheritance. Here, we selected genes relevant for NDD as those with organ specificity restricted to the annotation's "brain" or "cognition".

The DBD database

The Developmental Brain Disorder (DBD) gene database focuses on six disorders, including ID, ASD, ADHD, bipolar disorder, schizophrenia and epilepsy. A list of disease-causative genes has been generated by manual curation and classified according to the number of reported loss-of-function (LOF) mutations (>3, 2, or 1) and the mode of inheritance (*de novo* or inherited). DBD also provides a list of high-confidence genes with autosomal recessive inheritance (for more information, see Gonzalez-Mantilla and Colleagues (Gonzalez-Mantilla et al., 2016)). We restricted the DBD database to the most convincing disease-causing genetic variants, with missense variants, deletions encompassing more than one gene, and individuals with multiple LOF variants being excluded, in order to increase the specificity of the genes in the list. Interestingly, the DBD database highlights the phenotypic heterogeneity of developmental brain disorders, in which similar variations in a gene might cause different phenotypes (Myers et al., 2020).

The OMIM database

OMIM is the online version of Mendelian Inheritance in Man (MIM), a catalogue of mendelian traits and disorders that was initiated by Professor V. A. McKusick in the 1960s. OMIM provides information on phenotype and associated genotype of most of the reported mendelian disorders. The database contains information on more than 7000 disorders including approximately 700 related to NDD.

The genetic databases dedicated to epilepsy

Regarding epilepsy, a frequent comorbidity of patients with autism, few databases are available. For this review, we used *The Lafora Gene Mutation Database* (Ianzano et al., 2005), *The Epilepsy Genetic Association Database* (epiGAD) (Tan and Berkovic, 2010), *CarpeDB*, *EpilepsyGene*, *GenEpi* and *MeGene* (the two last are no longer maintained). Finally, we also included the gene list used by the Mayo Clinic Laboratories in the context of genetic diagnostic of epilepsy.

Table 1: Databases reporting gene-disease pairs related to NDD

Gene list	Phenotype	Gene classification	N of genes	HC-NDD genes	Candidates NDD genes	Website and where to found selection criteria
SPARK	ASD	SPARK	157	Yes	No	https://spark-sf.s3.amazonaws.com/SPARK_gene_list.pdf
SFARI	ASD	Category S* - Syndromic	84	No	Yes	https://gene.sfari.org/about-gene-scoring/
		Category 1 - High Confidence	194	Yes	No	
		Category 2 - Strong Candidate	207	No	Yes	
		Category 3 - Suggestive Evidence	507	No	Yes	
SysID	ID	Primary	1396	Yes	No	https://sysid.cmbi.umcn.nl/
		Candidates	1218	No	Yes	
DDG2P	DD	confirmed (Brain or Cognition)	722	Yes	No	https://www.ebi.ac.uk/gene2phenotype/terminology
		confirmed (other organ specificity)	505	No	Yes	
		probable	570	No	Yes	
		possible	291	No	Yes	
		both RD and IF	57	No	Yes	
		child IF	10	No	Yes	
DBD	ID, ASD, ADHD, SCZ, BD, & Epilepsy	Tier 1	152	Yes	No	https://dbd.geisingeradmi.org/
		Tier 2	70	No	Yes	
		Tier 3	108	No	Yes	
		Tier 4	156	No	Yes	
		AR	115	Yes	No	
HC-NDD	**NDD & Epilepsy**		**1586**			**https://genetrek.pasteur.fr/**

SPARK, Simons Foundation Powering Autism Reseach for Knowledge; SFARI, Simons Foundation Autism Research Initiative; SysID, Systems Biology Approaches to Intellectual Disability; DDG2P, developmental disorder genotype to phenotype; DBD, developmental brain disorders; ASD, Autism Spectrum Disorder; ID, Intellectual Disability; DD, Developmental Disorder; ADHD, Attention Deficit and Hyperactivity Disorder; SCZ, Schizophrenia; BD, Bipolar Disorder; RD, Relevant Disease; IF, incidental finding; AR, genes with Autosomal Recessive inheritance. * Category S mentioned here is not overlapping with categories 1 to 3 of SFARI genes. High Confidence NeuroDevelopmental Disorders (HC-NDD) genes or candidate-NDD genes are described in the main text, in the figure 1 and can be seen on https://genetrek.pasteur.fr/.

Specificity and common features of the genetic database for NDD

In this section, we first selected genes based on their high-confidence for their association with NDD (Supplementary Table 1; see also our website https://genetrek.pasteur.fr/). We then extracted a list of candidate genes for NDD, which are not yet robustly associated with NDD, but which include genes with (i) weak confidence from the NDD-related databases, (ii) high/specific expression in the brain or (iii) under strong selective pressure. These two lists of highly confident and candidate NDD genes can be used for filtering variants in exome/genome of patients and for exploring genes from transcriptomics/proteomics experiments. To centralise the gene set lists from all databases, we developed "GeneTrek" https://genetrek.pasteur.fr/, a website that can be queried in order to assess how your gene or your list of genes of interest are mapped across all databases. GeneTrek also provides links to other websites adding information on gene expression, genetic variations and possible contacts with clinical geneticists and researchers investigating the genes through the Human Disease Genes website series (HDG) (Dingemans et al., 2021).

High-confidence NDD genes

Several genetic databases such as SFARI Gene, SPARK, SysID, DDG2P and DBD provide key information to classify genes into categories that are useful for clinicians. Indeed, for diagnostic purposes, only genes with clear evidence for association can be used. Here, for the high confidence NDD gene list (HC-NDD genes), we selected all genes from the category 1 genes from the SFARI, the SPARK, the primary ID gene list from SysID, the "confirmed" gene list from the DDG2P (with "Brain" or "Cognition" as "organ specificity"), and the Tier 1 or autosomal recessive (AR) inheritance list from the DBD, gathering a total of 1,586 genes (Figure 1 & GeneTrek website).

