Prevalence of comorbidities in individuals with neurodevelopmental disorders from the aggregated phenomics data of 51,227 pediatric individuals

Dingemans, A.J.M.; Jansen, S.; Reeuwijk, J. van; Leeuw, N. de; Pfundt, R.; Schuurs-Hoeijmakers, J.; ... ; Vries, B.B.A. de

nature medicine

Article

https://doi.org/10.1038/s41591-024-03005-7

Prevalence of comorbidities in individuals with neurodevelopmental disorders from the aggregated phenomics data of 51,227 pediatric individuals

Received: 26 May 2023

Accepted: 16 April 2024

Published online: 14 May 2024

Check for updates

Alexander J. M. Dingemans[1], Sandra Jansen[1], Jeroen van Reeuwijk[1], Nicole de Leeuw[1], Rolph Pfundt[1], Janneke Schuurs-Hoeijmakers[1], Bregje W. van Bon[1], Carlo Marcelis[1], Charlotte W. Ockeloen[1], Marjolein Willemsen[1], Pleuntje J. van der Sluijs[2], Gijs W. E. Santen[2], R. Frank Kooy[3], Anneke T. Vulto-van Silfhout[1], Tjitske Kleefstra[1], David A. Koolen[1], Lisenka E. L. M. Vissers[1] & Bert B. A. de Vries[1]

The prevalence of comorbidities in individuals with neurodevelopmental disorders (NDDs) is not well understood, yet these are important for accurate diagnosis and prognosis in routine care and for characterizing the clinical spectrum of NDD syndromes. We thus developed PhenomAD-NDD, an aggregated database containing the comorbid phenotypic data of 51,227 individuals with NDD, all harmonized into Human Phenotype Ontology (HPO), with in total 3,054 unique HPO terms. We demonstrate that almost all congenital anomalies are more prevalent in the NDD population than in the general population, and the NDD baseline prevalence allows for an approximation of the enrichment of symptoms. For example, such analyses of 33 genetic NDDs show that 32% of enriched phenotypes are currently not reported in the clinical synopsis in the Online Mendelian Inheritance in Man (OMIM). PhenomAD-NDD is open to all via a visualization online tool and allows us to determine the enrichment of symptoms in NDD.

In 1938, Penrose was the first to describe the clinical features of individuals with neurodevelopmental disorders (NDDs) in *A Clinical and Genetic Study of 1280 Cases of Mental Defect*[1]. Although he mainly focused on inheritance patterns and looking at nature versus nurture effects, he also reported congenital malformations in this cohort. Despite almost a century of additional research, much of which has reported associated clinical features in individuals with NDD[2–15], only the prevalence of comorbid mental-health disorders and epilepsy (respectively up to 40% and 26%) have been determined[16,17]. The overall prevalence of most specific phenotypic characteristics eludes us, although these additional phenotypic features often prove essential when taking care of a child with developmental delays or intellectual disability (ID). These comorbid clinical features that assist in making a clinical diagnosis are crucial for adequate variant interpretation and provide information on prognosis and treatment. NDDs affect 2–3% of the population[18], and the absence of information on comorbidities can thus be perceived as more than just a general clinical problem. Following the introduction of next-generation sequencing, which facilitates better genetic diagnoses of individuals with rare disease, and NDDs in particular, there is an ongoing need for curation of more accurate and quantitative phenotypic

[1]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands. [2]Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands. [3]Department of Medical Genetics, University of Antwerp, Antwerp, Belgium. e-mail: Bert.deVries@radboudumc.nl

https://doi.org/10.1038/s41591-024-03005-7

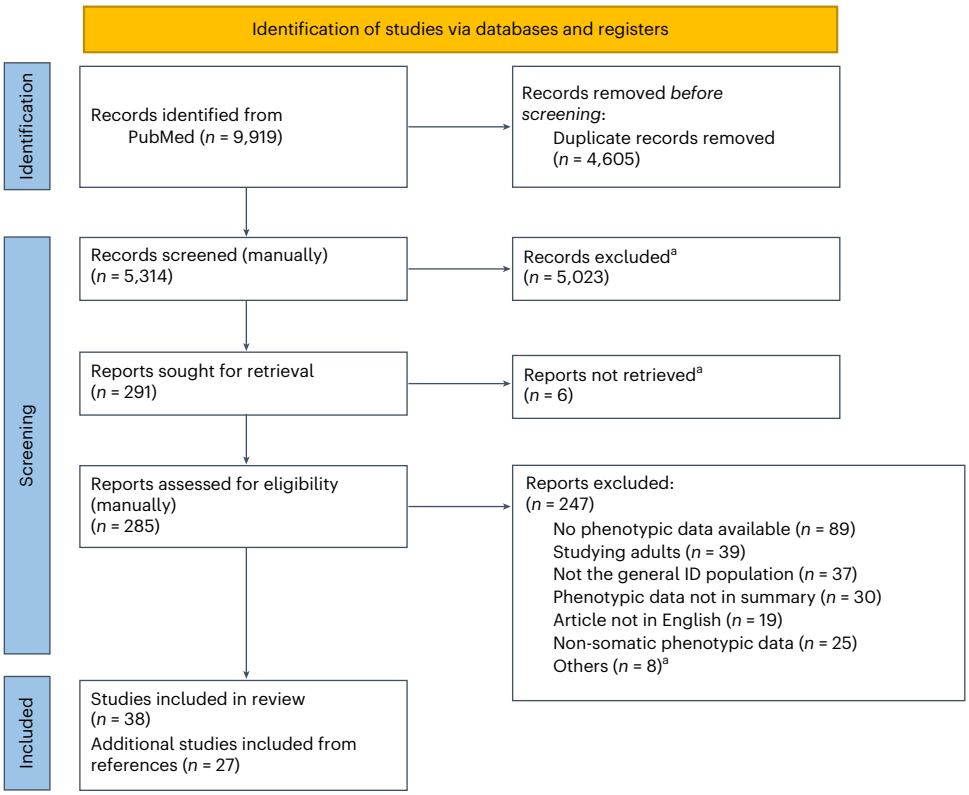


Fig. 1 | PRISMA 2020 flow diagram. The complete search strategy utilized in this study, in compliance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) reporting guideline[38]. [a]The reasons for exclusion are presented in Supplementary Table 1.

data for NDDs. With the rise of next-generation sequencing in the clinic, the 'Genome Aggregation Database (gnomAD) Consortium' has made progress at an unprecedented scale, providing an essential contribution to the field by capturing a wide range of genetic variation in more than 120,000 individuals, and today is used as a rich resource for assessing and prioritizing genetic variation in health and disease[19]. In part, the successes of gnomAD has relied on standardizing and aggregating a dataset of thousands of individuals[20]. Although advances have been made to allow for clinical phenotypes to be captured in standardized terminology, such as Human Phenotype Ontology (HPO)[21,22], phenomics data are mostly only captured at the level of individual patients, with some exceptions such as DECIPHER[23]. This prohibits not only systematic large-scale analyses, for example on the prevalence and existence of comorbidities in a given patient population, but also hampers accurate genotype–phenotype correlations. Without such a cohort prevalence, one cannot be confident that a particular anomaly is part of a specific disorder. In this study we therefore set out to describe the baseline prevalence of comorbid features in pediatric individuals with NDDs by aggregating our in-house clinical data with a systematic review of all available literature related to the pediatric clinical spectrum of NDD. These data and its architecture are freely available in PhenomAD-NDD, an online repository, which will not only be essential to determine what features are enriched for a particular genetic disorder, but will also improve our general understanding of NDDs and their associated comorbidities.

Results

Overview of PhenomAD-NDD

The PhenomAD-NDD repository includes 1,477 individuals with NDDs who were seen at the Radboud University Medical Center. Of these, 969 (66%) were male and 508 (34%) female, with a median age of 8.8 years (interquartile range of 7.2 years). A literature review strategy yielded 9,919 records reporting NDDs and their associated comorbidities, which, after exclusion of duplicates (n = 4,605), left 5,314 for manual screening (Supplementary Table 1). Two hundred and ninety-one reports were deemed possible inclusions, which, after assessment of eligibility, resulted in 38 included studies. Subsequent screening of the references identified another 27 studies for inclusion, bringing the total to 65 studies (Fig. 1). Extracting the phenotypic data and converting these to HPO terms provided data for 49,750 children with NDDs. Collectively, the total number of individuals included in PhenomAD-NDD was 51,227, with 3,054 unique HPO terms and, on average, 2.5 HPO terms per patient (12.0 when including parent-HPO-nodes). The distribution of HPO terms per individual for the individuals seen in Nijmegen showed a right-skewed pattern. This skewness indicates variability in the number of reported congenital anomalies across individuals. Despite the skewness, there was no evidence of substantial clustering within specific individuals that would suggest a non-random distribution of congenital anomalies (Extended Data Fig. 2).

