




SYSTEMATIC REVIEW

A systematic review and pooled analysis of penetrance estimates of copy-number variants associated with neurodevelopment



Shuxiang Goh^{1,2,*} , Lavvina Thiyagarajan^{1,2,3}, Tracy Dudding-Byth⁴, Mark Pinese^{1,5}, Edwin P. Kirk^{1,2,6,7}

¹School of Clinical Medicine, University of New South Wales, Sydney, NSW, Australia; ²New South Wales Health, NSW, Australia; ³The Children's Hospital at Westmead, Sydney, NSW, Australia; ⁴Hunter Genetics, Newcastle, NSW, Australia; ⁵Children's Cancer Institute, Sydney, NSW, Australia; ⁶Centre for Clinical Genetics, Sydney Children's Hospital, Sydney, NSW, Australia; ⁷NSW Health Pathology Randwick Genomics Laboratory, Sydney, NSW, Australia

ARTICLE INFO

Article history:

Received 18 March 2024

Received in revised form

23 July 2024

Accepted 24 July 2024

Available online 30 July 2024

Keywords:

Deletion

Duplication

Intellectual disability

Neurosusceptibility

Prenatal

ABSTRACT

Purpose: Many copy-number variants (CNVs) are reported to cause a variety of neurodevelopmental disabilities including intellectual disability, developmental delay, autism, and other phenotypes with incomplete penetrance. Therefore, not all individuals with a pathogenic CNV are affected. Penetrance estimates vary between studies. A systematic review was conducted to clarify CNV penetrance for 83 recurrent CNVs.

Methods: A systematic review using PRISMA guidelines (PROSPERO #CRD42021253955) was conducted to identify penetrance estimates for CNVs associated with neurodevelopment. Pooled analysis was performed using forest plots. The Ottawa Risk of Bias Assessment facilitated evaluation.

Results: Fifteen studies were reviewed in detail with 9 affected cohorts pooled and compared with the gnomAD v4.0 CNV control cohort of 269,885 individuals. Several CNVs previously associated with nonstatistically significant penetrance estimates now exhibit statistically significant differences, contributing to emerging evidence for their pathogenicity (15q24 duplication [A-D breakpoints], 15q24.2q24.5 deletion and duplication [*FBXO22*], 17q11.2 duplication [*NFI*], 17q21.31 duplication [*KANSL1*] and 22q11.2 distal duplication). Additionally, evidence is presented for the benign nature of some CNVs (15q11.2 duplication [*NIPAI*] and 2q13 proximal duplication [*NPHPI*]).

Conclusion: This is a large-scale systematic review of CNVs associated with neurodevelopment. A synopsis analyzing penetrance and pathogenicity is provided for each of the 83 recurrent CNVs.

© 2024 by American College of Medical Genetics and Genomics. Published by Elsevier Inc.

*Correspondence and requests for materials should be addressed to Shuxiang Goh, School of Clinical Medicine, University of New South Wales, High Street, Randwick, Sydney, NSW, Australia 2052. Email address: shuxiang.goh@health.nsw.gov.au

Introduction

Copy-number variants (CNVs) are additional or missing genetic information in a contiguous region of a chromosome. These changes are usually submicroscopic but can contribute to medical pathology in humans. Those that predispose to intellectual disability (ID), developmental delay (DD) and autism are termed neurodevelopmental CNVs. Many neurodevelopmental CNVs have also been implicated as causative of congenital malformations, personality differences, or behavioral issues.¹⁻¹⁶

There is no standard definition of a neurodevelopmental CNV. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition definition of a neurodevelopmental disorder includes ID, DD, autism spectrum disorder, attention deficit hyperactivity disorder (ADHD), communication disorder, specific learning disorder, and motor disorder (eg, tics).¹⁷ However, children are not usually tested with a chromosomal microarray for ADHD, motor disorder, tics, or mild autism spectrum disorder.¹⁻¹⁵ The affected cohorts in these studies also tend to be skewed toward individuals with severe disability because these are the individuals for whom doctors are more likely to order genetic testing. This means that the reported datasets are derived from samples submitted for diagnostic testing and are not representative of neurodevelopmental disorders as a whole. Some individuals harboring certain CNVs can manifest a medical condition, whereas others with the same CNV do not. This variation may be because of other genetic and environmental influences. Penetrance refers to the proportion of individuals with the CNV who are affected.^{4,15} A CNV with low penetrance affects only a few individuals who have it, a highly penetrant CNV affects the majority of those who have it, and a fully penetrant CNV affects all who have it.

There are thousands of CNVs,^{18,19} although many of these may be unique to 1 family. A central repository reporting on penetrance for all recurrent CNVs is lacking. Systematic reviews of a particular CNV are an excellent source of information,¹⁹⁻²¹ but these articles are relevant for only that particular CNV. Some websites publish information on multiple CNVs,^{16,18,22-26} but not all CNVs are listed,^{16,22-26} and their focus is not on penetrance.

A limitation in the existing literature is that penetrance estimates for CNVs sometimes vary considerably between studies. This can make clinical decisions challenging because reasons for these differences and the accuracy of such studies have not been systematically explored.

This systematic review was conducted to serve as a repository for all published penetrance estimates for recurrent neurodevelopmental CNVs and explores reasons for different estimates between studies.

Materials and Methods

This systematic review was prospectively registered (PROSPERO CRD42021253955).²⁷

Search strategy

A literature search was conducted for studies reporting CNV penetrance of neurodevelopmental conditions between 2006 and 2022 in Medline (PubMed), Embase, Prospero, and OpenGrey using the PRISMA²⁸ approach (Figure 1), with details in Supplemental Methods 1. Fifteen articles met inclusion criteria for this review.

Ottawa Risk of Bias Assessment

Following PRISMA guidelines,²⁸ the first author conducted an Ottawa risk of bias assessment²⁹ for all 15 studies, whereas the second author performed this assessment for 2. A high degree of internal consistency was observed between the authors, consistent with the high internal validity reported by the developers of the Ottawa Risk of Bias Assessment.²⁹

Synthesis methods

A case-by-case approach was chosen to determine if it was reasonable to compare CNVs from different studies, and duplicate affected cohorts were excluded from pooled analysis (Supplemental Analysis).

Three different methods for estimating penetrance were identified by this systematic review. The first was a Bayesian formula initially developed for penetrance of CNVs for schizophrenia,¹⁵ then adapted for neurodevelopmental disorders.^{1-6,8,9,11-14} This represents the most widely used method for penetrance estimation in case-control studies. The second method only works for cohort studies and is a more straightforward method for estimating penetrance.¹⁰ These 2 methods were utilized for their respective studies, and results are shown in Table 2 and Supplemental Analysis. The third method used a series of approximations to arrive at a penetrance estimate.⁷ After recalculating penetrance using the study's raw data with a Bayesian formula (without introducing any approximations), this study's findings are also reported in Table 2 and Supplemental Analysis.

The Bayesian formula used in this systematic review to estimate penetrance in 14 of 15 case-control studies^{1-9,11-15} is presented below

$$\text{Penetrance} = P(D|G) = \frac{P(D) \times P(G|D)}{P(G)} = \frac{P(D) \times P(G|D)}{P(G|D) \times P(D) + P(G|D^c) \times P(D^c)}$$