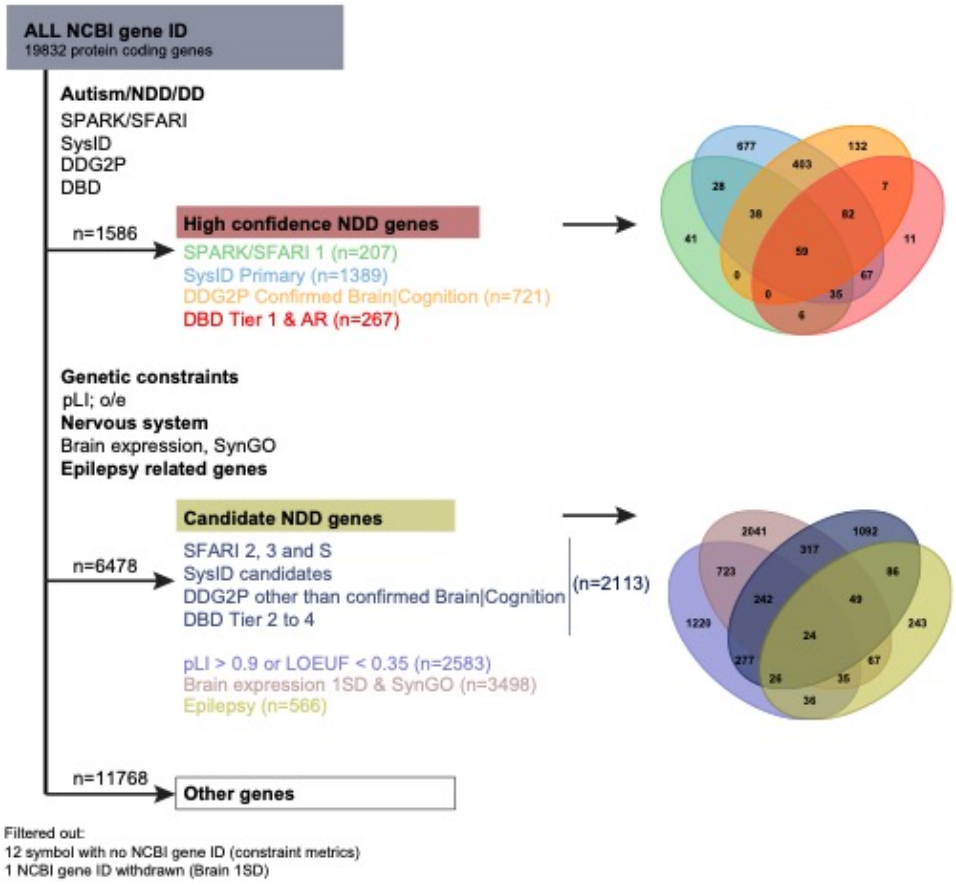

Figure 1. Selection of the high confidence and candidate NDD genes

The high confidence NDD (HC-NDD) genes were extracted from the following databases: SPARK/SFARI category 1, SysID primary, G2P DD confirmed and Brain|Cognition, and, DBD Tier1 and AR categories. The candidate genes are those from the SPARK/SFARI category 2, 3 and S, the SysID candidates, the DBD Tier 2 to 4, the G2P DD others than categories "confirmed" or "Brain|Cognition". To complete the candidate NDD gene list, genes under strong genetic constraints (pLI > 0.9 or LOEUF < 0.35), genes showing strong or specific expression (above one standard deviation) in fetal and/or adult human brain and/or genes encoding synaptic proteins from the Synaptic Gene Ontologies, or, genes related to epilepsy were added. The selection of genes was focused on protein coding gene only. ASD, Autism Spectrum Disorder; NDD, NeuroDevelopmental Disorder; DD, Developmental Disorder; SPARK, Simons Foundation Powering Autism Research for Knowledge; SFARI, Simons Foundation Autism Research Initiative; SysID, Systems Biology Approaches to Intellectual Disability; DDG2P, Development Disorders Gene2Phenotype; DBD, Developmental Brain Disorder; pLI, probability of loss of function intolerance; LOEUF, upper bound of the oe confidence interval with oe standing for observed / expected number of loss-of-function variants.

Candidate NDD genes

This list of 6,478 genes represents candidate risk genes for NDD (Figure 1), but with currently less confidence in their causality. The list can be used as a second filter to detect potential new genes for NDD or to reinforce the association of previously reported genes. It includes the genes in the categories "2", "3" and "syndromic" from the SFARI, the candidates ID genes from the SysID, those from the "confirmed" with organ specificity other than Brain or Cognition, "probable", "possible", "relevant disease" and "incidental finding" DDG2P categories, and, those from the Tier 2 to Tier 4 of the DBD list. We also included the genes related to epilepsy (Dunn et al., 2020; International League Against Epilepsy Consortium on Complex Epilepsies, 2018; Pfisterer et al., 2020; Wang et al., 2017) that were not listed as high confident genes in DDG2P and SysID. In addition, we included in the NDD candidate gene list, the genes displaying high or specific expression (above one standard deviation) in fetal and/or adult human brain and genes encoding synaptic proteins from the Synaptic Gene Ontologies (SYNGO) consortium (Koopmans et al., 2019). Finally, we included genes under strong selective pressure using two metrics that were developed from exome and genome sequencing of the Genome aggregation Database (gnomAD) to estimate intolerance to loss of function mutations: the probability of being loss-of-function intolerant score ($pLI > 0.9$)(Lek et al., 2016) or/and the observed/expected (oe) number of loss-of-function variants in that genes (LOEUF or upper bound of the oe confidence interval < 0.35) (Karczewski et al., 2020). The probability of loss-of-function intolerance (pLI) measures the probability for each gene to be tolerant to loss-of-function mutations, e.g. stop gains, splice variants or frameshifts, based on the observation of such mutations in control populations and weighted by the gene GC content and sequencing coverage (Lek et al., 2016). A cut-off of $pLI>0.9$ is classically used to define genes that are intolerant to such mutations. Similarly to the pLI, the loss-of-function observed/expected upper fraction (LOEUF) describes the intolerance of each gene to loss-of-

function mutations, based on the observed and expected numbers of such mutations in control populations, indicating strong selection against loss-of-function mutations in each gene (cut-off of LOEUF<0.35) (Karczewski et al., 2020).

Overlap between the databases

As illustrated in Figure 1 and table 2, a limited number of HC-NDD genes (n=59; 3.7%) are shared by the SPARK/SFARI, SysID, DDG2P and DBD gene lists, highlighting shared genetic pathways among NDD phenotypes, but also different selection criteria depending on the database (Gonzalez-Mantilla et al., 2016; Myers et al., 2020). This further points out the importance of considering a broad spectrum of genes for patients with NDD instead of a too restricted list of genes (Myers et al., 2020). A number of genes could be more related to a given disorder. For example, 41 genes seem to be more associated to ASD than with other conditions, such as *SHANK2*, *NRXN2/3* and *NLGN2*. However, their involvement in other NDDs cannot be excluded and requires further investigations. Within the large list of 6,478 candidate NDD genes, 24 (see Figure 1 and Table 2) pass all selection criteria – candidate genes in the NDD database, expressed in the brain and highly intolerant to loss of function mutations – making them strong candidate for being robustly associated with NDD in the future. There is finally a third category that includes 11,768 protein-coding genes in the human genome that are neither in the high-confidence nor candidate NDD gene lists. It is worth mentioning that the lack of scientific evidence supporting the role of such genes in NDD phenotypes is not definitive. Gene curation is still in progress.

Table 2: Overlapping HC-NDD or Candidate-NDD genes across databases

List of genes	Names	GO enrichment
Overlapping in HC-NDD genes (n=59)	*ADNP*, *ADSL*, *AHDC1*, *ANKRD11*, *ARID1B*, *AUTS2*, *BCL11A*, *CASK*, *CDKL5*, *CHAMP1*, *CHD2*, *CHD7*, *CREBBP*, *CTCF*, *CTNNB1*, *DDX3X*, *DNMT3A*, *DYRK1A*, *EBF3*, *EHMT1*, *EP300*, *FOXG1*, *FOXP1*, *GRIN2B*, *HIVEP2*, *HNRNPU*, *IQSEC2*, *KANSL1*, *KMT2A*, *KMT5B*, *MAGEL2*, *MECP2*, *NIPBL*, *NSD1*, *PCDH19*, *PHIP*, *POGZ*, *PTCHD1*, *PTEN*, *RAI1*, *SCN1A*, *SCN2A*, *SETBP1*, *SETD5*, *SLC6A1*, *SLC9A6*, *SON*, *SRCAP*, *STXBP1*, *SYNGAP1*, *TBCK*, *TBR1*, *TCF20*, *TCF4*, *TSC2*, *UBE3A*, *UPF3B*, *VPS13B*, *WAC*	- Positive regulation of transcription from RNA polymerase II promoter - Negative regulation of transcription from RNA polymerase II promoter - Positive regulation of transcription, DNA-templated - DNA methylation - Transcription, DNA-templated
Overlapping in Candidate-NDD genes (n=24)	*ADGRL2*, *AKAP6*, *ARNT2*, *ATP5F1A*, *CACNG2*, *CSNK1E*, *DIP2C*, *DNAJC6*, *FASN*, *GABRB1*, *GLRA2*, *GNAQ*, *KCNA4*, *KCND3*, *KCNV1*, *LGI1*, *PDE10A*, *PLXNA1*, *RBFOX1*, *RHEB*, *TENM2*, *UBR5*, *XPO1*, *ZNF423*	- Protein homo-oligomerization

Epilepsy-associated genes are underlined; GO: Gene Ontology

The biological pathways associated with the HC-NDD genes

Chromosomal location and mode of inheritance of HC-NDD genes

HC- NDD genes are distributed over all chromosomes (Figure 2). There are 1,452 genes located on the autosomes, 129 on the X chromosome, and five on the mitochondrial genome (*ATP6*, *COX1*, *COX2*, *COX3* and *ND5*). There is no enrichment of HC-NDD genes on one specific autosome, but a significant enrichment on the X chromosome (8.2% for HC-NDD genes vs 3.9% for the non HC-NDD genes ; Pearson's Chi-squared = 63.539, df = 1, p-value = 1.572e-15). A majority of the HC-NDD genes (N=968; 61%) display a recessive mode of inheritance, while 590 show a dominant mode of inheritance and 28 a complex (e.g. mitochondrial and imprinting) or uncertain mode of inheritance.