Looking at the contents of PhenomAD-NDD, epilepsy (21.26% [20.55–22.00%]) and behavioral problems (44.38% [43.47–45.30%]) were relatively common (Table 1). Autistic behavior was noted in a fifth (21.34% [20.39–22.33%]) of individuals, and delayed speech (52.02% [49.91–54.19%]) was as prevalent as a delay in the development of motor skills (51.93% [50.35–53.54%]). Some individuals displayed abnormality of the musculoskeletal system (25.77% [25.19–26.36%]), mainly scoliosis (28.54% [26.75–30.42%]) or joint hypermobility (13.5% [11.73–15.55%]). Morphological anomalies of the kidney (2.3% [1.59–3.22%]) and the heart (7.6% [7.09–8.13%]) were relatively rare. All individuals had a neurodevelopmental abnormality, with the vast majority (91.89% [90.83–92.96%]) diagnosed with ID. For 25.8% of these individuals, the severity of the ID was known, but over a third (44.81% [43.30–46.37%]) of this subgroup were diagnosed with mild ID, 24.99% [23.87–26.16%] had moderate ID, 19.62% [18.62–20.65%] had severe ID and 10.57% [9.85–11.34%] had profound ID (subgrouping based on the corresponding HPO terms for the severity of ID). Less than a quarter of our study population reported immunological issues

https://doi.org/10.1038/s41591-024-03005-7

Table 1 | Prevalence of some standard clinical features in the 51,227 individuals included in PhenomAD-NDD

Category	Clinical feature	PhenomAD-NDD	95% CI PhenomAD-NDD
Neurological	Neurodevelopmental abnormality	51,227/51,227 (100.0%)	99.14–100.87
	Motor delay	4,113/7,921 (51.93%)	50.35–53.54
	Delayed speech and language development	2,296/4,414 (52.02%)	49.91–54.19
	Intellectual disability	28,648/31,177 (91.89%)	90.83–92.96
	Intellectual disability severity specified	7,386/28,648 (25.8%)	–
	Intellectual disability, mild	3,310/7,386 (44.81%)	43.30–46.37
	Intellectual disability, moderate	1,846/7,386 (24.99%)	23.87–26.16
	Intellectual disability, profound	781/7,386 (10.57%)	9.85–11.34
	Intellectual disability, severe	1,449/7,386 (19.62%)	18.62–20.65
	Autistic behavior	1,887/8,842 (21.34%)	20.39–22.33
	Behavioral abnormality	9,096/20,498 (44.38%)	43.47–45.30
	Seizure	3,334/15,679 (21.26%)	20.55–22.00
Head circumference	Macrocephaly	260/9,031 (2.88%)	2.54–3.25
	Microcephaly	1,362/5,946 (22.91%)	21.71–24.16
Facial abnormalities	Abnormality of the ear	1,629/8,516 (19.13%)	18.21–20.08
	Abnormality of the eye	4,169/11,559 (36.07%)	34.98–37.18
	Abnormality of the nose	746/1,792 (41.63%)	38.70–44.73
	Abnormality of the mouth	1,561/3,333 (46.83%)	44.54–49.22
	Cleft lip	10/1,546 (0.65%)	0.31–1.19
	Cleft palate	31/1,546 (2.01%)	1.36–2.85
Cardiac	Abnormal heart morphology	828/10,896 (7.6%)	7.09–8.13
	Atrial septal defect	58/1,723 (3.37%)	2.56–4.35
	Ventricular septal defect	44/1,773 (2.48%)	1.80–3.33
Gastrointestinal	Abnormality of the gastrointestinal tract	472/13,449 (3.51%)	3.20–3.84
	Abdominal wall defect	62/1,738 (3.57%)	2.74–4.57
	Constipation	516/2,770 (18.63%)	17.06–20.31
Urogenital	Abnormal renal morphology	34/1,477 (2.3%)	1.59–3.22
	Abnormality of the genitourinary system	1,211/19,367 (6.25%)	5.91–6.62
	Hypospadias	27/1,527 (1.77%)	1.17–2.57
Skeleton	Joint hypermobility	200/1,477 (13.5%)	11.73–15.55
	Abnormality of the musculoskeletal system	7,424/28,810 (25.77%)	25.19–26.36
	Kyphosis	16/1,527 (1.05%)	0.60–1.70
	Scoliosis	944/3,308 (28.54%)	26.75–30.42
	Craniosynostosis	8/1,477 (0.5%)	0.23–1.07
Skin/hair/nails	Abnormality of the integument	1,327/5,103 (26.0%)	24.62–27.44
Visual and hearing impairments	Abnormality of vision	1,106/7,052 (15.68%)	14.77–16.64
	Hearing abnormality	506/7,186 (7.04%)	6.44–7.68
Immunological	Abnormality of the immune system	1,435/11,393 (12.6%)	11.95–13.26
Endocrinological	Abnormality of the endocrine system	117/3,493 (3.35%)	2.77–4.01
Neoplasia	Neoplasm	258/8,790 (2.94%)	2.59–3.32

In the PhenomAD-NDD column, the first number indicates the number of individuals for whom that symptom was positive: the number after the solidus (/) indicates the total number of individuals for whom that symptom was assessed.

(12.6% [11.95–13.26%]), which were more common than endocrinological problems (3.35% [2.77–4.01%]). A complete overview of all clinical features in PhenomAD-NDD is available online and in Supplementary Information (Fig. 3).

Subgroup analysis

When investigating several subgroups in our clinical cohort of 1,477 individuals for differences between sexes, behavioral abnormalities (69% versus 55%, $P < 0.00001$), autistic behavior (39% versus 21%, $P < 0.00001$) and urogenital problems (22% versus 10%, $P < 0.00001$) were all more common in boys than in girls. However, most clinical features, such as ID and its severity, were not statistically significantly different between the sexes (Fig. 4). When looking at the phenotype of individuals with different severity of ID, seizures were more common in the severe/profound ID group than in the mild/moderate ID group (48% versus 14%, $P < 0.00001$). However, behavioral abnormalities

https://doi.org/10.1038/s41591-024-03005-7

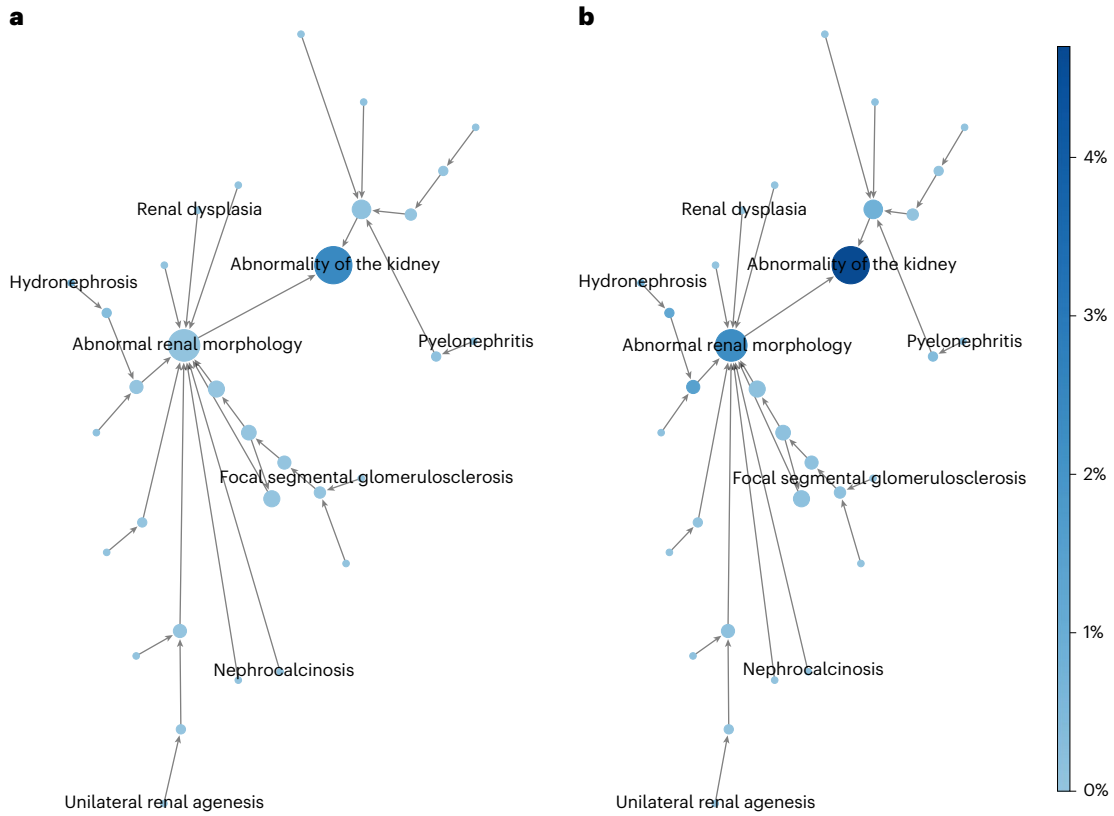


Fig. 2 | Prevalence of kidney-related HPO terms in PhenomAD-NDD. **a**,**b**, Prevalence of specific symptoms and the corresponding HPO terms (**a**), and an example where the HPO terms are counted as positive if one of the child nodes is positive (**b**), to minimize interobserver variability. For example, the difference when looking at 'abnormal renal morphology' is quite substantial (0.1% versus 2.3%).

were more present in the mild/moderate ID group than in the severe/profound ID group (78% versus 58%, $P = 0.0004$).

Comparison to EUROCAT

To determine whether congenital anomalies are more present in the NDD population than in the general population, the European network of population-based registries for the epidemiological surveillance of congenital anomalies (EUROCAT) was consulted. This network collects prevalence information on congenital anomalies in the general population[24–26] and is therefore suitable for use as a control group. In total, EUROCAT data for 11,616,332 individuals, collected from 37 institutions in 18 countries, were included in these analyses (Extended Data Fig. 1). These provide an extensive overview of the baseline prevalence of congenital anomalies in all births. For example, congenital heart defects are present in 0.7% of births, with ventricular septal defects accounting for roughly half, whereas the prevalence of these in individuals with NDDs is ten times higher (7.6%). Hypospadias are present in 0.2% of individuals in EUROCAT, whereas individuals with NDDs also report this symptom nine times more often (1.8%). For abdominal wall defects, the difference between EUROCAT and NDDs is almost 100-fold (0.04% versus 3.6%, respectively).