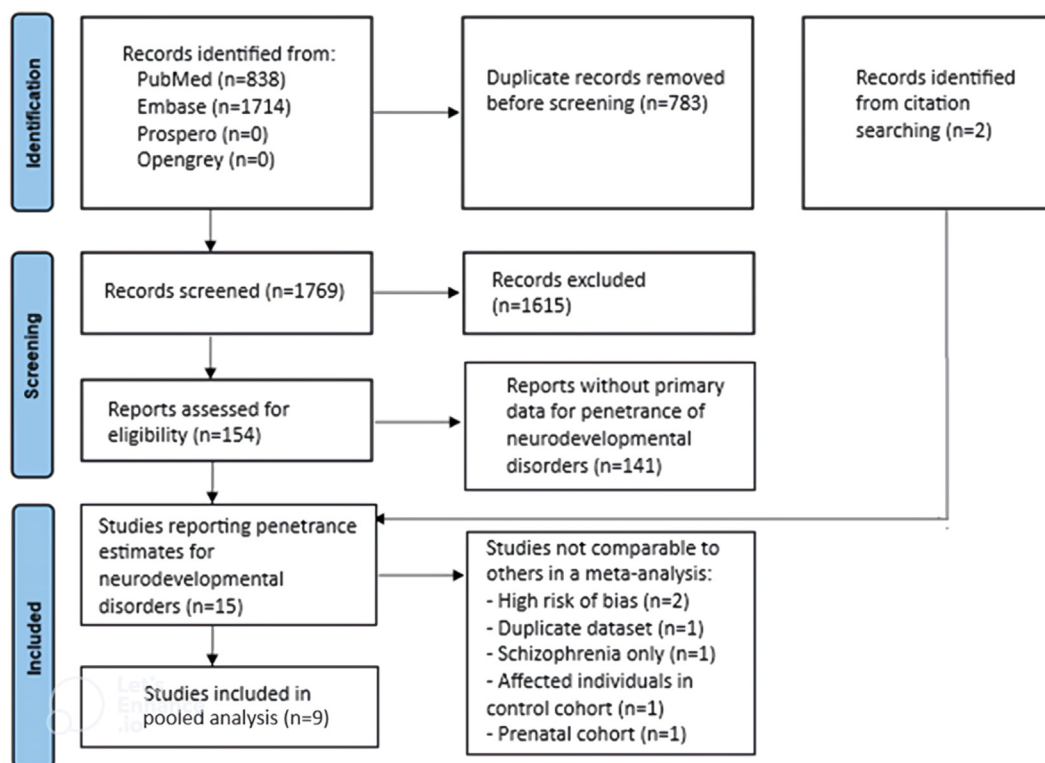


Figure 1 PRISMA search strategy for a systematic review of penetrance estimates for copy-number variants associated with neurodevelopment. The search included abstracts from published articles, conference proceedings, theses, and EPub ahead of print articles. The search terms are included in [Supplemental Methods 1](#). After removing duplicates, 1769 articles were identified. Upon reviewing their abstracts, 154 articles were flagged for full-text review. Of these, 13 articles^{1,3-7,9-15} were identified with evidence of penetrance estimates for copy-number variants involved in neurodevelopmental disorders. Two additional references^{2,8} were identified by citation searching, bringing the total to 15. A subsequent Ottawa Risk of Bias Assessment²⁹ led to the exclusion of 6 studies^{1,5,8,11,12,15} because their results were incomparable for a pooled analysis with the remaining 9 studies.^{2-4,6,7,9,10,13,14}

in which $P(D|G)$ is the probability of manifesting the disease if the genotype is present, $P(D)$ is the probability of manifesting the phenotype (or disease), $P(G|D)$ is the probability of having the genotype given the disease is present, $P(G)$ is the probability of having the genotype (or CNV), D^c refers to not having the disease and $P(G|D^c)$ is the probability of having the genotype given you do not have the disease.

Penetrance for the cohort study¹⁰ was estimated using the formula below:

$$\text{Penetrance} = \frac{\text{Number of affected individuals with CNV}}{\text{total number of individuals with CNV}}$$

Details regarding confidence intervals of penetrance estimates are provided in [Supplemental Methods 3](#).

Published penetrance estimates and their 95% confidence intervals for individual studies listed in this systematic review are calculated from the raw data provided by the original studies. These penetrance estimates may differ to that published by the original studies because of minor inconsistencies that some other studies introduced when using the formula ([Supplemental Methods 3.1](#)).

Forest plots were created for each of the 83 CNVs ([Supplemental Analysis](#)). A meta-analysis was not performed. Instead, pooled analysis of the affected cohorts compared with the much larger gnomAD v4.0 CNV control cohort ($n = 269,885$) was thought to provide a more satisfactory approach with overall benefits superior to that of a meta-analysis. This is mainly due to the gnomAD v4.0 control cohort, released in late 2023, being 10 times (or more) larger than many previous control cohorts, more standardized, and likely to contain fewer affected individuals compared with control cohorts in previous studies. The gnomAD cohort comprises the UK biobank, several other biobanks, and individuals flagged as controls in other studies. Detailed methods of the mathematics and considerations behind penetrance estimates for the pooled analysis and their 95% confidence intervals are provided in [Supplemental Methods 3.2](#).

Results

A PRISMA search²⁸ ([Figure 1](#)) identified 15 articles¹⁻¹⁵ reporting on penetrance estimates for 92 CNVs. Of these,

penetrance estimates from 9 studies were assessed as comparable, and their 83 partially overlapping CNVs are presented in a pooled analysis ([Supplemental Analysis](#)).

All studies used chromosomal microarray to determine copy-number variation. One study used exome sequencing and microarray.¹⁰ Although all studies were conducted in high-income countries,¹⁻¹⁵ most^{1-7,9-15} included participants who identified as members of a minority ethnicity.

[Table 1](#) highlights that the primary difference between studies lies in their choice of affected and control cohorts. One study defined their affected cohort as individuals who had ID, DD, or autism.² Others did this and also included individuals referred for congenital malformations,^{1,4,7,9,11,13,14} schizophrenia,^{5,8} or ADHD.⁶ Some included individuals with behavioral disorders^{4,6,10,14} or individuals with an unspecified phenotype.^{4,13} One studied neurodevelopmental disorders,¹⁰ 1 studied autism only,³ and 1 studied prenatal and postnatal anomalies.¹¹

Control cohorts may contain affected individuals. Some authors took steps to reduce the likelihood of this.^{3,8,10} However, the control cohorts of most studies on neurodevelopment are affected adult cohorts of other studies,^{2,4,9,10,13-15} such as outpatient ophthalmology clinics, patients with melanoma, a study on smoking, or blood bank donors, with the rationale being that these cohorts will be relatively depleted for severe childhood-onset conditions and can therefore act as controls for childhood-onset neurodevelopmental disability.

[Table 2](#) lists 83 CNVs, their breakpoints (converted to GRCh38), pooled penetrance estimates, and 95% confidence intervals from 9 studies.^{2-4,6,7,9,10,13,14} In general, studies used similar breakpoints, with exceptions and a commentary for each CNV provided in [Supplemental Analysis](#).

[Table 2](#) (and [Supplemental Analysis](#)) show that when raw data of multiple earlier studies reporting penetrance are pooled and compared with the large gnomAD v4.0 control cohort, penetrance estimates have tighter 95% confidence intervals, and in many cases, the tighter confidence intervals provide evidence for statistical significance and pathogenicity. Statistical significance is achieved when the lower 95% confidence interval for penetrance is greater than the value of P(D) in the study. This occurs when there is a statistically significant difference between the prevalence of the CNV in affected vs control cohorts.

There are many examples where updated penetrance estimates using this systematic review result in statistical significance for a CNV in which penetrance was previously ambiguous or nonsignificant. For example, the 22q11.2 distal duplication [*BCR*, *MAPK1*] was identified in 3 earlier studies to have strikingly different penetrance estimates of 0%,¹⁰ 16%,⁹ and 100%⁴ with correspondingly wide 95% confidence intervals of 0% to 100%,¹⁰ 6% to 100%,⁹ and 5.6% to 100%.⁴ This systematic review pooled the nonoverlapping affected cohorts and compared them with the gnomAD control cohort, resulting in a penetrance estimate of 51% and much tighter confidence intervals (95% CI 32%-77%), which is statistically significant, contributing to

emerging evidence for pathogenicity of this duplication. Likewise, some results published in earlier studies that did not achieve statistical significance individually, do achieve statistical significance using this pooled approach (eg, 15q24 duplication [*BBS4*, *PML*, and *SIN3A*] [A-D breakpoints], 15q24.2q24.5 deletion and duplication [*FBXO22* and *TSPAN3*], 17q11.2 duplication [*NF1*], 17q21.31 duplication [*MAPT* and *KANSL1*]), again contributing to emerging evidence for pathogenicity of these CNVs. Others that were statistically significant in some studies but not significant in other studies have a pooled estimate with tighter confidence intervals, which helps to clarify this disparity, suggesting either pathogenicity or non-pathogenicity (eg, 1q21.1 distal deletion and duplication [*GJA5*], 3q29 deletion and duplication [*DLG1*], 5q35.3 duplication [*NSDI*], 9q34 duplication [*EHMT1*], 15q11.2 [BP1-BP2] duplication [*NIPA1*, *NIPA2*], 15q24 deletion [*BBS4*, *PML*, and *SIN3A*], 16p11.2 distal deletion [*SH2B1*], 17p13.3 deletion [*YWHAE*], and 17p13.3 duplication [*PAFAH1B1*]). Lastly, some CNVs with penetrance estimates that were statistically significant have been recalculated to be statistically nonsignificant because of the larger gnomAD control cohort (eg, 2q13 proximal duplication [*NPHPI*]).

Some CNVs that are generally regarded as fully penetrant were identified in the gnomAD control cohort, which led to an estimation of penetrance that was incomplete (eg, 7q11.23 deletion [Williams-Beuren syndrome], 16p13.3 deletion [Rubinstein-Taybi syndrome], 17p11.2 duplication [Potocki-Lupski syndrome], 17q11.2 deletion [Neurofibromatosis Type 1], 22q11.2 deletion [Velocardiofacial syndrome], and 22q13.33 deletion [Phelan-McDermid syndrome]). Potential reasons for affected individuals of fully penetrant conditions being included in control cohorts include mosaicism, variable expressivity with subclinical symptoms, erroneous inclusion due to lack of proper documentation or phenotyping, data entry error, and laboratory (equipment) error.

Some studies did not specify if CNV prevalence in affected cohorts was significantly different from controls.^{1,8-11} In general, nonsignificant penetrance estimates for CNVs are not included in this review (because there are too many such benign CNVs). An exception is made if a study published penetrance estimates for that CNV, and significance was not listed in the original study. In this case, all studies, regardless of significance, are examined by a forest plot to reduce ascertainment bias.