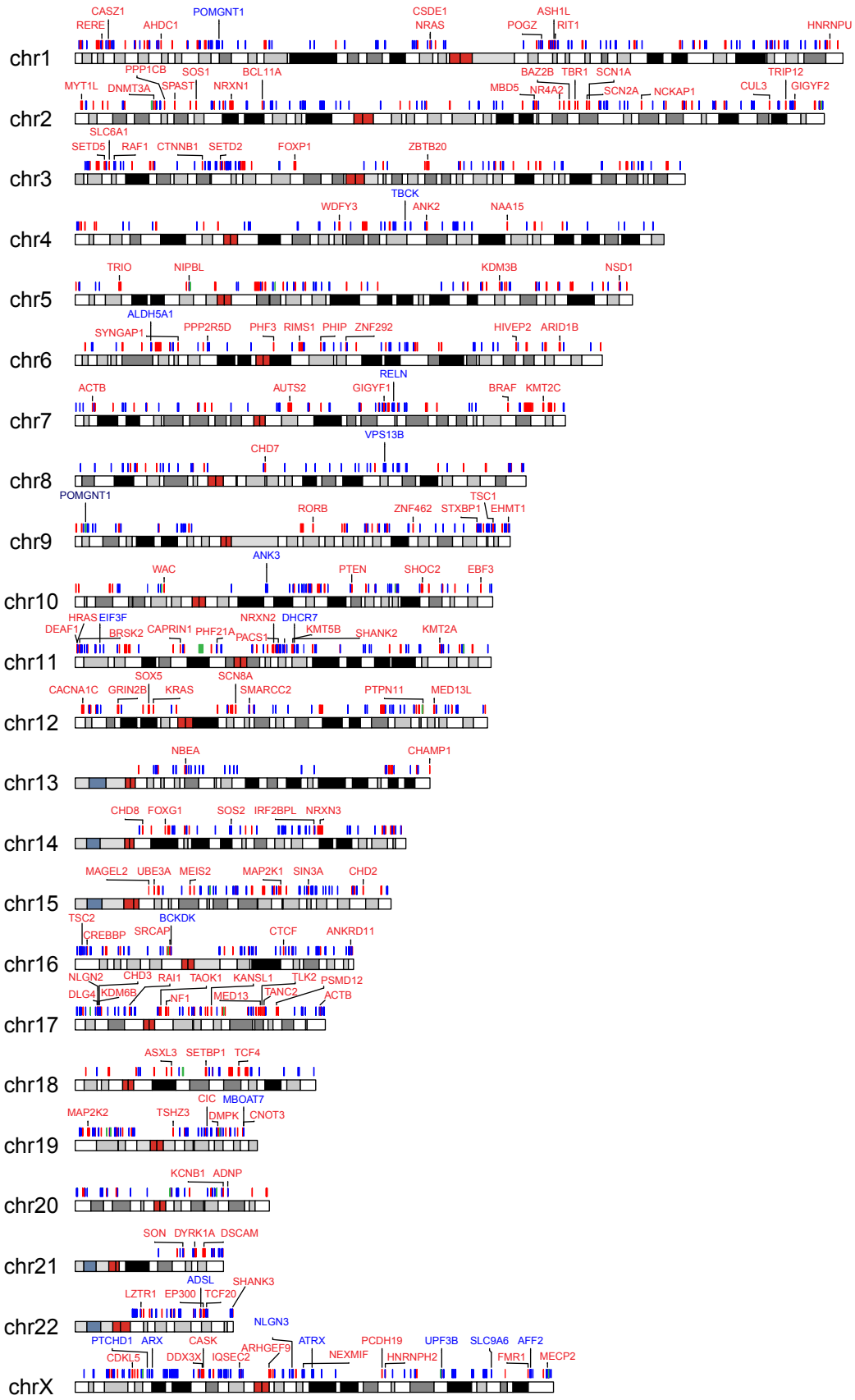

Figure 2. Distribution of the high confidence NDD genes on the human chromosomes

The 1,586 HC-NDD genes are indicated on the human chromosomes. The pattern of inheritance associated to HC-NDD genes is specified in color (red: dominant, blue: recessive & green: complex). Only the symbols of the SPARK genes are indicated above the chromosomes.

Selective constraints on HC-NDD genes in humans and during evolution

Genes associated with NDD are on average under strong selective pressure during evolution (Dumas et al., 2019) and intolerant to loss of function mutations in human population (Kosmicki et al., 2017). However, there are some interesting exceptions of HC-NDD genes showing strong signs of positive selective pressure. Examples of HC-NDD genes with protein divergence during primate evolution are *AHI1*, *AHSG*, *BRCA1*, *CNTNAP4*, *CSPP1*, *FAN1*, *FAM149B1*, *GLB1*, *NHS*, *TCTN1*. Some of them genes are associated with microcephaly, questioning the potential role of human-specific variants within these genes in the expansion of the human brain size (Dumas et al., 2021). In humans, 68% of the HC-NDD genes with a dominant inheritance model (403/590) are strongly intolerant to loss of function mutations considering pLI or LOEUF metrics.

Spatio-temporal expression of the HC-NDD genes in the human brain

To further characterize the HC-NDD genes, we performed a Specific Expression Analysis (SEA) on the human Brainspan collection (Dougherty et al., 2010). This provides a parallel analysis of transcripts enriched in particular human brain regions, and/or developmental periods. HC-NDD gene list was divided into two non-overlapping set of genes i) HC-NDD genes not listed in SynGO (N= 1,382 genes) and ii) HC-NDD genes listed in SynGO (N=204 genes). As illustrated in the figure 3, the HC-NDD genes not listed in SynGO are enriched in the very early fetal period (amygdala, cortex, hippocampus, striatum). In contrast, the HC-NDD genes with synaptic functions and listed in SynGO are more expressed after birth in the cerebellum and the cortex. Example of the SPARK genes for ASD expressed in specific regions

are *TBR1* and *SOX5* in early fetal cortex, *FOXP1* in the early mid fetal striatum, *SCN2A* and *STX1B* in the cerebellum and *KCNB1* and *SCN8A* in the cortex after birth.

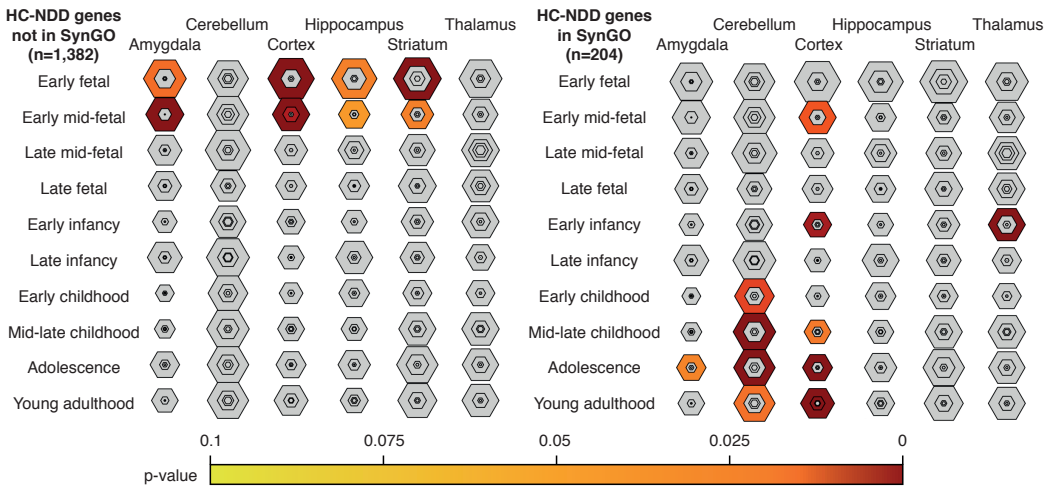

Figure 3. Specific Expression Analysis (SEA) across brain regions and development of the high confidence NDD genes.