All congenital anomalies are present in fewer than 1% of individuals in EUROCAT, and almost all investigated congenital anomalies are present more frequently in our NDD cohort when compared with their prevalence in EUROCAT (Table 2). Some abnormalities, such as congenital glaucoma and Hirschsprung's disease, were not reported at all in PhenomAD-NDD. For the 40 congenital anomalies that were reported in both PhenomAD-NDD and EUROCAT, all were more prevalent in PhenomAD-NDD, with 30 reaching statistical significance (75%) after Bonferroni correction. Furthermore, although PhenomAD-NDD exhibited a higher frequency of almost all congenital anomalies when compared to EUROCAT, it is important to acknowledge the potential ascertainment bias inherent to our findings. Certain anomalies may be less likely to come to clinical attention unless a more thorough clinical evaluation is performed, which is often triggered by the neurodevelopmental condition.

Enrichment analysis

To detect the phenotypic features that are predictive of a particular genetic syndrome and to demonstrate the additional power of PhenomAD-NDD, the enrichment of clinical features was calculated and compared to the HPO terms in the clinical synopsis of the Online Mendelian Inheritance in Man (OMIM) compendium for 33 NDD-related syndromes[27]. For these clinical features, high-quality data are also available in the Human Disease Genes websites for individuals with a clinical and molecularly proven (American College of Medical Genetics and Genomics likely pathogenic or pathogenic variants) diagnosis of these 33 disorders (Supplementary Table 2). When looking at all enriched symptoms in these 33 NDD-related syndromes with a minimal prevalence of 1%, 1,541 features were enriched according to PhenomAD-NDD: 1,041 (68%) of those were recorded in the clinical synopsis, indicating that 32% of relevant clinical features according to PhenomAD-NDD are currently missing in OMIM. This varies considerably per genetic syndrome, from 16% (MRD50, OMIM #617787) to 100% (NEDHELS, #617171, Extended Data Fig. 3) of concordance of the enriched features with the clinical synopsis. Furthermore, the symptoms that are not present in OMIM were on average enriched by a factor of 8.2 in the genetic NDD-related syndromes compared to PhenomAD-NDD.

We next investigated whether 3,350 clinical features described in both OMIM and the Human Disease Genes dataset for these 33 NDD-related syndromes are in fact enriched, and thus predictive

https://doi.org/10.1038/s41591-024-03005-7

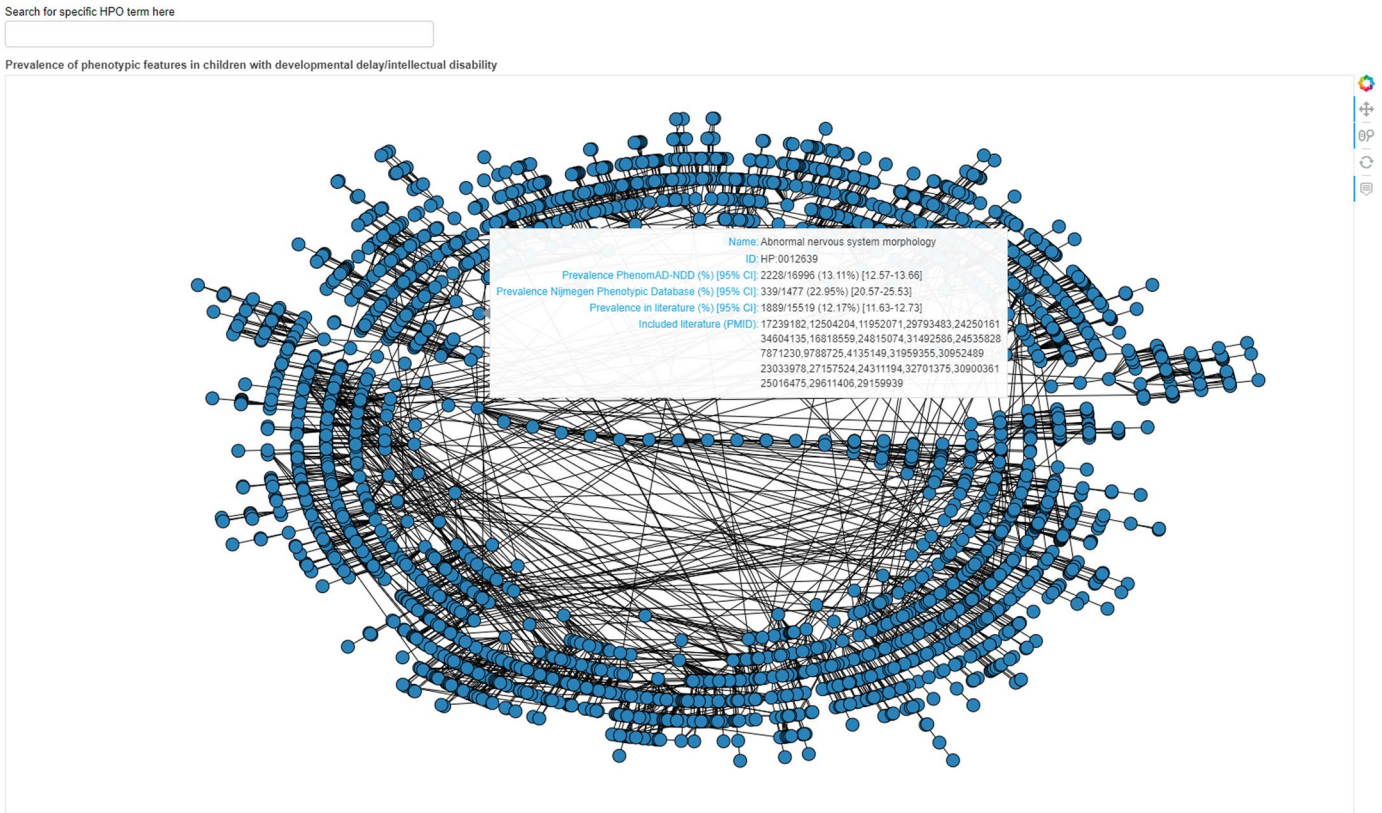


Fig. 3 | Visualization of the data in PhenomAD-NDD. A screenshot of one of the two produced tools. This one was developed with the Bokeh[39] library, which can be used to view the prevalence of any clinical feature in PhenomAD-NDD. A specific symptom can be looked up, and then the prevalence of that specific HPO term is given, including any child nodes as well. The tool is available as a supplement to this Article and on the Human Disease Genes website series (https://humandiseasegenes.nl/phenomadndd/)[37], and can be used as the definition of the baseline somatic comorbidity in pediatric individuals with NDDs.

features for the syndrome. The average enrichment for all these terms is 3.2, indicating that these clinical features are on average three times more present in these syndromes than in the general NDD population. However, 628 (19%) have an enrichment lower than 1.0, indicating that these features might not be enriched in the genetic syndrome. Of these, 59 (2%) reach statistical significance after Bonferroni correction, indicating that these are more prevalent in the general NDD population than in the genetic syndrome. An example of such a clinical feature is feeding difficulties for the disorder associated with heterozygous pathogenic variants in *DEAF1* (Vulto-van Silfhout-de Vries syndrome, OMIM #615828 (ref. 28). This syndrome is characterized by delayed development, behavioral abnormalities, hypotonia and seizures, with a wide range of symptoms described. OMIM indicates feeding difficulties as part of the phenotype, but the prevalence of feeding difficulties in the syndrome is only 27%, as opposed to 57% in PhenomAD-NDD (relative risk, 0.47, $P < 0.001$).

Discussion

Investigations of NDD comorbidities started almost 100 years ago[1–9]. Now, PhenomAD-NDD provides an aggregated detailed phenotypic description of an NDD cohort of considerable size and quality, containing data for 51,227 individuals with NDDs. Several other databases that collect phenotypic data do exist, such as LOVD[29], PhenoDis[30], GPCards[31], DECIPHER[23] and the Monarch Initiative[32]. However, only a limited number of these initiatives focus on NDDs specifically. A nice example of the latter is DECIPHER, which over the years has collected the data of more than 30,000 individuals with developmental disorders[23]. The collection of PhenomAD-NDD is, however, agnostic to a genetic diagnosis, and provides both the numerator and denominator for all features assessed. This difference is crucial to not only capture the prevalence of features in the broader NDD context, but also to allow us to determine the relevance of features when describing the phenotypic spectrum of syndromes and thus facilitate better insights into prognoses.