There is a strong potential for race or demographics to play a role in skewing penetrance estimates for CNVs because population frequencies for CNVs may vary between groups.^{4,30,31} The Ottawa Risk of Bias Assessment ([Supplemental Methods 2.2](#)) identified that affected and control cohorts differ in their racial composition in all studies that reported race, with a greater prevalence identified in affected compared with control cohorts.^{4,6}

One study published penetrance estimates for autism for the 15q11.2 [*NIPA1*, *NIPA2*] deletion and duplication, with

Table 1 Studies publishing penetrance estimates for CNVs associated with neurodevelopmental disability

First Author & Year	Type of Study	No. of CNVs	Control Cohort			Affected Cohort		
			Clinical Phenotype Controlled For	Cohort Size	Age Group	Clinical Phenotype	Cohort Size	Age Group
Al Shehi 2019 ^{1 c}	Case-control	1	Self-reported lack of psychiatric symptoms	65	Adult	Probands with neurodevelopmental delay, autism or congenital anomalies. Relatives of probands with self-reported psychiatric symptoms.	50	Children + Adult
Allach El Khattabi 2018 ²	Case-control	1	Unselected	16,132	Not specified	ID, developmental delay, autism.	16,013	Children + Adult
Chaste 2014 ³	Case-control	2	Autism	2,525 families	Adult + Children	Autistic individuals with relatively little ID.	2,525 families	Children
Cooper 2013 ⁴	Case-control	90 ^a	Unselected	8,329	Adult	73% (ID, autism, developmental delay), 15% (Congenital malformation, hypotonia, feeding difficulties, growth retardation, cardiovascular anomalies, renal anomalies, behavioural issues and other), 12% unspecified.	15,767	Children
Cosemans 2020 ^{5 c}	Case-control	1	ID, DD, autism	31,243	Not specified	ID, DD, autism, schizophrenia	174,223	Children + adult
Isles 2015 ⁶	Case-control	1	Unselected	149,780	Not specified	ID, Developmental delay, autism, attention deficit hyperactivity disorder, epilepsy, multiple congenital anomalies, unspecified. Schizophrenia. ^b	49,995 28,138	Children + adult Not specified
Jonch 2019 ⁷	Case-control	2	Unselected	151,619	Adult	Neurodevelopmental disability and/or malformation.	15,448	Not specified
Kendall 2019 ⁸	Case-control	33	ID, autism, schizophrenia	420,247	Adult	Unclear. See Supplemental Methods 2.2. ^c		^c
Kirov 2014 ⁹	Case-control	70	Unselected	17,873	Not specified	Developmental delay, autism, congenital anomalies. Schizophrenia. ^b	23,380-32,578 ^d 13,465-21,269 ^d	Children Not specified
Martin 2020 ¹⁰	Cohort	25	Same as affected	78,556	Children + adult	ID, communication disorder, autism, attention deficit hyperactivity disorder, specific language disorder, motor disorder, other neurodevelopmental disorder, schizophrenia, bipolar disorder, obsessive compulsive disorder, epilepsy and cerebral palsy.	12,039	Children + adult

(continued)

Table 1 Continued

First Author & Year	Type of Study	No. of CNVs	Control Cohort			Affected Cohort		
			Clinical Phenotype Controlled For	Cohort Size	Age Group	Clinical Phenotype	Cohort Size	Age Group
Maya 2020 ^{11 c}	Case-control	1	Low risk scans (prenatal cohort). Healthy family members (postnatal cohort).	5,962	Prenatal + children + adult	Prenatal anomaly (e.g. raised nuchal translucency) or postnatal anomaly (e.g. ID, autism, congenital malformation).	5,042	Prenatal + Children
Mohan 2019 ¹²	Case-control	2	Partially controlled for ID, DD, autism, epilepsy, dysmorphism.	38,564	Not specified	ID, DD, autism, epilepsy, dysmorphism.	38,564	Not specified
Rosenfeld 2013 ¹³	Case-control	13	Unselected	22,246	Adult	ID, developmental delay, epilepsy, autism, congenital anomalies, dysmorphic features and unspecified.	25,113-48,637 ^d	Not specified
Tropeano 2016 ¹⁴	Case-control	1	Unselected	12594	Not specified	Autism ^b	3631	Adult
						Other neurodevelopmental disorders, excluding those with a co-diagnosis of autism (e.g. developmental delay, neurocognitive disability, attention deficit hyperactivity disorder, psychoses, behavioural abnormalities, speech/language delay, learning disability, motor delay, microcephaly, macrocephaly, structural brain anomaly, seizures, abnormal muscle tone, other neurological problems)	18857	Not specified
Vassos 2010 ¹⁵	Case-control	7	Partially controlled for schizophrenia.	28,406-46,502	Not specified	Schizophrenia ^b	2,977-5,218 ^d	Not specified

Fifteen studies reporting penetrance estimates for 92 CNVs associated with neurodevelopmental disability.¹⁻¹⁵ The primary difference between studies was their definition or selection of an affected cohort and their choice of control cohorts. Nine studies reporting on 83 CNVs had comparable inclusion and exclusion criteria for their cohorts and can be combined in a pooled analysis for this systematic review.^{2-4,6,7,9,10,13,14}

^aCNVs in this study listed as not being statistically significant were excluded from the systematic review, unless another study published penetrance estimates for the CNV without indicating significance.

^bThis subgroup of the dataset was not included in this systematic review.

^cPublished penetrance estimates in these works were unable to be replicated using they study's data. Some of the data regarding the affected or control cohorts may be unclear.

^dThis study published penetrance estimates for multiple CNVs, some of which had different cohort sizes.

Table 2 Genomic coordinates of 83 CNVs from 9 studies^{2-4,6,7,9,10,13,14} and their pooled penetrance estimate compared with gnomAD controls

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
1p36 deletion and 1p36 duplication [GABRD]	Cooper	1:0.00-10.01	10.0	Del	78/32587	0/269885	100% (94-100) Caution: See Supplemental Analysis
	Kirov	1:0.00-10.01	10.0				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	16/15767	0/269885	100% (85-100) Caution: See Supplemental Analysis
1q21.1 proximal deletion and 1q21.1 proximal duplication [RBM8A] ^b	Cooper	(NCBI36) 1:144.0-144.3	Supplemental Analysis	Del	13/15767	53/269885	18% (9-27) Caution: See Supplemental Analysis
	Rosenfeld	(NCBI36) 1:144.0-144.5	Supplemental Analysis				
	gnomAD	1:145.7-146.0	0.33	Dup	85/48637	249/269885	9% (7-11) Caution: See Supplemental Analysis
1q21.1 distal deletion and 1q21.1 distal duplication [GJA5]^{d,b}	Cooper	1:147.10-147.92	0.82	Del	58/27806	71/269885	30% (23-38)
	Rosenfeld	1:147.06-148.41	1.35				
	Martin	1:147.11-147.92	0.81	Dup	42/27806	94/269885	19% (14-25)
	gnomAD	1:147.12-147.84	0.73				
2p16.3 deletion [NRXN1] ^b	Kirov	2:49.91-51.03	Supplemental Analysis	Del	12/6623	45/269885	37% (21-51) Caution: See Supplemental Analysis
	gnomAD	See Supplemental Analysis	Supplemental Analysis				
2q11.2 deletion [TMEM127]^b	Cooper	2:96.06-97.01	0.95	Del	2/15767	16/269885	10%^a (0-27) Not significant
	gnomAD	2:96.11-96.99	0.87				
2q13 proximal deletion and 2q13 proximal duplication [NPHP1] ^b	Cooper	2:110.07-110.23	0.16	Del	78/15767	1533/269885	4% ^a (3.6-5) Not significant
	gnomAD	2:110.12-110.21	0.08				
2q23.1 deletion [MBD5]	Kirov	2:148.0-148.5	0.55	Del	20/32587	0/269885	100% (78-100)
	gnomAD	See Supplemental Analysis	Supplemental Analysis				

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
2q37 deletion and 2q37 duplication [HDAC4]	Cooper	2:238.8-241.5	2.8	Del	20/32587	0/269885	100% (78-100)
	Kirov	2:238.8-241.5	2.8				Caution: See Supplemental Analysis
	gnomAD	2:239.1-241.9	2.8	Dup	2/32587	1/269885	47% ^a (0-100) Not significant. Caution: See Supplemental Analysis
3q29 deletion and 3q29 duplication [DLG1]	Cooper	3:196.02-197.63	1.6	Del	20/44626	6/269885	52% (33-80)
	Kirov	3:196.00-197.61	1.6				Caution: See Supplemental Analysis
	Martin	3:196.03-197.62	1.6	Dup	18/32587	4/269885	67% (45-92)
	gnomAD	3:196.05-197.55	1.5				Caution: See Supplemental Analysis
4p16.3 deletion (Wolf-Hirschhorn Syndrome) and 4p16.3 duplication	Cooper	4:1.87-2.01	0.14	Del	17/32587	0/269885	100% (74-100)
	Kirov	4:1.53-2.03	0.50	Dup	4/32587	3/269885	37% (8-100)
	gnomAD	4:1.80-2.30	0.33				
5q35.3 deletion (Sotos Syndrome) and 5q35.3 duplication [NSD1]	Cooper	5:176.29-177.63	1.3	Del	14/32587	0/269885	100% (69-100)
	Kirov	5:176.29-177.63	1.4	Dup	4/32587	0/269885	100% (31-100) Caution: See Supplemental Analysis
	gnomAD	5:176.31-178.00	1.7				
6p25 deletion and 6p25 duplication	Kirov	6:0.16-6.06	5.9	Del	23/32587	0/269885	100% (80-100)
	gnomAD	N/A	N/A	Dup	12/32587	0/269885	100% (69-100) Caution: See Supplemental Analysis
6q16 deletion and 6q16 duplication [SIM1]	Cooper	6:100.37-100.50	0.13	Del	1/23380	3/269885	17% ^a (0-100)
	Kirov	6:100.39-100.46	0.07				Not significant
	gnomAD	See Supplemental Analysis	See Supplemental Analysis	Dup	1/23380	5/269885	11% ^a (0-48) Not significant
7q11.23 deletion (Williams-Beuren Syndrome) and 7q11.23 duplication	Cooper	7:73.33-74.73	1.4	Del	86/44626	2/269885	93% ^c (85-100)
	Kirov	7:73.33-74.73	1.4				Caution: See Supplemental Analysis
	Martin	7:73.33-74.73	1.4	Dup	43/44626	7/269885	67% ^c (50-87)
	gnomAD	7:73.33-74.72	1.4				Caution: See Supplemental Analysis