For each brain regions and/or development stage, transcripts that are specifically expressed or enriched were shown using Bullseye plots. The enrichment was identified by Fisher's exact test with Benjamini-Hochberg (BH) correction and the BH corrected p-values are illustrated on the bottom of the figure. **The p-value for Specificity Index (pSI) is a probability threshold used to measure the specificity of the expression of genes in different regions or cell types (Dougherty et al., 2010). High pSI values (pSI>0.01) allow to measure the enrichment among larger sets of expressed genes while lower pSI values (pSI<0.01) allow to measure enrichment in smaller sets of genes specifically expressed in some regions or cell types.** Varying stringencies for enrichment in pSI thresholds are represented by the size of the hexagons going from least specific lists (outer hexagons with p-value threshold = 0.05) to most specific (center with p-value threshold = 0.0001). The regions most enriched in HC-NDD genes not listed in SynGO (left panel, for a pSI < 0.05 and after BH correction) are early fetal amygdala (q= 0.015), early mid fetal amygdala (q=8.070e-07), early fetal cortex (q=3.369e-04), early mid fetal cortex (q=3.369e-04), early fetal hippocampus (q= 0.027), early mid fetal hippocampus (q= 0.045), early fetal striatum (q=3.538e-06) and early mid fetal striatum (q= 0.027). The regions most enriched in HC-NDD genes listed in SynGO (right panel, for a pSI < 0.05 and after BH correction) are adolescence amygdala (q= 0.029), early childhood cerebellum (q= 0.009), middle late childhood cerebellum (q= 8.699e-05), adolescence cerebellum (q= 5.843e-06), young adulthood cerebellum (q= 0.017), early mid fetal cortex (q= 0.011), neonatal early infancy cortex (q= 0.003), middle late childhood cortex (q= 0.022), adolescence cortex (q= 5.817e-10), young adulthood cortex (q= 1.786e-16) and neonatal early infancy thalamus (q= 8.699e-05). Early fetal: before 12 pcw ; Early mid fetal 13-18 pcw. q = Benjamini-Hochberg corrected p value.

Biological functions of the high confidence NDD genes

To investigate main biological functions of the HC-NDD genes, we performed a gene ontology enrichment analysis and mapped a set of enriched biological processes onto the high-confidence protein-protein interaction network connecting the proteins encoded by HC-NDD genes (Figure 4). The functional enrichment points at groups of proteins involved in different

biological processes such as cell cycle (adjusted p = $7.6x10^{-10}$), mitochondrion organization (adjusted p = $1.5x10^{-3}$), chromosome organization (adjusted p = $2.5x10^{-14}$), transport (adjusted p = $1.2x10^{-14}$), transmembrane transport (adjusted p = $1.5x10^{-8}$) and nervous system development (adjusted p = $2.8x10^{-50}$). As indicated in figure 4B, genes listed in the "Nervous system/synapse" are associated with ASD, ID or both. Interestingly, there are more than 20 genes listed as part of the NLGN-NRXN pathway previously associated with ASD. Within this pathway, there are the genes coding for cell adhesion molecules (e.g. NLGN, NRXN), glutamate pre- and post-synaptic receptors (GRIN2A, GRIN2D, GRIA4), and scaffolding proteins at glutamatergic synapse (SHANK).

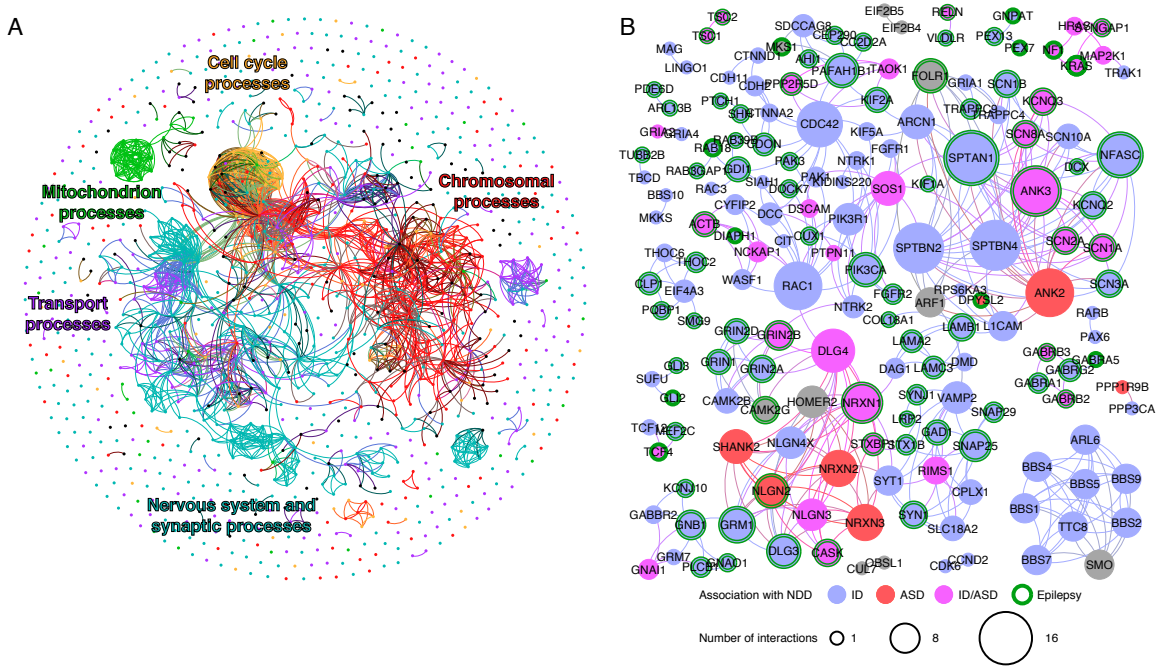

Figure 4. Protein-protein interaction network and functional annotation of HC-NDD genes.
(A) Pairs of proteins consist of physical interactions (experimental or database) from STRING with the highest confidence score (score > 0.9). Edge color represents interactions involving proteins that are part of selected enriched gene ontology terms. We merged the gene ontology terms into 5 large groups "Cell cycle", "Mitochondrion", "Chromosome/chromatin", "Nervous system/synapse" and "Transport". (B) Network of interacting proteins annotated in the "Nervous system/synapse" group in (A). Node colors represent the association of the genes with ASD, ID or both, node border color the association with epilepsy, and node size the number of interactions of each gene. Genes with a SFARI score of 1 and genes from the SPARK gene list were considered as ASD-associated, genes from the SysID database with primary evidence were considered as ID-associated, and genes from any epilepsy-related gene list were considered as epilepsy-associated.

Finally, we investigated further the roles of the HC-NDD genes at the synapse. We use the ontology terms provided by the SynGO consortium (Koopmans et al., 2019) (Figure 5, supplementary Table 2 and Table 3). Among the HC-NDD genes, 204 (12,9%) are mapped to SynGO annotated genes. Overall, the exploration of over-represented synaptic gene ontologies reveals 18 cellular component terms and 20 biological processes significantly enriched at 1% false discovery rate. Interestingly, there is an enrichment in both pre- (N=99) and post-synaptic (N=109) components. The synaptic processes that are enriched in the HC-NDD genes include ligand- and voltage-gated ion channel activity involved in regulation of pre-synaptic membrane potential as well as synaptic vesicle cycle and exocytosis and regulation of neurotransmitter receptor localization and levels to postsynaptic specialization membranes. Altogether, these cellular components and functional maps stress further the role of HC-NDD proteins in both pre-and postsynaptic glutamatergic synapse. A more precise map linking biological mechanisms to symptoms would require to have more information on the clinical trajectory of individuals with NDD.

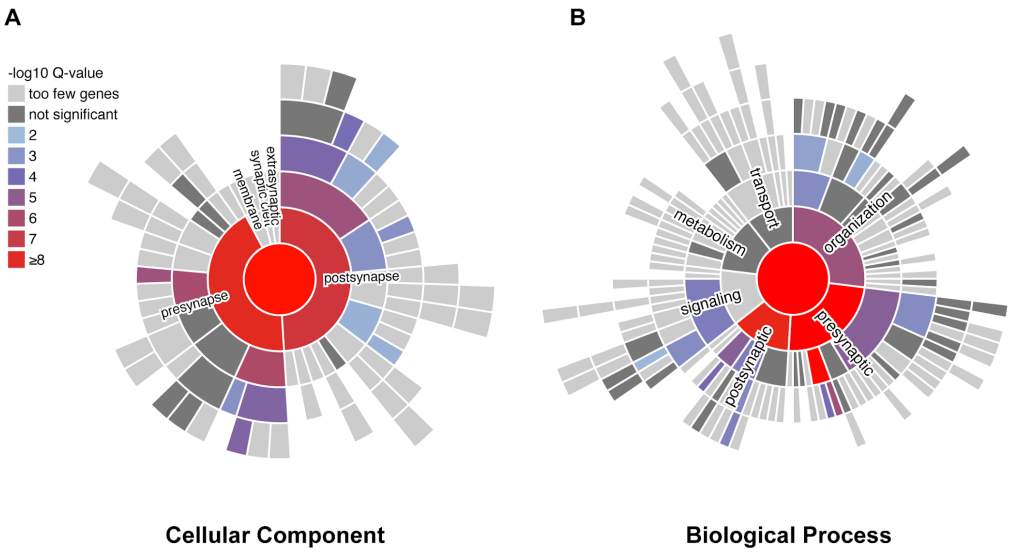

Figure 5. Synaptic location and function of the high confidence NDD genes.
SynGO sunburst plots show nested cellular components (A) and biological processes (B) of the synapse that are statistically enriched in HC-NDD genes. The circle in the center represents the root node, with the hierarchy moving outward from the center. A segment of the inner circle bears a hierarchical relationship to those segments