In PhenomAD-NDD, there are currently data with on average 2.5 HPO terms per individual (12.0 HPO terms including parent-HPO-nodes). These data, available in detail to all, can be used as a benchmark for future disease–gene discovery studies. By comparing the prevalence of a clinical feature to the prevalence in our cohort, one can determine whether a specific symptom is enriched for that particular patient group. For example, we show that epilepsy is expected in 21.3% of the NDD population, and autistic behavior in 21.3%—two examples of a clinical feature commonly described as a feature of a particular genetic syndrome. We show that if individuals with a pathogenic variant in a gene report epilepsy or autistic behavior at a level of ~20–25%, these numbers are the norm in the general NDD population and therefore should not be noted as a distinctive feature of that disorder. The determination of predictive clinical features is essential for the clinician when considering possible syndromic diagnoses in individuals with NDDs. The aggregation of multiple data sources sometimes leads to difficulties. Scoliosis, for instance, was present in 3.8% of the individuals seen in our outpatient clinic, but in almost 48.5% of individuals in the literature for whom scoliosis was reported, indicating a possible reporting bias in the included papers, and therefore elevating the prevalence in PhenomAD-NDD to 28.5%. Another limitation of using data from the literature is that only the summarized data were available. This constrains the depth of our analysis and interpretations; with data from the literature we cannot investigate the correlation of symptoms within patients and cannot correct for age and/or sex, whereas our regression analysis demonstrates that these could impact the results for a small number of syndromes. One should therefore always have

https://doi.org/10.1038/s41591-024-03005-7

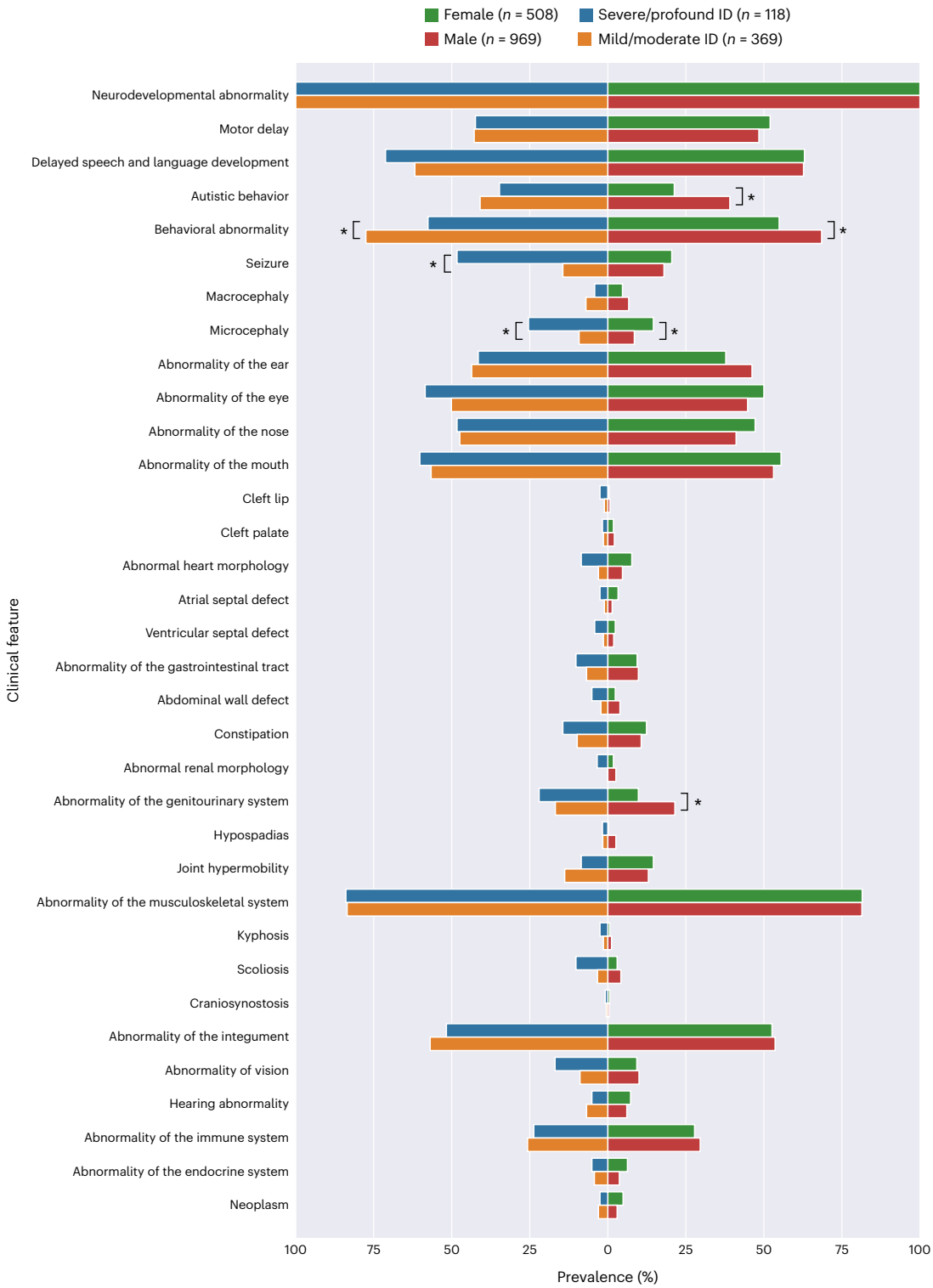


Fig. 4 | Subgroup analyses. Two subgroups were compared within the cohort, based on sex and the severity of ID when reported, and statistical significance was assessed using a two-sided χ^2 test with Bonferroni correction for multiple testing. Autistic behavior and moderate ID are more common in males, and congenital heart defects are more prevalent in females. Individuals with more severe ID report more microcephaly, seizures and congenital heart defects. *Statistically significant at the Bonferroni-corrected level.

these possible confounders in mind when performing enrichment analyses for genetic disorders.

However, the aggregation of all these data provides large benefits as well. In recent years, large-scale datasets of human sequence data have become freely accessible online, first with ExAC and more recently with gnomAD[19,33]. These genomic data assist us in the medical interpretation of genetic variants by illustrating healthy human genetic variation. Furthermore, these data are used to calculate the intolerance to variants in a specific gene (like the pLI score[34]), and have become essential when discovering new disease genes. A gene suspected to be associated with a newly identified genetic syndrome should not have similar (or even identical) diplotypes in the healthy population as in

https://doi.org/10.1038/s41591-024-03005-7

Table 2 | Prevalence of congenital anomalies in our NDD cohort compared with the prevalence of these anomalies in the EUROCAT registry

Anomaly	HPO ID	Prevalence PhenomAD-NDD (%)	Prevalence EUROCAT (%)	*P* value
Abdominal wall defect	HP:0010866	3.567	0.037	<0.001
Anencephalus and similar	HP:0002323	0.000	0.003	1
Annular pancreas	HP:0001734	0.000	0.002	1
Anophthalmos	HP:0000528	0.062	0.001	0.095
Anophthalmos/micropthalmos	HP:0000568; HP:0000528	0.406	0.008	<0.001
Ano-rectal atresia and stenosis	HP:0002023; HP:0002025	0.034	0.026	1
Anotia	HP:0009892	0.000	0.008	1
Aortic atresia/interrupted aortic arch	HP:0012304; HP:0011611	0.034	0.004	1
Aortic valve atresia/stenosis	HP:0010883; HP:0001650	0.169	0.013	<0.001
Arhinencephaly/holoprosencephaly	HP:0001360	0.068	0.003	1
Atresia of bile ducts	HP:0005912	0.000	0.003	1
Atresia or stenosis of other parts of small intestine	HP:0012848; HP:0012739	0.000	0.008	1
Atrial septal defect	HP:0001631	3.366	0.186	<0.001
Atrioventricular septal defect	HP:0006695	0.406	0.032	<0.001
Bilateral renal agenesis including Potter syndrome	HP:0010958	0.000	0.003	1
Bladder exstrophy and/or epispadia	HP:0002836	0.000	0.005	1
Choanal atresia	HP:0000453	0.000	0.008	1
Cleft lip	HP:0000202	1.906	0.079	<0.001
Cleft palate	HP:0000175	2.005	0.056	<0.001
Clubfoot (talipes equinovarus)	HP:0001762	1.422	0.056	<0.001
Coarctation of aorta	HP:0001680	0.339	0.036	<0.001
Common arterial truncus	HP:0001660	0.135	0.005	<0.001
Congenital cataract	HP:0000519	0.135	0.013	0.191
Congenital constriction bands/amniotic band	HP:0009775	0.000	0.002	1
Congenital glaucoma	HP:0008007; HP:0008041	0.000	0.003	1
Congenital heart defects	HP:0001627	7.599	0.710	<0.001
Congenital hydronephrosis	HP:0000126	0.948	0.140	<0.001
Craniosynostosis	HP:0001363	0.542	0.024	<0.001
Cystic adenomatous malfunction of lung	HP:0010959	0.000	0.008	1
Diaphragmatic hernia	HP:0000776	0.135	0.021	1
Double outlet right ventricle	HP:0001719	0.000	0.010	1
Duodenal atresia or stenosis	HP:0002247; HP:0100867	0.000	0.012	1
Ebstein's anomaly	HP:0010316	0.000	0.004	1
Encephalocele	HP:0002084	0.000	0.004	1
Gastroschisis	HP:0001543	0.000	0.023	1
Hip dislocation and/or dysplasia	HP:0002827; HP:0001385	1.557	0.063	<0.001
Hirschsprung's disease	HP:0002251	0.064	0.014	1
Hydrocephalus	HP:0000238	2.145	0.031	<0.001
Hypoplastic left heart	HP:0004383	0.068	0.015	1
Hypoplastic right heart	HP:0010954	0.000	0.003	1
Hypospadias	HP:0000047	1.768	0.175	<0.001
Indeterminate sex	HP:0000062	0.068	0.005	1
Limb reduction defects	HP:0009815	12.796	0.038	<0.001
Microcephaly	HP:0000252	22.906	0.024	<0.001
Mitral valve anomalies	HP:0001633	0.135	0.004	<0.001
Multicystic renal dysplasia	HP:0000003	0.000	0.029	1
Neural tube defects	HP:0410043	0.475	0.026	<0.001

https://doi.org/10.1038/s41591-024-03005-7

Table 2 (continued) | Prevalence of congenital anomalies in our NDD cohort compared with the prevalence of these anomalies in the EUROCAT registry