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
8p23.1 deletion and 8p23.1 duplication [<i>CLDN23, SOX7, GATA4</i>]	Cooper	8:8.24-12.04	3.8	Del	18/32587	0/269885	100% (77-100) See Supplemental Analysis
	Kirov	8:8.23-12.03	3.8				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	24/32587	0/269885	100% (79-100) Caution: See Supplemental Analysis
9q34 deletion (Kleefstra syndrome) and duplication [EHMT1]	Cooper	9:134.92-138.19	3.3	Del	18/32587	0/269885	100% (76-100)
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	8/32587	0/269885	100% (53-100) Caution: See Supplemental Analysis
10q23 deletion and 10q23 duplication [<i>NRG3, GRID1, BMPR1A</i>]	Cooper	10:80.20-87.04	6.8	Del	28/44626	2/269885	82%^c (62-100) Caution: See Supplemental Analysis
	Kirov	10:80.20-87.04	6.8				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	5/32587	6/269885	27% (6-61) Caution: See Supplemental Analysis
13q12 deletion [<i>CRYL1</i>] ^b	Cooper	13:20.24-20.44	0.2	Del	14/15767	227/269885	5% ^a (3-8) Not significant
gnomAD	See Supplemental Analysis	Supplemental Analysis					
15q11.2 deletion [BP1-BP2] and 15q11.2 duplication [NIPA1, NIPA2]	Cooper	15:22.78-23.07	0.29	Del	312/40561	956/269885	10% (9-12)
	Rosenfeld	15:22.62-23.12	0.50				
	Chaste	Unclear	Unclear				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	149/31215	1331/269885	5%^a (4-6) Not significant
15q11.2 deletion [BP1-BP3] (Prader-Willi Syndrome/Angelman Syndrome) and 15q11.2 duplication [NIPA1, NIPA2]	Isles	Unclear	Unclear	Del	4/12039	0/269885	100% (55-100)
	Martin	15:22.78-28.14	5.4				
	gnomAD	15:22.81-28.28	5.5	Dup	56/63034	11/269885	54% (40-73) Caution: See Supplemental Analysis

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
15q11q13 deletion [BP2-BP3] (Prader-Willi Syndrome/Angelman Syndrome) and 15q11.13 duplication^b	Cooper	15:24.57-28.18	3.6	Del	60/32587	0/269885	100% (92-100)
	Kirov	15:24.57-28.18	3.6				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	82/32587	11/269885	77% (65-88) Caution: See Supplemental Analysis
15q13.3 deletion [BP4-BP5] and 15q13.3 duplication [<i>CHRNA7</i> , <i>OTUD7A</i>] ^b	Cooper	15:30.84-32.19	1.35	Del	100/44626	27/269885	55% (45-66) Caution: See Supplemental Analysis
	Kirov	15:30.84-32.19	1.35				
	Martin gnomAD	15:30.84-32.15 15:30.63-32.11	1.31 1.49	Dup	27/32587	140/269885	8% (5-11) Caution: See Supplemental Analysis
15q13.3 smaller deletion and 15q13.3 smaller duplication [<i>CHRNA7</i> and <i>OTUD7A</i> only]^b	Kirov	15:31.72-32.16	0.44	Del	Supplemental Analysis	Supplemental Analysis	Caution: See Supplemental Analysis
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	Supplemental Analysis	Supplemental Analysis	Caution: See Supplemental Analysis
15q24 deletion and 15q24 duplication [<i>BBS4</i> , <i>PML</i> , <i>SIN3A</i>]	Cooper	15:72.62-75.50	2.88	Del	10/44626	0/269885	100% (53-100)
	Kirov	15:72.62-74.12	1.50				
	Martin gnomAD	15:72.67-75.68 15:72.69-75.21	3.01 2.53	Dup	4/32587	3/269885	37% (7-100) Caution: See Supplemental Analysis
15q24.2q24.5 deletion and 15q24.2q24.5 duplication [<i>FBXO22</i>, <i>TSPAN3</i>]	Cooper	15:75.68-77.91	2.23	Del	5/32587	1/269885	69% (28-100) Caution: See Supplemental Analysis
	Kirov	15:75.68-77.91	2.23				
	gnomAD	15:75.84-77.63	1.79	Dup	6/32587	1/269885	73% (31-100) Caution: See Supplemental Analysis

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
15q25.2 proximal deletion and 15q25.2 proximal duplication [RPS17, HOMER2, BNC1]	Cooper	15:82.51-84.07	1.56	Del	2/23380	0/269885	100% ^a (0-100)
	Kirov	15:82.51-84.07	1.56				Not significant
	gnomAD	15:82.54-84.04	1.50	Dup	4/23380	1/269885	71% (24-100) Caution: See Supplemental Analysis
16p13.3 deletion (Rubinstein-Taybi Syndrome) [CREBBP]	Cooper	16:3.73-3.81	0.08	Del	10/32587	1/269885	82%^c (48-100)
	Kirov	16:3.73-3.88	0.15				Caution: See Supplemental Analysis
	gnomAD	16:3.73-3.81	0.08				
16p13.11 deletion and 16p13.11 duplication [MYH11] ^d	Cooper	16:15.41-16.20	0.79	Del	77/45265	83/269885	23% (18-29)
	Rosenfeld	16:14.90-16.40	1.5				Caution: See Supplemental Analysis
	Kirov	16:15.42-16.21	0.79				
	Allach El Khattabi	See Supplemental Analysis	Supplemental Analysis	Dup	143/48600	398/269885	10% (8-11)
	Martin	16:15.42-16.20	0.78				Caution: See Supplemental Analysis
gnomAD	16:15.03-16.20	1.17					
16p12.2 deletion (previously 16p12.1 deletion) and 16p12.2 duplication (previously 16p12.1 duplication) [CDR2]^b	Cooper	16:21.93-22.45	0.52	Del	62/33226	139/269885	16% (12-20)
	Rosenfeld	16:21.93-22.48	0.55				Caution: See Supplemental Analysis
	Kirov	16:21.93-22.45	0.52	Dup	16/32587	135/269885	5%^a (3-7)
	gnomAD	16:21.95-22.37	0.42				Not significant
16p11.2p12.2 deletion (previously 16p11.2p12.1 deletion) and 16p11.2p12.2 duplication (previously 16p11.2p12.1 duplication)	Cooper	16:21.34-29.43	8.09	Del	20/32587	0/269885	100% (78-100)
	Kirov	16:21.52-29.09	7.57				Dup
	gnomAD	See Supplemental Analysis	Supplemental Analysis				

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
16p11.2 distal deletion and 16p11.2 distal duplication [SH2B1]^{b,d}	Cooper	16:28.76-29.10	0.34	Del	49/45265	32/269885	33% (24-44) Caution: See Supplemental Analysis
	Rosenfeld	16:28.73-29.08	0.35				
	Martin	16:28.81-29.04	0.23	Dup	35/33226	81/269885	16% (11-22) Caution: See Supplemental Analysis
	gnomAD	16:28.81-28.99	0.18				
16p11.2 proximal deletion and 16p11.2 proximal duplication [TBX6] ^{b,d}	Cooper	16:29.64-30.19	0.55	Del	163/45265	58/269885	47% (40-55) Caution: See Supplemental Analysis
	Rosenfeld	16:29.58-30.23	0.65				
	Martin	16:29.64-30.19	0.55	Dup	114/45265	73/269885	33% (28-41)
	gnomAD	16:29.66-30.19	0.52				
17p13.3 deletion and 17p13.3 duplication [YWHAE]	Cooper	17:0.65-1.45	0.80	Del	7/32587	0/269885	100% (47-100)
	Kirov	17:1.35-1.40	0.05				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	6/15767	8/269885	41% (15-65) Caution: See Supplemental Analysis
17p13.3 deletion and 17p13.3 duplication [PAFAH1B1]	Cooper	17:2.46-3.02	0.56	Del	8/32587	0/269885	100% (53-100)
	Kirov	17:2.59-2.69	0.10				
	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	4/15767	1/269885	79% (34-100) Caution: See Supplemental Analysis
17p12 deletion (Hereditary Neuropathy with Liability to Pressure Palsies) and 17p12 duplication (Charcot Marie Tooth) [PMP22]^b	Cooper	17:14.17-15.60	1.43	Del	3/15767	146/269885	2%^c (0-4) Caution: See Supplemental Analysis
	Martin	17:14.19-15.52	1.33				
	gnomAD	17:14.21-15.51	1.30	Dup	18/27806	75/269885	11%^c (6-17) Caution: See Supplemental Analysis