of the outer circle which lie within the angular sweep of the parent segment. More details can be found on the SynGO website : https://syngoportal.org/. A total of 204 HC-NDD genes were mapped to SynGO annotated genes, of which (A) 174 encoded proteins annotated in different cellular components and (B) 160 encoded proteins annotated in different biological processes (a gene may have multiple annotations). A total of 18 cellular component terms and 20 biological process terms are significantly enriched at 1% FDR (testing terms with at least three matching input genes). Enrichment analyses were performed using 1-sided Fisher exact tests ("with greater than" for the alternative hypothesis) and the False Discovery Rate (FDR) method was applied for multiple testing correction.

Perspectives

This overview of the different databases highlights the cumulative evidence of a strong genetic contribution to NDD. Regarding the achievements, it is now estimated that more than 1,500 genes are robustly associated with NDD. The use of whole exome/genome sequencing leads to a genetic diagnosis in more than a third of the patients and a recent modelling suggests that more than 1,000 genes associated with developmental disorders have not yet been described, many of which are likely to be less penetrant than the currently known genes (Kaplanis et al., 2020). Recent progresses have also been made to develop automated pipelines for filtering pathogenic variants among the large numbers of variants identified through exome/genome sequencing (Wright et al., 2020).

Regarding the limitations, for the majority of the genes/variants, we have few information about family history and on the impact of the mutations on the clinical trajectory and outcome of the carriers. Such fine-grained information are sometimes available in medical case-reports, but rarely in a way that the information is standardized and in machine readable format, hindering comprehensive genotype-phenotype analyses based on large-scale cohorts. Another limitation is that most of the studies are focused on patients despite that deleterious mutations are sometimes observed in unaffected relatives and in the general population. Initiatives investigating the general population such as UK-Biobank are therefore crucial in order to detect mutation carriers who are *Resilients* and who could inform us on possible protective factors pointing at new treatments.

Acknowledgements

This research has been conducted using public databases (SFARI, SysID, DDG2P, DBD, gnomAD, SYNGO, Lafora Gene Mutation Database, epiGAD, CarpeDB and EpilepsyGene). This work was funded by Institut Pasteur, the Bettencourt-Schueller Foundation, Université de Paris, the Conny-Maeva Charitable Foundation, the Cognacq Jay Foundation, the Orange Foundation, the Fondamental Foundation, AIMS-2-TRIALS which received support from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777394, the Human Brain Project and the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Highlights

- Databases overview listing gene-disease pairs related to neurodevelopmental disorder
- Genetic and phenotypic continuum of neurodevelopmental disorder
- List of high-confidence neurodevelopmental disorder genes
- Genes for neurodevelopmental disorders are expressed early in fetal brain
- Standardized genotype-phenotype data are required in neurodevelopmental disorders

Websites

Simons Foundation Autism Research Initiative (SFARI): https://gene.sfari.org/
Simons Foundation Powering Autism Research for Knowledge (SPARK): https://simonsfoundation.s3.amazonaws.com/share/SFARI/SPARK_Gene_List.pdf
Clinical Genome (ClinGen): https://www.clinicalgenome.org/
Clinical Variation (ClinVar): https://www.ncbi.nlm.nih.gov/clinvar/
Systems Biology Approaches to Intellectual Disability (SysID): https://sysid.cmbi.umcn.nl/
Gene2Phenotype (G2P): https://www.ebi.ac.uk/gene2phenotype

DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER): https://decipher.sanger.ac.uk/
Universal Protein Resource (UniProt): https://www.uniprot.org/
Developmental Brain Disorder (DBD): https://dbd.geisingeradmi.org/
Online Mendelian Inheritance in Man (OMIM): https://www.ncbi.nlm.nih.gov/omim
The Lafora Gene Mutation Databas: http://projects.tcag.ca/lafora/
The Epilepsy Genetic Association Database (epiGAD): http://www.epigad.org/
CarpeDB: http://carpedb.ua.edu/
EpilepsyGene: http://www.wzgenomics.cn/EpilepsyGene/index.php
Mayo Clinic Laboratories: https://www.mayocliniclabs.com/it-mmfiles/Targeted_Genes_and_Methodology_Details_for_Epilespy_Genetic_Panels.pdf
GeneTrek: https://genetrek.pasteur.fr/
National Center for Biotechnology Information (NCBI): https://www.ncbi.nlm.nih.gov/
Synaptic Gene Ontologies (SynGO): https://syngoportal.org/
Genome Aggregation Database (gnomAD): https://gnomad.broadinstitute.org/
Specific Expression Analysis (SEA): http://genetics.wustl.edu/jdlab/csea-tool-2/
Atlas of the developing human brain (Brainspan): https://www.brainspan.org/
Protein-Protein Interaction Networks & Functional Enrichment Analysis (STRING): https://string-db.org/
Human Disease Genes website series (HDG): https://humandiseasegenes.nl/

References

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., Packer, A., 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism 4, 36. https://doi.org/10.1186/2040-2392-4-36

Baron-Cohen, S., Scott, F.J., Allison, C., Williams, J., Bolton, P., Matthews, F.E., Brayne, C., 2009. Prevalence of autism-spectrum conditions: UK school-based population study. Br. J. Psychiatry J. Ment. Sci. 194, 500–509. https://doi.org/10.1192/bjp.bp.108.059345

Bourgeron, T., 2015. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat. Rev. Neurosci. 16, 551–563. https://doi.org/10.1038/nrn3992

Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu, Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179, 1469-1482.e11. https://doi.org/10.1016/j.cell.2019.11.020

Dumas, G., Malesys, S., Bourgeron, T., 2019. Systematic detection of divergent brain protein-coding genes in human evolution and their roles in cognition (preprint). Genetics. https://doi.org/10.1101/658658

Dunn, P.J., Maher, B.H., Albury, C.L., Stuart, S., Sutherland, H.G., Maksemous, N., Benton, M.C., Smith, R.A., Haupt, L.M., Griffiths, L.R., 2020. Tiered analysis of whole-exome sequencing for epilepsy diagnosis. Mol. Genet. Genomics MGG 295, 751–763. https://doi.org/10.1007/s00438-020-01657-x

Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., Carter, N.P., 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. 84, 524–533. https://doi.org/10.1016/j.ajhg.2009.03.010

Gillberg, C., 2010. The ESSENCE in child psychiatry: Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations. Res. Dev. Disabil. 31, 1543–1551. https://doi.org/10.1016/j.ridd.2010.06.002

Gonzalez-Mantilla, A.J., Moreno-De-Luca, A., Ledbetter, D.H., Martin, C.L., 2016. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. JAMA Psychiatry 73, 275–283. https://doi.org/10.1001/jamapsychiatry.2015.2692

Hansen, B.H., Oerbeck, B., Skirbekk, B., Petrovski, B.É., Kristensen, H., 2018. Neurodevelopmental disorders: prevalence and comorbidity in children referred to mental health services. Nord. J. Psychiatry 72, 285–291. https://doi.org/10.1080/08039488.2018.1444087

Ianzano, L., Zhang, J., Chan, E.M., Zhao, X.-C., Lohi, H., Scherer, S.W., Minassian, B.A., 2005. Lafora progressive Myoclonus Epilepsy mutation database-EPM2A and NHLRC1 (EPM2B) genes. Hum. Mutat. 26, 397. https://doi.org/10.1002/humu.9376

Idring, S., Rai, D., Dal, H., Dalman, C., Sturm, H., Zander, E., Lee, B.K., Serlachius, E., Magnusson, C., 2012. Autism spectrum disorders in the Stockholm Youth Cohort: design, prevalence and validity. PloS One 7, e41280. https://doi.org/10.1371/journal.pone.0041280