Anomaly	HPO ID	Prevalence PhenomAD-NDD (%)	Prevalence EUROCAT (%)	*P* value
Oesophageal atresia with or without tracheo-oesophageal fistula	HP:0002032	0.068	0.023	1
Omphalocele	HP:0001539	0.203	0.013	<0.001
Oro-facial clefts	HP:0002006	7.893	0.135	<0.001
PDA as only CHD in term infants (≥37 weeks)	HP:0001643	1.151	0.035	<0.001
Polydactyly	HP:0010442	0.406	0.087	0.014
Posterior urethral valve and/or prune belly	HP:0010957; HP:0004392	0.000	0.009	1
Pulmonary valve atresia	HP:0010882	0.000	0.009	1
Pulmonary valve stenosis	HP:0001642	0.609	0.042	<0.001
Single ventricle	HP:0001750	0.000	0.004	1
Situs inversus	HP:0011616; HP:0011538; HP:0003363; HP:0001696; HP:0031592	0.000	0.005	1
Skeletal dysplasias	HP:0002652	0.000	0.005	1
Spina bifida	HP:0002414	0.475	0.019	<0.001
Syndactyly	HP:0001159	6.892	0.044	<0.001
Tetralogy of Fallot	HP:0001636	0.393	0.029	<0.001
Total anomalous pulmonary venous return	HP:0005160	0.000	0.006	1
Transposition of great vessels	HP:0001669	0.339	0.031	<0.001
Tricuspid atresia and stenosis	HP:0011662; HP:0010446	0.000	0.005	1
Ventricular septal defect	HP:0001629	2.482	0.338	<0.001

Reported *P* values were calculated using two-sided χ^2 tests and are adjusted for Bonferroni correction.
PDA, patent ductus arteriosus; CHD, congenital heart disease.

affected individuals[35,36]. PhenomAD-NDD may as such be considered an equivalent of gnomAD for the interpretation of individuals at the phenotypic level, in the broader context of NDDs. An ultimate synergy may be obtained when both phenotypic and genotypic data become available at this unprecedented scale, with more initiatives developed to collect the data (for example, the European '1+ Million Genomes' Initiative). Further incorporating these different datasets and adding additional modalities of patient data to PhenomAD-NDD, as well as investigating further opportunities to unify worldwide existing initiatives on the systematic collection of phenotype–genotype information, could further increase the utility of the database. However, because these additional modalities are currently not systematically captured together in a uniform manner in clinical practice, this would require a different manner of capturing the (phenotypic) data.

We utilized the data collected in this study in several analyses. First, we proved that the data enable the determination of enriched clinical features for 33 genetic NDD-related syndromes. We demonstrated that PhenomAD-NDD is able to recover the phenotypic features that are known in the literature remarkably well. Furthermore, because enrichment analysis is a quantitative analysis, it is possible to calculate the relative prevalence of the clinical features compared with the general NDD population in PhenomAD-NDD. Comparing these with the current standard in clinical practice, OMIM, shows that the investigated clinical synopses are currently missing 32% of statistically significant enriched clinical features. Furthermore, we demonstrated that these symptoms are, on average, enriched by a factor of 8.2, indicating the clinical significance of these missing phenotypic features. The described symptoms in OMIM are generally correct, with only 2% of the clinical features described more prevalent in the general NDD population than in the genetic syndrome according to PhenomAD-NDD. This indicates that only a small proportion of the clinical features described in OMIM exhibit a higher prevalence in PhenomAD-NDD than in the genetic syndrome population. Examining the concordance of the enriched features according to our enrichment analyses and the clinical synopses of OMIM, we see that there is a substantial variation between genetic syndromes, as some, such as NEDHELS (OMIM #617171), have much more and more detailed phenotypic information than others (such as MRD50, OMIM #617787) described in OMIM. This is probably due to the nature of OMIM, which relies on manual curation of the medical literature, and some pages will thus be more up to date than others.

These analyses demonstrate the actual predicting clinical features for these genetic syndromes by quantifying their differences in prevalence. This enables the prediction of clinical features for genetic syndromes such that healthcare professionals can know which symptoms are characteristic of a specific disorder when seeing an individual with an NDD. Although we show the results for 33 genetic NDD-related disorders, anyone can establish the enriched clinical features for any genetic disorder of interest using the developed tool (provided online or in Supplementary Information). It is important to note here that the data in OMIM are still relevant. If one is wondering whether a clinical feature is part of a genetic syndrome (that is, the prevalence is elevated compared to that in the general population), the clinical synopsis of OMIM provides an excellent overview. However, if one wishes to compare the prevalence of a specific clinical feature to that in the general NDD population, for example for variant interpretation or to describe the characteristic features in a syndrome, our data are better suited to provide answers.

Next to these analyses, we confirmed our clinical hypothesis that congenital anomalies are more present in pediatric individuals with NDDs than in the general population. Not surprisingly, almost every investigated anomaly is significantly more present in individuals with NDDs, confirming that clinicians should be monitoring the development of children with a congenital anomaly closely.

The subgroup, EUROCAT and enrichment analyses are only some of the numerous examples of what is possible with a dataset of this size

and quality. To further understand rare diseases, global open science and data-sharing are essential. Accordingly, PhenomAD-NDD is freely available online (Fig. 3) on the Human Disease Genes website series (https://humandiseasegenes.nl/phenomadndd/)[37]. The latter also includes a tool allowing clinicians and researchers to calculate the enrichment of clinical features for syndromes of interest.

In conclusion, pediatric individuals with NDDs report problems in every organ system, and almost all congenital anomalies are more prevalent in these individuals than in the general population. The exact prevalence of each HPO term described in our cohort of 51,227 individuals can be found in PhenomAD-NDD. Clinicians and researchers can use these data and a developed online tool to determine whether a clinical feature is truly enriched when describing a genetic NDD-related syndrome. We have utilized these data to demonstrate that 32% of enriched clinical features in 33 NDD syndromes are currently missing in the clinical synopses of OMIM.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-024-03005-7.

References

1. Penrose, L. S. et al. *A Clinical and Genetic Study of 1280 Cases of Mental Defect* (HMSO, 1938).
2. Sarason, S. B. *Psychological Problems in Mental Deficiency* (Harper, 1949).
3. Yannet, H. Diagnostic classification of patients with mental deficiency: distribution of 1,330 institutionalized patients, with review of incidence of convulsive disorders and noncerebral developmental anomalies. *Am. J. Dis. Child.* **70**, 83–88 (1945).
4. Kallmann, F. J. *Heredity in Health and Mental Disorder; Principles of Psychiatric Genetics in the Light of Comparative Twin Studies* (Norton, 1953).
5. Pitt, D. & Roboz, P. A survey of 782 cases of mental deficiency. *J. Ment. Defic. Res.* **9**, 4–23 (1965).
6. Gustavson, K. H., Holmgren, G., Jonsell, R. & Son Blomquist, H. K. Severe mental retardation in children in a northern Swedish county. *J. Ment. Defic. Res.* **21**, 161–180 (1977).
7. Laxova, R., Ridler, M. A. & Bowen-Bravery, M. An etiological survey of the severely retarded Hertfordshire children who were born between January 1, 1965 and December 31, 1967. *Am. J. Med. Genet.* **1**, 75–86 (1977).
8. Göstason, R. Psychiatric illness among the mentally retarded. A Swedish population study. *Acta Psychiatr. Scand. Suppl.* **318**, 1–117 (1985).
9. Reid, A. H. Schizophrenia in mental retardation: clinical features. *Res. Dev. Disabil.* **10**, 241–249 (1989).
10. Deb, S. & Joyce, J. Psychiatric illness and behavioural problems in adults with learning disability and epilepsy. *Behav. Neurol.* **11**, 125–129 (1999).
11. Holland, T. & Murphy, G. Behavioural and psychiatric disorder in adults with mild learning difficulties. *Int. Rev. Psychiatry* **2**, 117–135 (1990).
12. Mirabi, M. in *Child and Adolescent Psychiatry, Mental Retardation and Geriatric Psychiatry* (eds Pichot, P. et al.) 177–181 (Springer, 1985).
13. Corbett, J. A. in *Scientific Studies in Mental Retardation* (eds Dobbing, J. et al.) 479–506 (Palgrave Macmillan, 1984).
14. Linna, S. L. et al. Psychiatric symptoms in children with intellectual disability. *Eur. Child Adolesc. Psychiatry* **8**, 77–82 (1999).
15. Farag, T. I. et al. Disease profile of 400 institutionalized mentally retarded patients in Kuwait. *Clin. Genet.* **44**, 329–334 (1993).
16. Bratek, A., Krysta, K. & Kucia, K. Psychiatric comorbidity in older adults with intellectual disability. *Psychiatr. Danub.* **29**, 590–593 (2017).
17. McGrother, C. W. et al. Epilepsy in adults with intellectual disabilities: prevalence, associations and service implications. *Seizure* **15**, 376–386 (2006).
18. Daily, D. K., Ardinger, H. H. & Holmes, G. E. Identification and evaluation of mental retardation. *Am. Fam. Physician* **61**, 1059–1067 (2000).
19. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
20. Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X. & Greene, C. S. Responsible, practical genomic data sharing that accelerates research. *Nat. Rev. Genet.* **21**, 615–629 (2020).
21. Robinson, P. N. et al. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* **83**, 610–615 (2008).
22. Köhler, S. et al. The human phenotype ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).
23. Firth, H. V. et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
24. Boyd, P. A. et al. Paper 1: the EUROCAT network—organization and processes. *Birth Defects Res. A* **91**, S2–S15 (2011).
25. Dolk, H. EUROCAT: 25 years of European surveillance of congenital anomalies. *Arch. Dis. Child. Fetal Neonatal Ed.* **90**, F355–F358 (2005).
26. Greenlees, R. et al. Paper 6: EUROCAT member registries: organization and activities. *Birth Defects Res. A* **91**, S51–S100 (2011).
27. McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
28. Nabais Sá, M. J. et al. De novo and biallelic DEAF1 variants cause a phenotypic spectrum. *Genet. Med.* **21**, 2059–2069 (2019).
29. Fokkema, I. F. A. C. et al. LOVD v.2.0: the next generation in gene variant databases. *Hum. Mutat.* **32**, 557–563 (2011).
30. Adler, A. et al. PhenoDis: a comprehensive database for phenotypic characterization of rare cardiac diseases. *Orphanet J. Rare Dis.* **13**, 22 (2018).
31. Li, B. et al. GPCards: an integrated database of genotype-phenotype correlations in human genetic diseases. *Comput. Struct. Biotechnol. J.* **19**, 1603–1611 (2021).
32. Mungall, C. J. et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.* **45**, D712–D722 (2017).
33. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
34. Fuller, Z. L., Berg, J. J., Mostafavi, H., Sella, G. & Przeworski, M. Measuring intolerance to mutation in human genetics. *Nat. Genet.* **51**, 772–776 (2019).
35. Lelieveld, S. H. et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).
36. Kaplanis, J. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020).
37. Dingemans, A. J. M. et al. Human disease genes website series: an international, open and dynamic library for up-to-date clinical information. *Am. J. Med. Genet. A* **185**, 1039–1046 (2021).
38. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int. J. Surg.* **88**, 105906 (2021).
39. Bokeh Development Team. *Bokeh: Python Library for Interactive Visualization* https://bokeh.pydata.org/en/latest/ (2021).