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
17p11.2 deletion (Smith-Magenis Syndrome) and 17p11.2 duplication (Potocki-Lupski Syndrome) [<i>RAI1</i>]	Cooper	17:16.81-20.58	3.77	Del	35/44626	0/269885	100% (82-100)
	Kirov	17:16.92-18.38	1.46	Dup	25/32587	1/269885	92% ^c (75-100) Caution: See Supplemental Analysis
	Martin	17:16.91-20.30	3.39				
	gnomAD	17:16.93-20.31	3.38				
17q11.2 deletion (Neurofibromatosis Type 1) and 17q11.2 duplication [<i>NF1</i>]	Cooper	17:30.84-31.89	1.05	Del	27/44626	2/269885	81% ^c (61-100) Caution: See Supplemental Analysis
	Kirov	17:30.77-31.95	1.18	Dup	35/44626	4/269885	74% (54-93) Caution: See Supplemental Analysis
	Martin	17:30.78-31.94	1.16				
	gnomAD	17:30.78-32.01	1.22				
17q12 deletion (Renal Cysts and Diabetes) and 17q12 duplication [<i>HNF1B</i>] ^{b,d}	Cooper	17:36.46-37.85	1.39	Del	32/45265	6/269885	63% (45-85) See Supplemental Analysis
	Rosenfeld	17:36.36-37.87	1.51	Dup	48/45265	61/269885	20% (14-27)
	Martin	17:36.46-37.85	1.39				
	gnomAD	17:36.49-37.75	1.26				
17q21.31 deletion (Koolen-de Vries Syndrome) and 17q21.31 duplication [<i>MAPT, KANSL1</i>]	Cooper	17:45.63-46.11	0.48	Del	42/32587	0/269885	100% (88-100)
	Kirov	17:45.62-46.10	0.48	Dup	5/32587	1/269885	69% (31-100)
	gnomAD	See Supplemental Analysis	Supplemental Analysis				
17q23 deletion and 17q23 duplication [<i>TBX2, TBX4</i>]	Kirov	17:60.17-62.21	2.04	Del	6/32587	0/269885	100% (47-100)
	gnomAD	17:60.04-62.14	2.10	Dup	1/32587	1/269885	31% ^a (0-100) Not significant Caution: See Supplemental Analysis
19p13.12 deletion	Kirov gnomAD	19:12.97-16.59 See Supplemental Analysis	3.62 Supplemental Analysis	Del	13/32587	0/269885	100% (67-100)

(continued)

Table 2 Continued

Copy Number Variant	Study	Genomic Coordinates (GRCh38) Mb	Size (Mb)	Deletion/Duplication	Pooled Affected Cohort	gnomAD Control Cohort	Penetrance (95% Confidence Interval) %
22q11.2 deletion (Velocardio-facial Syndrome) and 22q11.2 duplication [TBX1] ^b	Cooper	22:19.03-20.30	1.27	Del	192/44626	10/269885	86% ^c (79-94) Caution: See Supplemental Analysis
	Rosenfeld	22:18.34-21.22	2.88				
22q11.2 distal deletion and 22q11.2 distal duplication [BCR, MAPK1]	Kirov	22:19.03-20.27	1.24	Dup	167/60676	167/269885	19% (16-23) Caution: See Supplemental Analysis
	Martin	Unclear	Unclear				
22q11.2 distal deletion and 22q11.2 distal duplication [BCR, MAPK1]	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	167/60676	167/269885	19% (16-23) Caution: See Supplemental Analysis
	Cooper	22:21.56-23.31	1.75				
22q13.33 deletion Phelan-McDermid Syndrome and 22q13.33 duplication [SHANK3]	Kirov	22:21.56-23.31	1.75	Dup	18/35419	7/269885	51% (32-77) Caution: See Supplemental Analysis
	Martin	22:21.44-23.31	1.87				
Xp22.3 duplication [SHOX]	gnomAD	22:21.56-23.38	1.82	Dup	Supplemental Analysis	Supplemental Analysis	Caution: See Supplemental Analysis
	Tropeano	X:0.62-0.66	0.04				
22q13.33 deletion Phelan-McDermid Syndrome and 22q13.33 duplication [SHANK3]	Cooper	22:50.67-50.73	0.06	Del	45/15767	3/269885	93% ^c (84-100) Caution: See Supplemental Analysis
	Kirov	22:42.60-50.72	8.12				
Xp22.3 duplication [SHOX]	gnomAD	See Supplemental Analysis	Supplemental Analysis	Dup	Supplemental Analysis	Supplemental Analysis	Caution: See Supplemental Analysis
	Tropeano	X:0.62-0.66	0.04				

Penetrance is for a combination of ID, DD, autism, congenital malformations and miscellaneous other phenotypes, depending on the studies involved in contributing data. For each of these 83 CNVs, the [Supplemental Analysis](#) provides a forest plot showing each study that contributed to this final pooled penetrance estimate, an analysis of pathogenicity and penetrance, as well as any special considerations relevant for that CNV. The bolding in the table is to help with readability.

NCBI36: Coordinates provided by original study in NCBI36/hg18, which does not lift over appropriately to hg38 (See [Supplemental Analysis](#) for the 1q21.1 proximal deletion and duplication [RBM8A]).

^aThe prevalence of this CNV in control and affected cohorts are not significantly different ($P > .05$). Equivalently, the lower 95% confidence interval for penetrance is less than or equal to 5.08%, with further details provided in [Supplemental Methods 3.2](#).

^bOne or more studies publishing penetrance estimates for this CNV have been excluded from this systematic review as outlined in [Supplemental Methods 2.2](#).

^cThis CNV is likely to be fully penetrant. See forest plot and accompanying synopsis for this CNV in the [Supplemental Analysis](#).

^dThe study by Kirov et al.⁹ is not included because the affected cohort for this CNV is identical to a different study.¹³

their affected cohort comprising individuals who had, on average, “moderate to severe autistic symptoms with relatively little intellectual disability.”^{3,32} Penetrance estimates were significantly lower than that of other studies that had affected cohorts comprising individuals with ID, DD, autism, and congenital malformations. These differences are shown in the forest plot in [Supplemental Analysis](#) for the relevant CNVs and a discussion provided to justify exclusion from pooled analysis where appropriate.

Comparing different studies in a systematic review has been noted to be challenging because of inherent differences in methodology and study population, but exclusion of too many studies defeats the purpose of conducting a systematic review. Overall, the majority of individuals in affected cohorts had learning difficulties,^{1-10,12-14} justifying the use of a pooled analysis to compare penetrance estimates between cohorts in the 9 selected studies.^{2-4,6,7,9,10,13,14}

Discussion

Composition of affected and control cohorts

Direct comparisons between studies can be challenging because different studies enrolled affected patients with different types of neurodevelopmental conditions. The DSM-V definition of a neurodevelopmental disorder includes ID, DD, autism, ADHD, communication disorder, specific learning disorder, and motor disorder (eg, tics). However, because penetrance estimates of CNVs are primarily based on chromosomal microarray data, and this diagnostic test is not commonly used for conditions such as ADHD, motor disorder, or mild autism,¹⁻¹⁵ it is inaccurate to assume that a CNV predisposes to conditions matching the DSM-V definition of a neurodevelopmental disorder. Instead, the evidence available regarding penetrance for the CNV is limited to the condition observed in the affected cohort of the study.

All affected cohorts identified by this systematic review contain at least some individuals with a neurological phenotype.¹⁻¹⁵ Within an affected cohort, most individuals have ID, DD, and/or autism.^{1-10,12-14} A minority of individuals in these affected cohorts may have congenital malformations,^{1,4,7,11,13,14} behavioral disorders,^{4,6,10,14} or an unspecified phenotype^{4,13} without ID/DD/autism. It is unclear why those with an unspecified phenotype would be included in an affected cohort and whether this may include healthy parents of children who have a CNV, or individuals referred for microarray without a listed phenotype on the pathology form. Both reasons have potential to skew penetrance estimates upward, making comparison between studies challenging.

Comparison between studies can also be challenging because of different choices of control cohorts. Affected individuals can sometimes be erroneously included in control cohorts because of phenotypes that are undocumented in the medical record¹⁰ or if study investigators do not directly

assess the phenotype of those in the control cohort.^{2,4,7,9,13-15}

This can be particularly problematic because most control cohorts used by studies in this systematic review comprise adults attending outpatient departments for conditions unrelated to neurodevelopment,^{2,4,9,10,13-15} but outpatient departments may be enriched for individuals with congenital malformations or comorbid ID. Other reasons for affected individuals being present in control cohorts may include mosaicism for the CNV,³³ variable expressivity, mild or subclinical phenotypes, presymptomatic individuals (for adult-onset conditions), preanalytic or analytic error (wrong blood sample or incorrect assay result), or data entry error. Any of these reasons can explain the identification of fully penetrant CNVs in control populations. Less likely possibilities could include an expansion of the phenotype beyond which is currently appreciated in the medical literature, to include individuals who are truly nonpenetrant, or complex reasons for nonpenetrance that may include imprinting, digenic, sex-based, or genetic-environmental associations that have not been well studied yet for CNVs. Including affected individuals in control cohorts will reduce penetrance estimates that reference these datasets. This is particularly problematic for CNVs that have neurodevelopmental and nonneurodevelopmental phenotypes (eg, the 17q12 deletion [*HNF1B*] that causes renal cysts and diabetes) because individuals with renal cysts and diabetes may be included in the control cohort. However, it is also possible that there could be an effect in the opposite direction: individuals with nonneurodevelopmental phenotypes could be depleted from control cohorts because of their condition (eg, renal failure), and this could result in falsely elevated penetrance estimates.