International League Against Epilepsy Consortium on Complex Epilepsies, 2018. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat. Commun. 9, 5269. https://doi.org/10.1038/s41467-018-07524-z

Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y., Gallone, G., Lelieveld, S.H., Martin, H.C., McRae, J.F., Short, P.J., Torene, R.I., de Boer, E., Danecek, P., Gardner, E.J., Huang, N., Lord, J., Martincorena, I., Pfundt, R., Reijnders, M.R.F., Yeung, A., Yntema, H.G., Deciphering Developmental Disorders Study, Vissers, L.E.L.M., Juusola, J., Wright, C.F., Brunner, H.G., Firth, H.V., FitzPatrick, D.R., Barrett, J.C., Hurles, M.E., Gilissen, C., Retterer, K., 2020. Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature 586, 757–762. https://doi.org/10.1038/s41586-020-2832-5

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., Genome Aggregation Database Consortium, 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

Kjeldsen, M.J., Kyvik, K.O., Christensen, K., Friis, M.L., 2001. Genetic and environmental factors in epilepsy: a population-based study of 11900 Danish twin pairs. Epilepsy Res. 44, 167–178. https://doi.org/10.1016/s0920-

1211(01)00196-6
Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M.A.W., Kleefstra, T., Kramer, J.M., Webber, C., Huynen, M.A., Schenck, A., 2016. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am. J. Hum. Genet. 98, 149–164. https://doi.org/10.1016/j.ajhg.2015.11.024
Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P., Hussain, N.K., Imig, C., de Jong, A.P.H., Jung, H., Kohansalnodehi, M., Kramarz, B., Lipstein, N., Lovering, R.C., MacGillavry, H., Mariano, V., Mi, H., Ninov, M., Osumi-Sutherland, D., Pielot, R., Smalla, K.-H., Tang, H., Tashman, K., Toonen, R.F.G., Verpelli, C., Reig-Viader, R., Watanabe, K., van Weering, J., Achsel, T., Ashrafi, G., Asi, N., Brown, T.C., De Camilli, P., Feuermann, M., Foulger, R.E., Gaudet, P., Joglekar, A., Kanellopoulos, A., Malenka, R., Nicoll, R.A., Pulido, C., de Juan-Sanz, J., Sheng, M., Südhof, T.C., Tilgner, H.U., Bagni, C., Bayés, À., Biederer, T., Brose, N., Chua, J.J.E., Dieterich, D.C., Gundelfinger, E.D., Hoogenraad, C., Huganir, R.L., Jahn, R., Kaeser, P.S., Kim, E., Kreutz, M.R., McPherson, P.S., Neale, B.M., O'Connor, V., Posthuma, D., Ryan, T.A., Sala, C., Feng, G., Hyman, S.E., Thomas, P.D., Smit, A.B., Verhage, M., 2019. SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. Neuron 103, 217-234.e4. https://doi.org/10.1016/j.neuron.2019.05.002
Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., Buxbaum, J.D., Neale, B.M., MacArthur, D.G., Wall, D.P., Robinson, E.B., Daly, M.J., 2017. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat. Genet. 49, 504–510. https://doi.org/10.1038/ng.3789
Lai, M.-C., Lombardo, M.V., Baron-Cohen, S., 2014. Autism. Lancet Lond. Engl. 383, 896–910. https://doi.org/10.1016/S0140-6736(13)61539-1
Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C., Lyoshin, V., Maddipatla, Z., Maiti, R., Mitchell, J., O'Leary, N., Riley, G.R., Shi, W., Zhou, G., Schneider, V., Maglott, D., Holmes, J.B., Kattman, B.L., 2020. ClinVar: improvements to accessing data. Nucleic Acids Res. 48, D835–D844. https://doi.org/10.1093/nar/gkz972
Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Exome Aggregation Consortium, 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–291. https://doi.org/10.1038/nature19057
Maenner, M.J., Shaw, K.A., Baio, J., EdS1, Washington, A., Patrick, M., DiRienzo, M., Christensen, D.L., Wiggins, L.D., Pettygrove, S., Andrews, J.G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C.R., Lee, L.-C., Harrington, R.A., Huston, M., Hewitt, A., PhD-7, Esler, A., Hall-Lande, J., Poynter, J.N., Hallas-Muchow, L., Constantino, J.N., Fitzgerald, R.T., Zahorodny, W., Shenouda, J., Daniels, J.L., Warren, Z., Vehorn, A., Salinas, A., Durkin, M.S., Dietz, P.M., 2020. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. Morb. Mortal. Wkly. Rep. Surveill. Summ. Wash. DC 2002 69, 1–12. https://doi.org/10.15585/mmwr.ss6904a1
McHugh, J.C., Delanty, N., 2008. Chapter 2 Epidemiology and Classification of Epilepsy, in: International Review of Neurobiology. Elsevier, pp. 11–26. https://doi.org/10.1016/S0074-7742(08)00002-0
Miller, L.L., Pellock, J.M., DeLorenzo, R.J., Meyer, J.M., Corey, L.A., 1998. Univariate genetic analyses of epilepsy and seizures in a population-based twin study: the Virginia Twin Registry. Genet. Epidemiol. 15, 33–49. https://doi.org/10.1002/(SICI)1098-2272(1998)15:1<33::AID-GEPI3>3.0.CO;2-5
Myers, S.M., Challman, T.D., Bernier, R., Bourgeron, T., Chung, W.K., Constantino, J.N., Eichler, E.E., Jacquemont, S., Miller, D.T., Mitchell, K.J., Zoghbi, H.Y., Martin, C.L., Ledbetter, D.H., 2020. Insufficient Evidence for "Autism-Specific" Genes. Am. J. Hum. Genet. 106, 587–595. https://doi.org/10.1016/j.ajhg.2020.04.004
Pfisterer, U., Petukhov, V., Demharter, S., Meichsner, J., Thompson, J.J., Batiuk, M.Y., Asenjo-Martinez, A., Vasistha, N.A., Thakur, A., Mikkelsen, J., Adorjan, I., Pinborg, L.H., Pers, T.H., von Engelhardt, J., Kharchenko, P.V., Khodosevich, K., 2020. Identification of epilepsy-associated neuronal subtypes and gene expression underlying epileptogenesis. Nat. Commun. 11, 5038. https://doi.org/10.1038/s41467-020-18752-7
Saemundsen, E., Magnússon, P., Georgsdóttir, I., Egilsson, E., Rafnsson, V., 2013. Prevalence of autism