Methods

Nijmegen Phenotypic Database

Within the Radboud Biobanks 'Intellectual Disability' and 'Genetics and Rare Disease'[40], phenotypic and molecular data were prospectively and systematically captured for individuals with NDDs who were referred to the Radboud University Medical Center between 2009 and 2021. For all individuals under the age of 18 years for whom consent was available, the phenotypes were manually curated and converted into HPO terms[21]. Sex was determined by self-reporting. The medical ethical review board of the Radboud University Medical Center approved the use of data for the purpose of this study (#2018-4733).

Review of literature

We searched PubMed in February 2021 to find previous studies looking at the associated clinical features of pediatric individuals with NDDs. The following search terms were used: mental retardation comorbidity; intellectual disability comorbidity; developmental delay comorbidity; 'intellectual disability' AND 'congenital anomalies'; 'mental retardation' AND 'congenital anomalies'; 'developmental delay' AND 'congenital anomalies'. The titles of the studies resulting from this search were screened for English language, content and age of the included individuals, after which we made a selection. We read the abstracts of these selected studies to assess eligibility, resulting in the inclusion of studies describing the associated somatic comorbidity of a cohort of individuals aged <18 years with an NDD (Fig. 1 and Supplementary Table 1).

Because we were interested in the phenotype of the general NDD population, studies focusing on a specific genetic syndrome were excluded. Studies with a primary focus on psychiatry were also excluded, although studies where psychiatry was described as a comorbidity of an NDD were included. We checked the references of all identified articles for possible further inclusions. If this led to additional inclusions, the references of these papers were reviewed as well, ensuring the most complete dataset achievable. For every included study, which clinical features it reported were checked to obtain a realistic denominator for each. For example, if a study reported on ophthalmological symptoms in individuals with NDDs, the study was included and used to determine the prevalence of ophthalmological symptoms, but not when calculating the prevalence of, for instance, congenital heart disease, as this was not assessed.

To provide a homogeneous dataset, the clinical data of all papers selected for inclusion were converted to HPO terms. First, all data were extracted from the tables in the papers that described the phenotypic data (Supplementary Table 1, collected_data sheet). Then, one clinician (to ensure uniformity and reliability for the whole dataset) converted these clinical features one by one into HPO terms: if there was not a direct match for a specific HPO term, usually the best-matching term based on expert opinion was chosen. The terms that were chosen for every clinical feature are also described in Supplementary Table 1. These data were then used to create PhenomAD-NDD: the aggregated dataset of the prevalence of all clinical features in individuals with NDDs. For the assessment of positively scored HPO terms in all of PhenomAD-NDD, parental ancestor terms were included to complete the clinical information and to minimize interobserver variability (Fig. 2).

In the process of synthesizing data for our meta-analysis, a critical challenge was the inherent limitation of working with summarized data from each study, as we did not have access to individual-level data. Specifically, we were able to discern the prevalence of individual clinical features (represented as HPO terms) in a given study cohort, but the overlap or co-occurrence of these features within individual participants remained elusive. To address this, while aggregating data, each clinical feature was added as a distinct HPO term. However, given that HPO terms can be nested, with specific symptoms falling under broader categories, there was an inherent risk of inadvertently inflating the prevalence figures. To mitigate this, we included the safeguard that the combined prevalence of an HPO term and its related parent or ancestor terms could never exceed the total patient count from the respective study. Additional mitigations in the synthesis included the following: (1) if a study provided only percentages, the number of patients that had that specific term was calculated using the total number of individuals, and was rounded if there was ambiguity; (2) in cases where a given feature corresponded to multiple HPO terms (for instance, when endocrine and metabolic features were described as one), that feature was excluded.

Subgroup analysis

Subgroup analysis to determine whether sex (male versus female) and severity of ID (mild/moderate versus severe/profound) confounded the phenotype was limited to individuals from the Nijmegen Phenotypic Database, as this information was not retrievable from the literature-included studies. The prevalences of symptoms between subgroups were compared, and statistical significance was assessed using a two-sided χ^2 test with Bonferroni correction for multiple testing, leading to thresholds for significance of $P < 0.05/39$ (the number of features investigated) = 0.00128 and $P < 0.05/35$ (the number of features investigated in this analysis) = 0.00143 for sex difference and severity of ID, respectively.

Comparison to EUROCAT

We consulted EUROCAT for the baseline prevalence of congenital anomaly subgroups in the general population[24–26]. We included all congenital anomalies reported in live births between 2000 and 2016. Only registries that reported data on all congenital anomalies (referred to as 'full registries' by EUROCAT) were eligible. Registries in which only a subset were evaluated were excluded from analysis. Two-sided χ^2 tests with Bonferroni correction for multiple testing were used to assess statistical significance, yielding $P < 0.05/65 = 0.000769$ (65 was chosen as the denominator as that is the number of congenital anomalies investigated).

Enrichment analysis

To demonstrate the power of PhenomAD-NDD, a proof-of-concept study was performed assessing the clinically enriched features for 33 NDD-related syndromes. Enriched clinical features were defined as symptoms that are significantly more prevalent in a specific genetic disorder than in the general NDD population. The data for these syndromes were collected using the Human Disease Genes website series[37], an initiative to collect phenotypic information in a standardized manner using HPO terms for patients with a molecularly (likely) pathogenic variant (according to the American College of Medical Genetics and Genomics guidelines[41]) in the gene of interest, and then compared to the prevalence in PhenomAD-NDD. A two-sided Fisher's exact test was used, in combination with Bonferroni correction for multiple testing (dividing 0.05 by the number of clinical features that are tested for that specific syndrome), to determine statistically significant enrichments. To minimize spurious correlations, the enrichment analysis was performed with different cutoff values for the minimal prevalence (that is, clinical features with a prevalence lower than the cutoff value in either the syndrome or PhenomAD-NDD were excluded in that specific analysis). A higher threshold leads (by definition) to fewer included features that are usually more specific for that genetic syndrome. Furthermore, to ensure a fair comparison, the individuals diagnosed with these syndromes were not included in the dataset of PhenomAD-NDD.

The available HPO terms in the clinical synopses in OMIM[27] were also collected for these 33 syndromes using the OMIM application programming interface, which is the de facto standard of phenotypic information in clinical practice. For all enriched symptoms, it was noted whether the clinical synopsis actually included that specific clinical feature, or whether it was missing. To ensure a fair comparison, all

parent-HPO-nodes of all included HPO terms in the clinical synopsis were first made positive.

Finally, because individual sex and age data were not available for the data collected from the literature and we aimed to investigate a potential confounding effect of these features, we carried out a logistic regression analysis for the in-house data (for which individual sex and age data were available). However, using all HPO terms including parent-HPO-nodes was not feasible, because the high multicollinearity between these features leads to issues in regression (a well-known problem[42]).

Therefore, for every feature that was statistically significantly enriched according to Fisher's exact test, we checked the in-house data using a logistic regression for the odds ratio (including 95% confidence intervals) and added these to the results (Supplementary Table 2). If the HPO term was a parent-HPO-node and therefore not included in the regression analysis, we noted 'not available' in the results. Finally, for every enrichment analysis, we added the sex and age as separate features to the results to demonstrate the possible influence these had on the results.