Unless studies specifically screen their affected cohorts to include and their control cohorts to exclude neurodevelopmental phenotypes, the true penetrance may lie outside their published 95% confidence intervals.

The impact of familial variants and ethnicity in cohorts

There is potential for benign CNVs to be reported as highly penetrant because of racial or demographic differences between affected and control cohorts. When race in the studies identified by this systematic review was reported on, affected cohorts were always documented to contain a higher prevalence of minority races compared with controls.^{4,6,19} The reasons for this are complex but likely revolve around a more racially diverse cohort of affected young children compared with a less racially diverse control cohort of adults. Sampling the affected cohort from 1 city while pooling the control cohort from different cities may result in a benign founder CNV being enriched in the affected cohort. This will result in the CNV being incorrectly assessed as being pathogenic and could artificially inflate penetrance estimates. Similarly, the methods section in most articles do not specify if steps were taken to ensure

that multiple family members were excluded. Some CNVs reporting only 2 or 3 individuals in 1 arm of the study could be due to 1 or more of these biases, potentially leading to erroneous estimations of high penetrance for a benign CNV.

Although minor differences are expected, large differences in penetrance estimates between studies suggest that their cohorts may not be drawn from the same population. Examples are the 17q12 deletion [*HNF1B*] for renal cysts and diabetes, 3q29 deletion [*DLG1*], and 15q13.3 deletion [BP4-BP5] [*CHRNA7*], which have very different prevalences of the CNV in different studies, with more details provided in [Supplemental Analysis](#).

The gnomAD v4.0 CNV control dataset used in this systematic review is also subject to this bias. Racial breakdowns are shown in [Supplemental Methods 3.3](#).

Choice of the value of P(D), the prevalence of disease in the general population

The chosen value of P(D) represents the proportion of the general population, which has the disease D. Across different studies, P(D) was noted to take a value of 13%,¹⁰ 5.3%,⁴ 5.12%,^{11,13} 4%,^{6,8,9,14} 3%,⁵ or 1%.³ The choice of the parameter P(D) likely represents the single largest source of potential inaccuracy when using the Bayesian formula for penetrance. This range of values is problematic because penetrance estimates are challenging to compare across studies that use different values of P(D). For example, if 2 studies have identical data, the choice of a P(D) value of 5% in 1 study would yield an approximately 5-fold higher penetrance estimate compared with a value of 1% in another study. Consequently, researchers' choice of this parameter can introduce substantial variation in reported penetrance estimates. Notably, penetrance for most CNVs in this systematic review were computed using P(D) values between 4% to 5.3%, potentially leading to a 4- or 5-fold higher penetrance compared with those using 1%.

Correcting for this in a systematic review is challenging and best performed during study recruitment because P(D) depends on the study's choice of an affected cohort. Within the framework of a PRISMA approach, we have aimed to identify why this is a problem and to inform future studies. For now, this systematic review used the study's chosen value of P(D) when reporting penetrance using their data and the median value of P(D) for calculations involving the pooled analysis. Reasons are provided in [Supplemental Methods 3.2](#).

Publication bias

One limitation of systematic reviews is their tendency to favor published data, potentially leading to publication bias. Some rare CNVs (eg, 2q11.2 deletion [*TMEM127*]), have been published with penetrance estimates in only 1 or 2 studies. As a rare CNV, there may only be a few individuals in the study, and minor differences in data can result in very different penetrance estimates. If data for these CNVs are

not included in some studies, either because of small sample sizes, nonstatistically significant results, or because penetrance was 0% or 100% (in some studies, CNVs with 0% or 100% penetrance are excluded^{8,13}), then a systematic review of published penetrance estimates may be analyzing data from outlier results, rather than the more accurate (but unpublished) 0% or 100% observation from other studies. For example, a single study⁹ reported conflicting results regarding the 15q13.3 deletion [*OTUD7A*, *CHRNA7*] with a 1.5 Mb deletion showing data supporting an estimated 37% penetrance, whereas a smaller 0.44 Mb deletion was reported to be 100% penetrant. Biologically, it is hard to imagine a smaller deletion being fully penetrant when the larger deletion (encompassing the smaller deletion) is not. There are many instances of suspected publication biases identified in this systematic review, and these are reviewed in [Supplemental Analysis](#) for each relevant CNV.

Publication bias may also be present when prevalences of the CNV are similar in affected vs control cohorts (because many studies will not publish such data), whereas those observing a significant difference may be more likely to be published. This may be the case with the 15q11.2 duplication [NIPA1, NIPA2],^{3,4,7} 2q13 proximal deletion and duplication [*NPHPI*],⁴ 13q12 deletion [*CRYLI*],⁴ 15q13.3 smaller duplication [*CHRNA7*],⁹ Xp22.3 duplication [*SHOX*],¹⁴ and 16p13.11 duplication [*MYH11*].^{2,4,9} A case-by-case analysis is provided for each of these CNVs in their respective synopses ([Supplemental Analysis](#)).

It is interesting to note that a large population-based study involving 24,877 individuals did not find an association for ID or autism between individuals harboring at least 1 neurodevelopmental CNV and those who did not have a CNV (95% CI for odds ratio 0.7-24.7),³⁴ although confidence intervals are noted to be wide. This suggests that many common neurodevelopmental CNVs have very low penetrance or are nonpenetrant for ID or autism. In contrast, the same dataset showed a minor statistically significant difference for congenital malformations (95% CI for odds ratio 1.1-3.5).³⁴

Penetrance estimates should be interpreted cautiously for minimally penetrant CNVs and for CNVs in which control or affected cohorts have very few or 0 reported individuals, especially if comparable large studies do not report penetrance data for the CNV.

CNVs with statistically insignificant penetrance estimates

Some studies reported CNV penetrance without clearly specifying that these results were not statistically significant (eg, 100% penetrance of 15q24 deletion [*BBS4*, *PML*, and *SIN3A*] with CI 0%-100%¹⁰). In the context of genome-wide testing of many CNVs, even a requirement of $P > 0.5$ would result in numerous false positives. Future studies could consider stronger standards for p-values to reduce false-positive errors by using a Bonferroni correction, Holm-Bonferroni method, or Benjamini-Hochberg method.

CNVs with low penetrance that meet statistical significance

Statistical significance is achieved when the 95% confidence interval for a CNV is strictly greater than the value of the background risk of disease, P(D).

As a case example, the 16p13.11 duplication [*MYH11*] is reported with penetrance estimates of 6.6%,⁴ 8%,² and 9%.⁹ After pooling nonoverlapping affected cohorts and comparing this with the gnomAD control cohort, a 10% (95% CI 8.0%-10%) penetrance is estimated, which is statistically significant. However, this should be interpreted cautiously. All 3 studies use a high value of 4% to 5.3% for P(D), which could lead to inflated penetrance estimates as previously discussed. In the study reporting 8% penetrance,² the affected cohort were individuals with ID, DD, and autism. The other 2 studies^{4,9} used a similar value of P(D) for a much wider range of conditions that included congenital malformations and miscellaneous and unspecified issues. A smaller choice of P(D) may have been more appropriate for all 3 studies, especially for the study using only individuals with ID/DD/ASD.² Although there is no easy way to correct for this after data collection by the studies, it is possible that a lower penetrance estimate for this CNV may not meet significance when limitations in the way penetrance have been calculated are taken into account. Large prospective studies of *MYH11* haploinsufficiency or triplosensitivity are rare, but it is interesting to note that of 16 fetuses identified with an *MYH11* deletion or duplication in 1 study, none had ID or autism on follow-up 6 months to 4 years afterward,³⁵ supporting caution when assigning pathogenicity to low-penetrant CNVs, especially in the prenatal setting.¹¹

Clinically, there have been many instances in which individuals with a low-penetrant CNV have subsequently had genomic testing showing a more likely monogenic cause for their neurodevelopmental issues.^{36,37} Care should therefore be used when assigning pathogenicity to low-penetrant CNVs,³⁸ even when their penetrance estimates meet statistical criteria for significance.

CNVs with no detailed case reports describing the phenotype

Penetrance of CNVs with no supporting case reports to document pathogenicity should be interpreted cautiously. For example, penetrance of the 6q16 duplication [*SIMI*] is estimated at 11% (95% CI 0%-48%). There are no publications that describe individuals with this CNV and the implied pathogenicity, and a penetrance of 11% may represent a false-positive result. In general, it may be safer to interpret such data as having an unclear penetrance of 0% to 48%, rather than as a static number.