spectrum disorders in an Icelandic birth cohort. BMJ Open 3, e002748. https://doi.org/10.1136/bmjopen-2013-002748
Saito, M., Hirota, T., Sakamoto, Y., Adachi, M., Takahashi, M., Osato-Kaneda, A., Kim, Y.S., Leventhal, B., Shui, A., Kato, S., Nakamura, K., 2020. Prevalence and cumulative incidence of autism spectrum disorders and the patterns of co-occurring neurodevelopmental disorders in a total population sample of 5-year-old children. Mol. Autism 11, 35. https://doi.org/10.1186/s13229-020-00342-5
Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., Larsson, H., Reichenberg, A., 2017. The Heritability of Autism Spectrum Disorder. JAMA 318, 1182. https://doi.org/10.1001/jama.2017.12141
Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Larsson, H., Hultman, C.M., Reichenberg, A., 2014. The Familial Risk of Autism. JAMA 311, 1770. https://doi.org/10.1001/jama.2014.4144
Shen, Y.-M., Chan, B.S.M., Liu, J.-B., Zhou, Y.-Y., Cui, X.-L., He, Y.-Q., Fang, Y.-M., Xiang, Y.-T., Luo, X.-R., 2018. The prevalence of psychiatric disorders among students aged 6~ 16 years old in central Hunan, China. BMC Psychiatry 18, 243. https://doi.org/10.1186/s12888-018-1823-7
Sokolova, E., Oerlemans, A.M., Rommelse, N.N., Groot, P., Hartman, C.A., Glennon, J.C., Claassen, T., Heskes, T., Buitelaar, J.K., 2017. A Causal and Mediation Analysis of the Comorbidity Between Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). J. Autism Dev. Disord. 47, 1595–1604. https://doi.org/10.1007/s10803-017-3083-7
Speed, D., O'Brien, T.J., Palotie, A., Shkura, K., Marson, A.G., Balding, D.J., Johnson, M.R., 2014. Describing the genetic architecture of epilepsy through heritability analysis. Brain J. Neurol. 137, 2680–2689. https://doi.org/10.1093/brain/awu206
Srivastava, S., Love-Nichols, J.A., Dies, K.A., Ledbetter, D.H., Martin, C.L., Chung, W.K., Firth, H.V., Frazier, T., Hansen, R.L., Prock, L., Brunner, H., Hoang, N., Scherer, S.W., Sahin, M., Miller, D.T., NDD Exome Scoping Review WorkGroup, 2019. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet. Med. Off. J. Am. Coll. Med. Genet. 21, 2413–2421. https://doi.org/10.1038/s41436-019-0554-6
Stefanski, A., Calle-López, Y., Leu, C., Pérez-Palma, E., Pestana-Knight, E., Lal, D., 2020. Clinical sequencing yield in epilepsy, autism spectrum disorder, and intellectual disability: A systematic review and meta-analysis. Epilepsia. https://doi.org/10.1111/epi.16755
Tan, N.C.K., Berkovic, S.F., 2010. The Epilepsy Genetic Association Database (epiGAD): analysis of 165 genetic association studies, 1996-2008. Epilepsia 51, 686–689. https://doi.org/10.1111/j.1528-1167.2009.02423.x
Taylor, M.J., Rosenqvist, M.A., Larsson, H., Gillberg, C., D'Onofrio, B.M., Lichtenstein, P., Lundström, S., 2020. Etiology of Autism Spectrum Disorders and Autistic Traits Over Time. JAMA Psychiatry 77, 936. https://doi.org/10.1001/jamapsychiatry.2020.0680
The Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H.K., Walters, R.K., Bras, J., Duncan, L., Escott-Price, V., Falcone, G.J., Gormley, P., Malik, R., Patsopoulos, N.A., Ripke, S., Wei, Z., Yu, D., Lee, P.H., Turley, P., Grenier-Boley, B., Chouraki, V., Kamatani, Y., Berr, C., Letenneur, L., Hannequin, D., Amouyel, P., Boland, A., Deleuze, J.-F., Duron, E., Vardarajan, B.N., Reitz, C., Goate, A.M., Huentelman, M.J., Kamboh, M.I., Larson, E.B., Rogaeva, E., St George-Hyslop, P., Hakonarson, H., Kukull, W.A., Farrer, L.A., Barnes, L.L., Beach, T.G., Demirci, F.Y., Head, E., Hulette, C.M., Jicha, G.A., Kauwe, J.S.K., Kaye, J.A., Leverenz, J.B., Levey, A.I., Lieberman, A.P., Pankratz, V.S., Poon, W.W., Quinn, J.F., Saykin, A.J., Schneider, L.S., Smith, A.G., Sonnen, J.A., Stern, R.A., Van Deerlin, V.M., Van Eldik, L.J., Harold, D., Russo, G., Rubinsztein, D.C., Bayer, A., Tsolaki, M., Proitsi, P., Fox, N.C., Hampel, H., Owen, M.J., Mead, S., Passmore, P., Morgan, K., Nöthen, M.M., Schott, J.M., Rossor, M., Lupton, M.K., Hoffmann, P., Kornhuber, J., Lawlor, B., McQuillin, A., Al-Chalabi, A., Bis, J.C., Ruiz, A., Boada, M., Seshadri, S., Beiser, A., Rice, K., van der Lee, S.J., De Jager, P.L., Geschwind, D.H., Riemenschneider, M., Riedel-Heller, S., Rotter, J.I., Ransmayr, G., Hyman, B.T., Cruchaga, C., Alegret, M., Winsvold, B., Palta, P., Farh, K.-H., Cuenca-Leon, E., Furlotte, N., Kurth, T., Ligthart, L., Terwindt, G.M., Freilinger, T., Ran, C., Gordon, S.D., Borck, G., Adams, H.H.H., Lehtimäki, T., Wedenoja, J., Buring, J.E., Schürks, M., Hrafnsdottir, M., Hottenga, J.-J., Penninx, B., Artto, V., Kaunisto, M., Vepsäläinen, S., Martin, N.G., Montgomery, G.W., Kurki, M.I., Hämäläinen, E., Huang, H., Huang, J., Sandor, C., Webber, C., Muller-Myhsok, B., Schreiber, S., Salomaa, V., Loehrer, E., Göbel, H., Macaya, A., Pozo-Rosich, P., Hansen, T., Werge, T., Kaprio, J., Metspalu, A., Kubisch, C., Ferrari, M.D., Belin, A.C., van den Maagdenberg, A.M.J.M., Zwart, J.-A., Boomsma, D., Eriksson, N., Olesen, J., Chasman, D.I., Nyholt, D.R., Anney, R., Avbersek, A., Baum, L., Berkovic, S., Bradfield, J., Buono, R.J., Catarino, C.B., Cossette, P., De Jonghe, P., Depondt, C., Dlugos, D., Ferraro, T.N., French, J., Hjalgrim, H., Jamnadas-Khoda, J., Kälviäinen, R., Kunz, W.S., Lerche, H., Leu, C., Lindhout, D., Lo, W., Lowenstein, D., McCormack, M., Møller, R.S., Molloy, A., Ng, P.-W., Oliver, K., Privitera, M., Radtke, R., Ruppert, A.-K., Sander, T., Schachter, S., Schankin, C., Scheffer, I., Schoch, S., Sisodiya, S.M., Smith, P., Sperling, M., Striano, P., Surges, R., Thomas, G.N., Visscher, F., Whelan, C.D., Zara, F., Heinzen, E.L., Marson, A., Becker, F., Stroink, H., Zimprich, F., Gasser, T., Gibbs, R., Heutink, P., Martinez, M., Morris, H.R., Sharma, M., Ryten, M., Mok, K.Y., Pulit, S., Bevan, S., Holliday,