Ethics statement

In this study, data from the Biobank 'Intellectual Disability', part of the Radboud Biobank Initiative, were used. This research complies with all relevant ethical regulations and the use of this dataset was approved by the ethical committee of Radboud University Medical Center (#2018-4733).

Statistical analysis statement

All relevant statistical tests are described in Methods. When the analysis is described, the corresponding statistical test is mentioned. All analyses were done using SciPy 1.9.3, NumPy 1.23.5, NetworkX 2.8.8 and Pandas 1.5.1 in Python 3.10. The interactive plot of PhenomAD-NDD was developed using Bokeh 2.4.3 and for the enrichment analysis tool, PyScript 1.0 was used.

Statistics and reproducibility

No statistical method was used to predetermine sample size, as all individuals seen at our center were eligible, and for the literature search we aimed to include as much as possible. Because we included over 50,000 individuals and therefore had a very large dataset, this seems sufficient. This is also shown in the 95% confidence intervals from the prevalences of the clinical features. No data were excluded from the analyses, but some studies were excluded from the meta-analysis during collection of the data. For every study we considered and screened, the reason for exclusion (when excluded) is noted in Supplementary Table 2. The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this Article.

Data availability

The contents of PhenomAD-NDD are freely available online at https://humandiseasegenes.nl/phenomadndd/, as well as a tool to calculate the enrichment of symptoms for your own dataset. The summarized data are available at https://github.com/ldingemans/PhenomAD_NDD_data/. PubMed (https://pubmed.ncbi.nlm.nih.gov/) was used for the literature search in this study.

Code availability

If the tools provided online are insufficient, or if others would like to implement the enrichment analyses themselves, all code and summarized data used and developed in this study are available at https://github.com/ldingemans/PhenomAD_NDD_data/ (ref. 43).

References

40. Manders, P., Lutomski, J. E., Smit, C., Swinkels, D. W. & Zielhuis, G. A. The Radboud Biobank: a central facility for disease-based biobanks to optimise use and distribution of biomaterial for scientific research in the Radboud University Medical Center, Nijmegen. *Open J. Bioresour.* **5**, 2 (2018).
41. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
42. Daoud, J. Multicollinearity and regression analysis. *J. Phys. Conf. Ser.* **949**, 012009 (2017).
43. Dingemans, L. ldingemans/PhenomAD_NDD_data: original release as in the paper. *Zenodo* https://doi.org/10.5281/zenodo.10908427 (2024).

Acknowledgements

We are grateful to the Dutch Organisation for Health Research and Development for ZON-MW grant no. 912-12-109 (to B.B.A.d.V. and L.E.L.M.V.), Donders Junior researcher grant no. 2019 (to B.B.A.d.V. and L.E.L.M.V.) and Aspasia grant no. 015.014.066 (to L.E.L.M.V.). The aims of this study contribute to the Solve-RD project (to L.E.L.M.V.), which has received funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement no. 779257. Multiple authors of this publication are members of the European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability ERN-ITHACA (EU Framework Partnership Agreement ID: 3HP-HPFPA ERN-01-2016/739516). The funder(s) of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Author contributions

This research included worldwide researchers who were involved throughout the research process and each had their own role and responsibility. The study was designed by L.E.L.M.V. and B.B.A.d.V., and implemented by A.J.M.D., L.E.L.M.V. and B.B.A.d.V. The data are owned by the various independent researchers (A.J.M.D., S.J., J.v.R., N.d.L., R.P., J.S.-H., B.W.v.B., C.M., C.W.O., M.W., P.J.v.d.S., G.W.E.S., R.F.K., A.T.V.-v.S., T.K., D.A.K., L.E.L.M.V. and B.B.A.d.V.), whereas the intellectual property lies with the leading research group from the Radboud University Medical Center. The original draft was written by A.J.M.D., L.E.L.M.V. and B.B.A.d.V. All authors were involved in reviewing and editing the paper.

Competing interests

The authors declare no competing interests.

Additional information

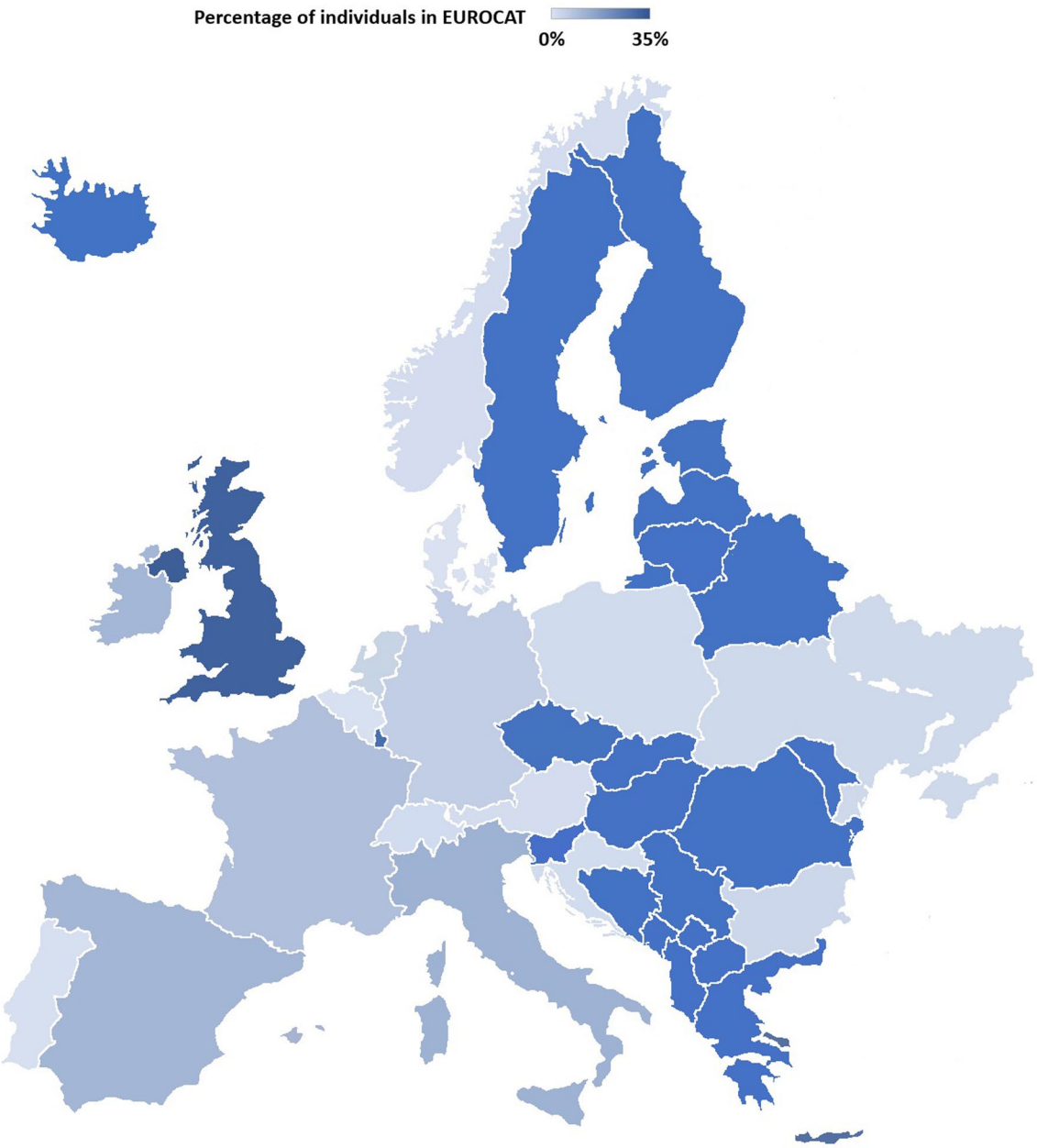
Extended data is available for this paper at https://doi.org/10.1038/s41591-024-03005-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-024-03005-7.

Correspondence and requests for materials should be addressed to Bert B. A. de Vries.

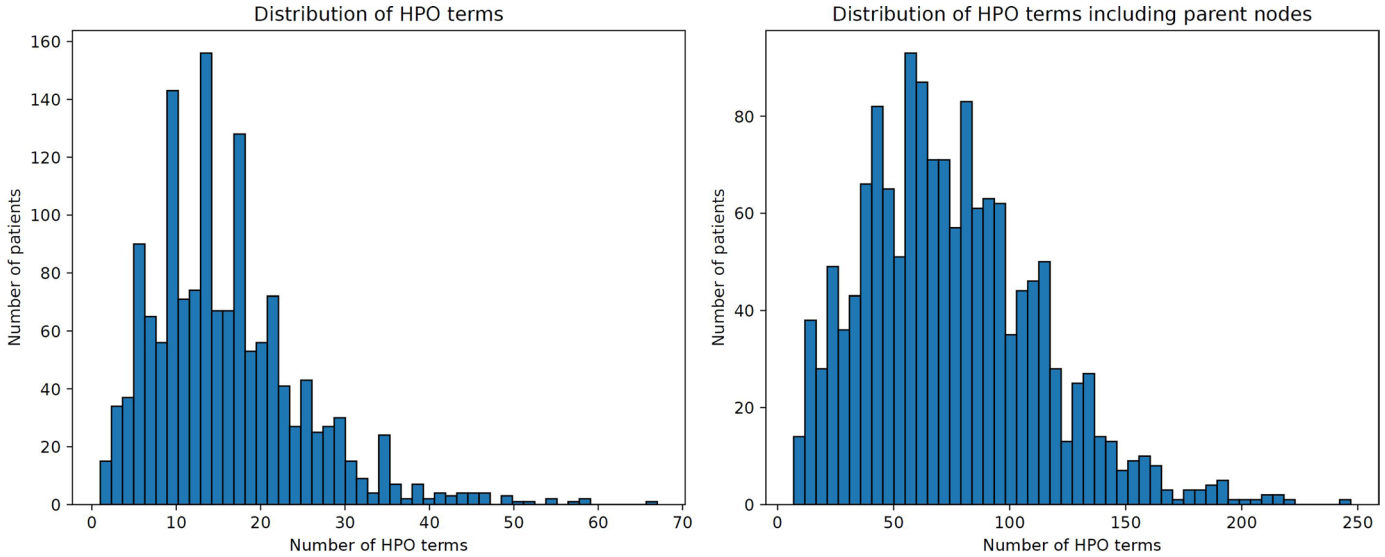
Peer review information *Nature Medicine* thanks Kristopher Kahle and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Sonia Muliyil, in collaboration with the *Nature Medicine* team.