More broadly, many recurrent CNVs do not have a specific reported phenotype. Instead, many are said to

occasionally cause some combination of DD, ID, autism, congenital cardiac disease, renal, ophthalmic, or other problems.¹⁶ For many low-penetrant but pathogenic CNVs, good health could be considered the most common phenotype or the most common outcome. Examples of low-penetrant CNVs include the 16p13.11 [*MYH11*] deletion or duplication, 16p11.2 proximal [*TBX6*] deletion or duplication, 16p11.2 distal [*SH2B1*] duplication, and the 15q13.3 [*CHRNA7*] deletion or duplication, with penetrance data suggesting that it is much more common to find individuals with these CNVs in control population datasets than in affected cohorts.^{4,10,13,16} Taking these factors into account, it is suggested that clinical interpretation of whether a CNV with low penetrance is responsible for a nonspecific phenotype in a patient is carefully considered because a different cause may be possible.

Age-related (adult) onset for the phenotype

The 17p12 deletion and duplication [*PMP22*] cause Hereditary Neuropathy with Liability to Pressure Palsies (HNPP) and Charcot Marie Tooth Disease, respectively. These conditions are well studied and have been reported to be fully penetrant.^{39,40} The gnomAD v4.0 CNV control database includes many individuals with these CNVs, leading to penetrance estimates of 2% (95% CI 0%-4%) and 11% (95% CI 6%-17%), respectively. This error is due to the inclusion in gnomAD of individuals who are presymptomatic or whose phenotype is mild (HNPP in particular would be unlikely to result in the exclusion of an individual from gnomAD). As a result, this method of determining penetrance cannot be used for phenotypes with age-related onset or relatively mild conditions. Although beyond the scope of this systematic review, this observation may extend to penetrance estimates for other adult-onset conditions, such as schizophrenia, depression, and anxiety.

Challenges in translating penetrance estimates derived in a research setting to clinical practice

An important aspect of clinical practice concerns prospective inheritance of a CNV with incomplete penetrance and risk counseling regarding this. Consider a family with an affected child who has a CNV with 20% penetrance that was inherited from a parent with a mild phenotype. In contrast, a different family may have the same CNV with no affected family members. Some may wonder if the likelihood of a relevant phenotype occurring in a newborn child identified prenatally with this CNV is 20% regardless of the scenario, whereas others might wonder if there are known mechanisms that could affect the chance of penetrance for this CNV. For instance, it is possible that other genetic variants, inherited from either parent, may interact with the CNV in a way that increases the likelihood of a phenotype — in other words, the condition may be recessive (eg, thrombocytopenia absent radius syndrome with a 1q21.2 [*RBM8A*]

deletion), oligogenic or polygenic. Additionally, chromosomal duplications may cause monogenic disease by insertion into a gene that is potentially on a different chromosome (thereby disrupting gene function) or be associated with nearby complex structural rearrangements, such as an inversion which disrupts gene function. Such scenarios will likely result in 100% penetrance for the CNV within the family but could be as low as 0% for families without an associated monogenic gene disruption. CNVs may also alter topologically associated domains, and this disruption could lead to disease.⁴¹ Finally, some CNVs are flanked by variations in repetitive sequences that have been postulated to modulate the phenotype of the CNV.⁴² In current clinical care, such scenarios are rarely investigated because of technical difficulties with current technology and suspected low yield.

These examples demonstrate the possibility that the apparent incomplete penetrance of CNVs could be due to some families harboring highly penetrant pathogenic changes that are co-inherited in a dominant manner with the CNV. The studies in this systematic review are unable to determine if gene disruption resulting in 100% penetrance in a minority of families with this CNV may average out to appear as low penetrance when including benign versions of the CNVs in other families. Clarifying this is urgently needed if penetrance estimates of CNVs are to be useful in the prenatal setting.

Similarly, CNV penetrance for a deletion may vary depending on the prevalence of hypomorphic alleles in that geographic region or ethnic group. The 1q21.2 [*RBM8A*] deletion for recessive thrombocytopenia-absent radius syndrome could be one such example, with multiple hypomorphic *RBM8A* alleles common in European (but less common in other) populations.^{22,43} Therefore, penetrance may not necessarily be a property of the CNV per se but could be more related to differing prevalences of hypomorphic alleles in racial subgroups for some CNVs.

In many clinical settings and countries, an individual with ID and a low-penetrant CNV may be a less likely candidate to receive genomic testing for a monogenic cause. If future research clarifies that the penetrance of the CNV is very low or nonsignificant, then the individual may be a candidate for further genomic testing.

Imprecise penetrance estimates have the potential to be detrimental by limiting diagnostic options for affected individuals. They may also cause undue harm to families in the prenatal setting by complicating prenatal care and causing unnecessary anxiety. Multiple issues need to be clarified before these estimates can be used optimally in clinical practice.

Future research

The most accurate study of penetrance would involve a prospective study to follow an unselected birth cohort until at least 5 years of age to determine the phenotype associated with CNVs. Prospective genetic studies have been

performed in this field,^{44,45} although the high dropout rate over many years was noted to introduce a bias that made penetrance difficult to estimate.⁴⁵

Other avenues of research include using existing data with more robust mathematics to inform more accurate penetrance estimation. Standardizing criteria for defining the breakpoints and phenotype of a CNV can help with the comparison of the CNV between different studies. The affected cohort's inclusion criteria should relate to childhood neurodevelopmental issues. Conditions which primarily affect adults (eg, schizophrenia, depression, and anxiety) are probably best studied separately using cohorts that exclude children and by using a different mathematical formula more suited to this purpose.

Conclusion

This systematic review identified 15 studies that published penetrance estimates for neurodevelopmental CNVs, with 9 using comparable methodologies for 83 CNVs.^{2-4,6,7,9,10,13,14} Although these studies represent the best available evidence to inform penetrance estimates of CNVs, all studies have limitations. The main limitations identified by this systematic review include the choice of affected and control cohorts, potential racial or regional demographic variations, small sample sizes for some CNVs, miscellaneous forms of biases, and challenges in determining statistical significance.

Additional challenges include translating penetrance estimates derived in a research setting to that of clinical care. It may be more appropriate to consider the penetrance of a CNV as lying within a range of possible percentages with 95% confidence, rather than as a precise number. Even then, the 95% confidence interval for these ranges may be incorrect because of imperfect study data or publication bias. The pathogenicity of low-penetrant CNVs should therefore be interpreted with caution if this may influence patient care. In the prenatal setting, low-penetrant CNVs may not be pathogenic, and some guidelines suggest not reporting these.^{11,46} In the postnatal setting, identification of a low-penetrant CNV should not preclude an individual from further genomic testing. CNVs reported as being penetrant in the literature without a consistent published phenotype should be interpreted with caution. CNVs reported as penetrant because of a small number of affected or control individuals in some studies (but which remain unreported in other comparable studies in the field) should also be interpreted with caution because of the potential for publication bias.

The pooled analysis provided by this systematic review provides supporting evidence of pathogenicity for the 15q24 duplication [*BBS4*, *PML*, *SIN3A*] [A-D breakpoints], 15q24.2q24.5 deletion and duplication [*FBXO22*, *TSPAN3*], 17q11.2 duplication [*NFI*], 17q21.31 duplication [*MAPT*, *KANSL1*] and 22q11.2 distal duplication [*BCR*, *MAPK1*]. It provides evidence for the benign nature of the 15q11.2 duplication [*NIPA1*, *NIPA2*] and 2q13 proximal duplication [*NPH1*].

Overall, data on CNVs from large-scale studies can help to inform the assessment of pathogenicity and provide a guide for penetrance estimation. However, there may be limitations that can sometimes affect their interpretation in a clinical setting. Future studies may help bridge the gap between research and clinical utility.

Data Availability

All relevant data are listed within the article and its supporting information files ([Supplemental Table 1](#) and [Supplemental Analysis](#)).

Funding

No sources of funding were used in this study.

Author Contributions

Conceptualization: S.G., E.K., M.P.; Methodology: S.G., E.K., M.P.; Data Curation: S.G., L.T.; Formal Analysis: S.G., L.T.; Investigation: S.G., L.T.; Resources: S.G.; Software: S.G.; Supervision: E.K., M.P., T.D.; Validation: S.G., L.T.; Visualization: S.G.; Writing-original draft: S.G.; Writing-review and editing: L.T., E.K., M.P., T.D.

Ethics Declaration

All data used in this study are publicly available. There are no other relevant ethical declarations.

Conflict of Interest

All authors declare no conflicts of interest.

Additional Information

The online version of this article (<https://doi.org/10.1016/j.gim.2024.101227>) contains supplemental material, which is available to authorized users.