E., Attia, J., Battey, T., Boncoraglio, G., Thijs, V., Chen, W.-M., Mitchell, B., Rothwell, P., Sharma, P., Sudlow, C., Vicente, A., Markus, H., Kourkoulis, C., Pera, J., Raffeld, M., Silliman, S., Boraska Perica, V., Thornton, L.M., Huckins, L.M., William Rayner, N., Lewis, C.M., Gratacos, M., Rybakowski, F., Keski-Rahkonen, A., Raevuori, A., Hudson, J.I., Reichborn-Kjennerud, T., Monteleone, P., Karwautz, A., Mannik, K., Baker, J.H., O'Toole, J.K., Trace, S.E., Davis, O.S.P., Helder, S.G., Ehrlich, S., Herpertz-Dahlmann, B., Danner, U.N., van Elburg, A.A., Clementi, M., Forzan, M., Docampo, E., Lissowska, J., Hauser, J., Tortorella, A., Maj, M., Gonidakis, F., Tziouvas, K., Papezova, H., Yilmaz, Z., Wagner, G., Cohen-Woods, S., Herms, S., Julià, A., Rabionet, R., Dick, D.M., Ripatti, S., Andreassen, O.A., Espeseth, T., Lundervold, A.J., Steen, V.M., Pinto, D., Scherer, S.W., Aschauer, H., Schosser, A., Alfredsson, L., Padyukov, L., Halmi, K.A., Mitchell, J., Strober, M., Bergen, A.W., Kaye, W., Szatkiewicz, J.P., Cormand, B., Ramos-Quiroga, J.A., Sánchez-Mora, C., Ribasés, M., Casas, M., Hervas, A., Arranz, M.J., Haavik, J., Zayats, T., Johansson, S., Williams, N., Elia, J., Dempfle, A., Rothenberger, A., Kuntsi, J., Oades, R.D., Banaschewski, T., Franke, B., Buitelaar, J.K., Arias Vasquez, A., Doyle, A.E., Reif, A., Lesch, K.-P., Freitag, C., Rivero, O., Palmason, H., Romanos, M., Langley, K., Rietschel, M., Witt, S.H., Dalsgaard, S., Børglum, A.D., Waldman, I., Wilmot, B., Molly, N., Bau, C.H.D., Crosbie, J., Schachar, R., Loo, S.K., McGough, J.J., Grevet, E.H., Medland, S.E., Robinson, E., Weiss, L.A., Bacchelli, E., Bailey, A., Bal, V., Battaglia, A., Betancur, C., Bolton, P., Cantor, R., Celestino-Soper, P., Dawson, G., De Rubeis, S., Duque, F., Green, A., Klauck, S.M., Leboyer, M., Levitt, P., Maestrini, E., Mane, S., De-Luca, D.M.-, Parr, J., Regan, R., Reichenberg, A., Sandin, S., Vorstman, J., Wassink, T., Wijsman, E., Cook, E., Santangelo, S., Delorme, R., Rogé, B., Magalhaes, T., Arking, D., Schulze, T.G., Thompson, R.C., Strohmaier, J., Matthews, K., Melle, I., Morris, D., Blackwood, D., McIntosh, A., Bergen, S.E., Schalling, M., Jamain, S., Maaser, A., Fischer, S.B., Reinbold, C.S., Fullerton, J.M., Grigoroiu-Serbanescu, M., Guzman-Parra, J., Mayoral, F., Schofield, P.R., Cichon, S., Mühleisen, T.W., Degenhardt, F., Schumacher, J., Bauer, M., Mitchell, P.B., Gershon, E.S., Rice, J., Potash, J.B., Zandi, P.P., Craddock, N., Ferrier, I.N., Alda, M., Rouleau, G.A., Turecki, G., Ophoff, R., Pato, C., Anjorin, A., Stahl, E., Leber, M., Czerski, P.M., Edenberg, H.J., Cruceanu, C., Jones, I.R., Posthuma, D., Andlauer, T.F.M., Forstner, A.J., Streit, F., Baune, B.T., Air, T., Sinnamon, G., Wray, N.R., MacIntyre, D.J., Porteous, D., Homuth, G., Rivera, M., Grove, J., Middeldorp, C.M., Hickie, I., Pergadia, M., Mehta, D., Smit, J.H., Jansen, R., de Geus, E., Dunn, E., Li, Q.S., Nauck, M., Schoevers, R.A., Beekman, A.T., Knowles, J.A., Viktorin, A., Arnold, P., Barr, C.L., Bedoya-Berrio, G., Bienvenu, O.J., Brentani, H., Burton, C., Camarena, B., Cappi, C., Cath, D., Cavallini, M., Cusi, D., Darrow, S., Denys, D., Derks, E.M., Dietrich, A., Fernandez, T., Figee, M., Freimer, N., Gerber, G., Grados, M., Greenberg, E., Hanna, G.L., Hartmann, A., Hirschtritt, M.E., Hoekstra, P.J., Huang, A., Huyser, C., Illmann, C., Jenike, M., Kuperman, S., Leventhal, B., Lochner, C., Lyon, G.J., Macciardi, F., Madruga-Garrido, M., Malaty, I.A., Maras, A., McGrath, L., Miguel, E.C., Mir, P., Nestadt, G., Nicolini, H., Okun, M.S., Pakstis, A., Paschou, P., Piacentini, J., Pittenger, C., Plessen, K., Ramensky, V., Ramos, E.M., Reus, V., Richter, M.A., Riddle, M.A., Robertson, M.M., Roessner, V., Rosário, M., Samuels, J.F., Sandor, P., Stein, D.J., Tsetsos, F., Van Nieuwerburgh, F., Weatherall, S., Wendland, J.R., Wolanczyk, T., Worbe, Y., Zai, G., Goes, F.S., McLaughlin, N., Nestadt, P.S., Grabe, H.-J., Depienne, C., Konkashbaev, A., Lanzagorta, N., Valencia-Duarte, A., Bramon, E., Buccola, N., Cahn, W., Cairns, M., Chong, S.A., Cohen, D., Crespo-Facorro, B., Crowley, J., Davidson, M., DeLisi, L., Dinan, T., Donohoe, G., Drapeau, E., Duan, J., Haan, L., Hougaard, D., Karachanak-Yankova, S., Khrunin, A., Klovins, J., Kučinskas, V., Lee Chee Keong, J., Limborska, S., Loughland, C., Lönnqvist, J., Maher, B., Mattheisen, M., McDonald, C., Murphy, K.C., Murray, R., Nenadic, I., van Os, J., Pantelis, C., Pato, M., Petryshen, T., Quested, D., Roussos, P., Sanders, A.R., Schall, U., Schwab, S.G., Sim, K., So, H.-C., Stögmann, E., Subramaniam, M., Toncheva, D., Waddington, J., Walters, J., Weiser, M., Cheng, W., Cloninger, R., Curtis, D., Gejman, P.V., Henskens, F., Mattingsdal, M., Oh, S.-Y., Scott, R., Webb, B., Breen, G., Churchhouse, C., Bulik, C.M., Daly, M., Dichgans, M., Faraone, S.V., Guerreiro, R., Holmans, P., Kendler, K.S., Koeleman, B., Mathews, C.A., Price, A., Scharf, J., Sklar, P., Williams, J., Wood, N.W., Cotsapas, C., Palotie, A., Smoller, J.W., Sullivan, P., Rosand, J., Corvin, A., Neale, B.M., 2018. Analysis of shared heritability in common disorders of the brain. Science 360, eaap8757. https://doi.org/10.1126/science.aap8757

The UniProt Consortium, 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212. https://doi.org/10.1093/nar/gku989

Thomas, S., Hovinga, M.E., Rai, D., Lee, B.K., 2017. Brief Report: Prevalence of Co-occurring Epilepsy and Autism Spectrum Disorder: The U.S. National Survey of Children's Health 2011-2012. J. Autism Dev. Disord. 47, 224–229. https://doi.org/10.1007/s10803-016-2938-7

Tick, B., Bolton, P., Happé, F., Rutter, M., Rijsdijk, F., 2016. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J. Child Psychol. Psychiatry 57, 585–595. https://doi.org/10.1111/jcpp.12499

Wang, J., Lin, Z.-J., Liu, L., Xu, H.-Q., Shi, Y.-W., Yi, Y.-H., He, N., Liao, W.-P., 2017. Epilepsy-associated genes. Seizure 44, 11–20. https://doi.org/10.1016/j.seizure.2016.11.030

Wright, C.F., Eberhardt, R.Y., Constantinou, P., Hurles, M.E., FitzPatrick, D.R., Firth, H.V., DDD Study, 2020. Evaluating variants classified as pathogenic in ClinVar in the DDD Study. Genet. Med. Off. J. Am. Coll. Med. Genet. https://doi.org/10.1038/s41436-020-01021-9

Zablotsky, B., Black, L.I., Blumberg, S.J., 2017. Estimated Prevalence of Children With Diagnosed Developmental Disabilities in the United States, 2014-2016. NCHS Data Brief 1–8.
Zablotsky, B., Black, L.I., Maenner, M.J., Schieve, L.A., Danielson, M.L., Bitsko, R.H., Blumberg, S.J., Kogan, M.D., Boyle, C.A., 2019. Prevalence and Trends of Developmental Disabilities among Children in the United States: 2009-2017. Pediatrics 144. https://doi.org/10.1542/peds.2019-0811
Zarrei, M., Burton, C.L., Engchuan, W., Young, E.J., Higginbotham, E.J., MacDonald, J.R., Trost, B., Chan, A.J.S., Walker, S., Lamoureux, S., Heung, T., Mojarad, B.A., Kellam, B., Paton, T., Faheem, M., Miron, K., Lu, C., Wang, T., Samler, K., Wang, X., Costain, G., Hoang, N., Pellecchia, G., Wei, J., Patel, R.V., Thiruvahindrapuram, B., Roifman, M., Merico, D., Goodale, T., Drmic, I., Speevak, M., Howe, J.L., Yuen, R.K.C., Buchanan, J.A., Vorstman, J.A.S., Marshall, C.R., Wintle, R.F., Rosenberg, D.R., Hanna, G.L., Woodbury-Smith, M., Cytrynbaum, C., Zwaigenbaum, L., Elsabbagh, M., Flanagan, J., Fernandez, B.A., Carter, M.T., Szatmari, P., Roberts, W., Lerch, J., Liu, X., Nicolson, R., Georgiades, S., Weksberg, R., Arnold, P.D., Bassett, A.S., Crosbie, J., Schachar, R., Stavropoulos, D.J., Anagnostou, E., Scherer, S.W., 2019. A large data resource of genomic copy number variation across neurodevelopmental disorders. NPJ Genomic Med. 4, 26. https://doi.org/10.1038/s41525-019-0098-3