https://doi.org/10.1038/s41591-024-03005-7



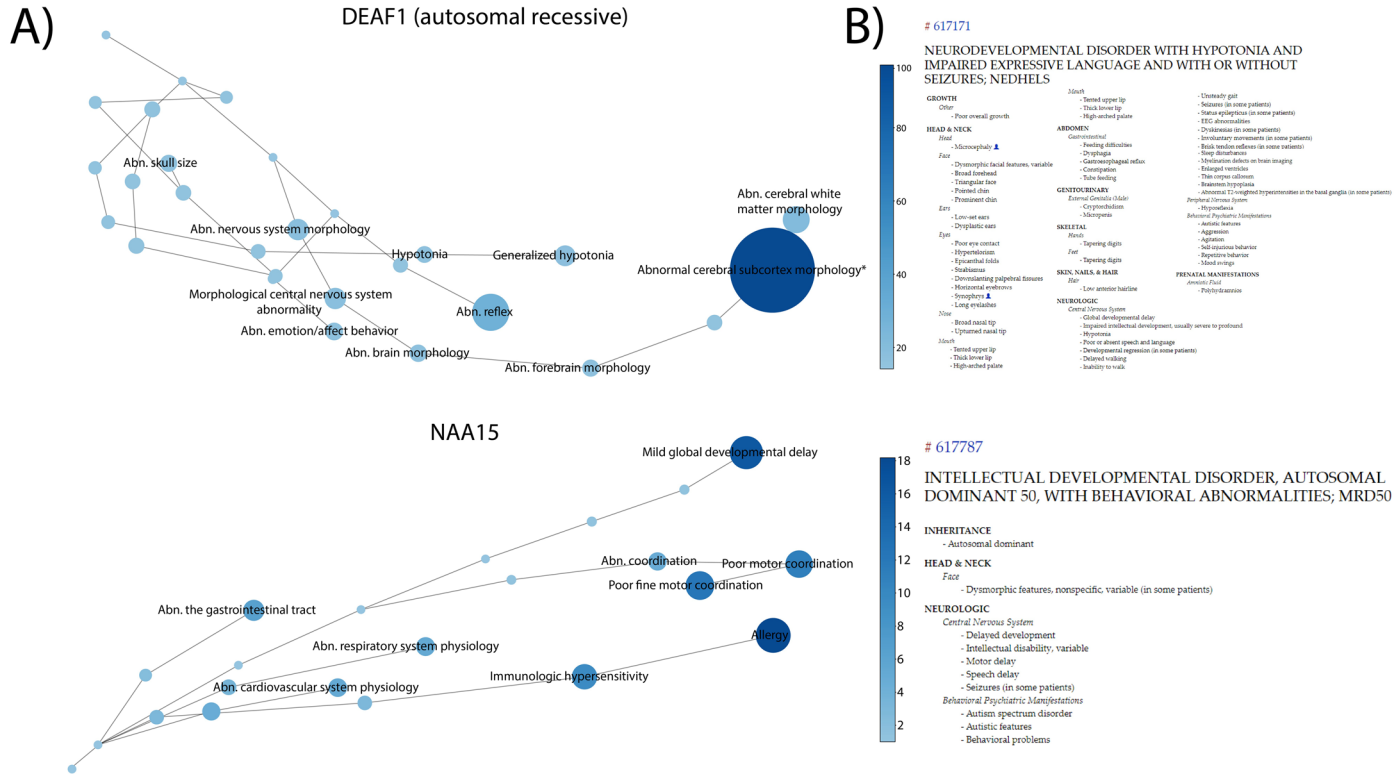
Extended Data Fig. 1 | Countries contributing to EUROCAT. The 18 countries whose institutions contribute data to the EUROCAT registry, with their relative contribution in various shades of blue. Of note, the total number of individuals in the EUROCAT registry is 11,616,332.

https://doi.org/10.1038/s41591-024-03005-7



Extended Data Fig. 2 | Distribution of number of clinical features. The distribution of the HPO terms of the individuals seen at the outpatient clinic in the Radboud University Medical Center.

https://doi.org/10.1038/s41591-024-03005-7



Extended Data Fig. 3 | Result of enrichment analysis. A) The results of the top 10 results of the enrichment analysis for two of the 33 investigated genetic syndromes are shown here: MRD50 (OMIM #617787), for which 16% of the symptoms are described in the clinical synopsis in OMIM, and NEDHELS (OMIM #617171, 100% concordance). Shown are the 10 clinical features that are the most enriched: that is with the highest ratio when dividing the prevalence in the syndrome with the prevalence in PhenomAD-NDD. The size of the nodes in the network corresponds to the prevalence in that specific genetic syndrome. The phenotypic data of the investigated syndromes was gathered from the specific website of the Human Disease Gene website series, with the prevalence rate in the syndrome referring to the proportion of individuals with the specific genetic syndrome who exhibit a particular clinical feature. The color corresponds to the enrichment ratio, which compares the prevalence in the syndrome to the prevalence in PhenomAD-NDD. To avoid spurious enrichments, only clinical features with a prevalence of at least 1% are shown in this figure. Using the developed tool in either the electronic supplement or online at the Human Disease Genes website series (https://humandiseasegenes.nl/phenomadndd/)[6], others are able to determine enriched clinical features and generate these figures for their syndrome of interest. **B**) The clinical synopsis of OMIM[8] as shown for these two genetic syndromes.

nature portfolio

Corresponding author(s): Bert B.A. de Vries

Last updated by author(s): Apr 8, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We conducted the searches for literature on PubMed (https://pubmed.ncbi.nlm.nih.gov/) and checked the references of included paper for more data.
Data analysis	All data were analysed in Python 3.10 using NumPy 1.23.5, NetworkX 2.8.8 and Pandas 1.5.1. The tool was created with Bokeh 2.4.3 and PyScript 1.0. Summarized data and generated code are available at https://github.com/ldingemans/PhenomAD_NDD_data/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Excel 365 and PubMed were used to collect the data and create the data set for meta-analysis. The contents of PhenomAD-NDD are freely available online at

https://humandiseasegenes.nl/phenomad-ndd/. There, a tool to calculate enrichment of symptoms for your own data set is available as well.Furthermore, summarized data and generated code are available as well at https://github.com/ldingemans/PhenomAD_NDD_data/.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex was used in the study and reported on in a subcohort (the data collected in our own clinic), to enable comparison between sexes.
Reporting on race, ethnicity, or other socially relevant groupings	These were not reported as these were not deemed relevant.
Population characteristics	Age, sex, but mainly phenotypic information.
Recruitment	In our center, any pediatric individual with a neurodevelopmental disorder was eligible for inclusion. For the literature review and meta-analysis, we looked for studies reporting on somatic comorbidity in pediatric individuals with a neurodevelopmental disorder. Studies focussing on or reporting only on psychiatric diseases were excluded.
Ethics oversight	The medical ethical review board of the Radboud university medical center approved the use of data for the purpose of this study (#2018-4733).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size as calculated, as all individuals seen at our center were eligible, and for the literature search we aimed to include as much as possible. Since we included over 50,000 individuals and therefore have a very large data set, this seems sufficient. This is also shown in the 95% confidence intervals from the prevalences of the clinical features.
Data exclusions	No data were excluded from analyses, but some studies were excluded from the meta-analysis during the collection of the data. For every study we considered and screened, the reason for exclusion (when excluded) is noted in supplementary table 2.
Replication	We verified the enrichment analysis in 33 separate syndromes. Replicating this is hard, as one needs another data source with the same high-quality phenotypic data for the genetic syndromes, that is currently (to our knowledge) not available.
Randomization	We performed some subgroup analyses: in these cases, the subgroup covariate determined how participants were allocated into the groups (i.e. the phenotype of females compared to males). In another comparison, we compared the phenotypic data of a specific genetic syndrome to that in our dataset, and thus the genetic diagnosis determined the allocation. Randomization was therefore not applicable.
Blinding	All analyses were done on the complete groups, there were no exclusions during the analysis and therefore there was no blinding. Blinding was not relevant to data collection, as characteristics such as sex- and age are not subjective.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Not applicable since it is not a trial.
Study protocol	Not applicable since it is not a trial.
Data collection	For the literature review, PubMed was accessed in February 2021. For the data collected in our center, individuals who were referred to the Radboud university medical center between 2009 and 2021 for neurodevelopmental disorders were eligible for inclusion.
Outcomes	Primary outcome measure is prevalence of the clinical features.