References

- Al Shehhi M, Forman EB, Fitzgerald JE, et al. NRXN1 deletion syndrome; phenotypic and penetrance data from 34 families. *Eur J Med Genet.* 2019;62(3):204-209. <http://doi.org/10.1016/j.ejmg.2018.07.015>
- Allach El Khattabi L, Heide S, Caberg JH, et al. 16p13.11 microduplication in 45 new patients: refined clinical significance and genotype-phenotype correlations. *J Med Genet.* 2018;57(5):301-307. <http://doi.org/10.1136/jmedgenet-2018-105389>
- Chaste P, Sanders SJ, Mohan KN, et al. Modest impact on risk for autism spectrum disorder of rare copy number variants at 15q11.2, specifically breakpoints 1 to 2. *Autism Res.* 2014;7(3):355-362. <http://doi.org/10.1002/aur.1378>
- Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet.* 2013;132(10):1077-1130. <http://doi.org/10.1007/s00439-013-1331-2>
- Cosemans N, Vandenhove L, Vogels A, et al. The clinical relevance of intragenic NRXN1 deletions. *J Med Genet.* 2020;57(5):347-355. <http://doi.org/10.1136/jmedgenet-2019-106448>
- Isles AR, Ingason A, Lowther C, et al. Parental origin of interstitial duplications at 15q11.2-q13.3 in schizophrenia and neurodevelopmental disorders. *PLoS Genet.* 2016;12(5):e1005993. <http://doi.org/10.1371/journal.pgen.1005993>
- Jønch AE, Douard E, Moreau C, et al. Estimating the effect size of the 15Q11.2 BP1-BP2 deletion and its contribution to neurodevelopmental symptoms: recommendations for practice. *J Med Genet.* 2019;56(10):701-710. <http://doi.org/10.1136/jmedgenet-2018-105879>
- Kendall KM, Bracher-Smith M, Fitzpatrick H, et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *Br J Psychiatry.* 2019;214(5):297-304. <http://doi.org/10.1192/bjp.2018.301>
- Kirov G, Rees E, Walters JTR, et al. The penetrance of copy number variations for schizophrenia and developmental delay. *Biol Psychiatry.* 2014;75(5):378-385. <http://doi.org/10.1016/j.biopsych.2013.07.022>
- Martin CL, Wain KE, Oetjens MT, et al. Identification of neuropsychiatric copy number variants in a health care system population. *JAMA Psychiatry.* 2020;77(12):1276-1285. <http://doi.org/10.1001/jamapsychiatry.2020.2159>
- Maya I, Perlman S, Shohat M, et al. Should we report 15q11.2 BP1-BP2 deletions and duplications in the prenatal setting? *J Clin Med.* 2020;9(8):11. <http://doi.org/10.3390/jcm9082602>
- Mohan KN, Cao Y, Pham J, et al. Phenotypic association of 15q11.2 CNVs of the region of breakpoints 1-2 (BP1-BP2) in a large cohort of samples referred for genetic diagnosis. *J Hum Genet.* 2019;64(3):253-255. <http://doi.org/10.1038/s10038-018-0543-7>
- Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet Med.* 2013;15(6):478-481. <http://doi.org/10.1038/gim.2012.164>
- Tropeano M, Howley D, Gazzellone MJ, et al. Microduplications at the pseudoautosomal SHOX locus in autism spectrum disorders and related neurodevelopmental conditions. *J Med Genet.* 2016;53(8):536-547. <http://doi.org/10.1136/jmedgenet-2015-103621>
- Vassos E, Collier DA, Holden S, et al. Penetrance for copy number variants associated with schizophrenia. *Hum Mol Genet.* 2010;19(17):3477-3481. <http://doi.org/10.1093/hmg/ddq259>
- Rare chromosome & gene disorder guides and other resources. Understanding Rare Chromosome and Gene Disorder Guides. Accessed September 23, 2023. <https://rarechromo.org/disorder-guides/>
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 5th ed. American Psychiatric Association Publishing; 2013.
- Firth HV, Richards SM, Bevan AP, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am J Hum Genet.* 2009;84(4):524-533. <http://doi.org/10.1016/j.ajhg.2009.03.010>
- Rare coding CNVs from exome sequenced individuals in gnomAD v4. gnomAD browser. Accessed November 3, 2023. <https://gnomad.broadinstitute.org/news/2023-11-v4-copy-number-variants/>
- Castronovo P, Baccarin M, Ricciardello A, et al. Phenotypic spectrum of NRXN1 mono- and bi-allelic deficiency: a systematic review. *Clin Genet.* 2020;97(1):125-137. <http://doi.org/10.1111/cge.13537>
- Cortés-Martín J, Peñuela NL, Sánchez-García JC, Montiel-Troya M, Díaz-Rodríguez L, Rodríguez-Blanque R. Deletion syndrome 22q11.2:

- a systematic review. *Children (Basel)*. 2022;9(8). <http://doi.org/10.3390/children9081168>
22. Rehm HL, Berg JS, Brooks LD, et al. ClinGen—the clinical genome resource. *N Engl J Med*. 2015;372(23):2235-2242. <http://doi.org/10.1056/NEJMsrl406261>
 23. Orphanet. Accessed November 3, 2023. <https://www.orpha.net/consor/cgi-bin/index.php?lng=EN>
 24. GARD. Genetic and Rare Diseases Information Center. Accessed November 3, 2023. <https://rarediseases.info.nih.gov/>
 25. Online Mendelian Inheritance in Man. Accessed November 3, 2023. <https://www.omim.org/>
 26. GeneReviews®. Accessed November 3, 2023. <https://www.ncbi.nlm.nih.gov/books/NBK11116/>
 27. Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. *Syst Rev*. 2012;1:7. <http://doi.org/10.1186/2046-4053-1-7>
 28. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. <http://doi.org/10.1136/bmj.n71>
 29. The Newcastle–Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analysis. The Ottawa Hospital. Accessed May 2, 2021. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
 30. Birnbaum R, Mahjani B, Loos RJF, Sharp AJ. Clinical characterization of copy number variants associated with neurodevelopmental disorders in a large-scale multi-ancestry biobank. *JAMA Psychiatry*. 2022;79(3):250-259. <http://doi.org/10.1001/jamapsychiatry.2021.4080>
 31. Yilmaz F, Null M, Astling D, et al. Genome-wide copy number variations in a large cohort of bantu African children. *BMC Med Genomics*. 2021;14(1):129. <http://doi.org/10.1186/s12920-021-00978-z>
 32. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron*. 2010;68(2):192-195. <http://doi.org/10.1016/j.neuron.2010.10.006>
 33. Tuke M, Tyrrell J, Ruth KS, et al. Large copy-number variants in UK Biobank caused by clonal hematopoiesis may confound penetrance estimates. *Am J Hum Genet*. 2020;107(2):325-329. <http://doi.org/10.1016/j.ajhg.2020.06.001>
 34. Birnbaum R, Heon Shin J, Stoltz J, et al. An investigation of the prevalence of copy number variants associated with neurodevelopmental disorders in a multi-ancestry brain biobank. *Neuropsychopharmacology*. 2022;47:368-369.
 35. Cai M, Que Y, Chen X, et al. 16p13.11 microdeletion/microduplication in fetuses: investigation of associated ultrasound phenotypes, genetic anomalies, and pregnancy outcome follow-up. *BMC Pregnancy Childbirth*. 2022;22(1):913. <http://doi.org/10.1186/s12884-022-05267-w>
 36. Chen X, Jiang Y, Chen R, et al. Clinical efficiency of simultaneous CNV-seq and whole-exome sequencing for testing fetal structural anomalies. *J Transl Med*. 2022;20(1):10. <http://doi.org/10.1186/s12967-021-03202-9>
 37. Demily C, Lesca G, Poisson A, et al. Additive effect of variably penetrant 22q11.2 duplication and pathogenic mutations in autism spectrum disorder: to which extent does the tree hide the forest? *J Autism Dev Disord*. 2018;48(8):2886-2889. <http://doi.org/10.1007/s10803-018-3552-7>
 38. Maya I, Sharony R, Yacobson S, et al. When genotype is not predictive of phenotype: implications for genetic counseling based on 21,594 chromosomal microarray analysis examinations. *Genet Med*. 2018;20(1):128-131. <http://doi.org/10.1038/gim.2017.89>
 39. Bird TD. Charcot-Marie-Tooth hereditary neuropathy overview. In: Adam MP, Everman DB, Mirzaa GM, et al., eds. *GeneReviews*; 1993.
 40. Chrestian N. Hereditary neuropathy with liability to pressure palsies. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews*; 1993.
 41. de Bruijn SE, Fiorentino A, Ottaviani D, et al. Structural variants create new topological-associated domains and ectopic retinal enhancer-gene contact in dominant retinitis pigmentosa. *Am J Hum Genet*. 2020;107(5):802-814. <http://doi.org/10.1016/j.ajhg.2020.09.002>
 42. Edwards SD, Schulze KV, Rosenfeld JA, et al. Clinical characterization of individuals with the distal 1q21.1 microdeletion. *Am J Med Genet A*. 2021;185(5):1388-1398. <http://doi.org/10.1002/ajmg.a.62104>
 43. Albers CA, Newbury-Ecob R, Ouwehand WH, Ghevaert C. New insights into the genetic basis of TAR (thrombocytopenia-absent radii) syndrome. *Curr Opin Genet Dev*. 2013;23(3):316-323. <http://doi.org/10.1016/j.gde.2013.02.015>
 44. Smajlagić D, Lavrichenko K, Berland S, et al. Population prevalence and inheritance pattern of recurrent CNVs associated with neurodevelopmental disorders in 12,252 newborns and their parents. *Eur J Hum Genet*. 2021;29(1):205-215. <http://doi.org/10.1038/s41431-020-00707-7>
 45. Muys J, Jacquemyn Y, Blaumeiser B, et al. Prenatally detected copy number variants in a national cohort: a postnatal follow-up study. *Prenat Diagn*. 2020;40(10):1272-1283. <http://doi.org/10.1002/pd.5751>
 46. Muys J, Janssens K, Vanakker O, et al. Belgian MicroArray prenatal (BEMAPRE) database. *Prenat Diagn*. 2015;35:31-32.