medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Inherited variants in *CHD3* demonstrate variable expressivity in Snijders Blok-Campeau syndrome

Jet van der Spek,[1,2,37] Joery den Hoed,[3,4,37] Lot Snijders Blok,[1,2,3] Alexander J. M. Dingemans,[1,2] Dick Schijven,[3] Christoffer Nellaker,[5,6,7] Hanka Venselaar,[8] Tahsin Stefan Barakat,[9] E. Martina Bebin,[10] Stefanie Beck-Wödl,[11] Gea Beunders,[12] Natasha J. Brown,[13,14] Theresa Brunet,[15] Han G. Brunner,[1,2,16] Philippe M. Campeau,[17,18] Goran Čuturilo,[19,20] Christian Gilissen,[1,21] Tobias B. Haack,[11] Ralf A. Husain,[22] Benjamin Kamien,[23] Sze Chern Lim,[13] Luca Lovrecic,[24] Janine Magg,[25] Ales Maver,[24] Valancy Miranda,[17,18] Danielle C. Monteil,[26] Charlotte W. Ockeloen,[1] Lynn S. Pais,[27] Vasilica Plaiasu,[28] Laura Raiti,[13] Christopher Richmond,[13,29,30] Angelika Rieß,[11] Eva M. C. Schwaibold,[31] Marleen E. H. Simon,[32] Stephanie Spranger,[33] Tiong Yang Tan,[13,14] Michelle L. Thompson,[34] Bert B.A. de Vries,[1,2] Ella J. Wilkins,[13] Marjolein H. Willemsen,[1] Clyde Francks,[3,35] Lisenka E. L. M. Vissers,[1,2] Simon E. Fisher,[3,35] Tjitske Kleefstra,[1,2,36*]

[1]Department of Human Genetics, Radboud University Medical Center, 6500 HB Nijmegen, Netherlands

[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, 6500 GL Nijmegen, Netherlands

[3]Language and Genetics, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, Netherlands

[4]International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, Netherlands

[5]Nuffield Department of Women's and Reproductive Health, University of Oxford, Women's Centre, John Radcliffe Hospital, Oxford OX3 9DU, UK

[6]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[7]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK

[8]Centre for Molecular and Biomolecular Informatics, Radboud University Nijmegen Medical Centre, 6525 GA, Nijmegen, Netherlands

[9]Department of Clinical Genetics, Erasmus MC University Medical Center, 3015 CN Rotterdam, Netherlands

[10]Department of Neurology, University of Alabama at Birmingham, Birmingham, AL, 35233, United States

[11]Department of Medical Genetics and Applied Genomics, University of Tübingen, Calwerstr. 7, 72076 Tübingen, Germany

[12]Department of Genetics, University Medical Center Groningen, 9700 RB Groningen, Netherlands

[13]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, Victoria, 3052, Australia

[14]University of Melbourne, Department of Pediatrics, Royal Children's Hospital, Parkville, Victoria, 3052, Australia

[15]Institute of Human Genetics, School of Medicine, Technical University Munich, 81675, Munich, Germany

[16]Maastricht University Medical Center, Department of Clinical Genetics, GROW School for Oncology and Developmental Biology, and MHeNS School for Mental health and Neuroscience, PO Box 5800, 6202 AZ, Maastricht, Netherlands

[17]CHU Sainte-Justine Research Center, Montreal, QC H3T 1C5, Canada

[18]Sainte-Justine Hospital, University of Montreal, Montreal, QC H3T 1C5, Canada

[19]University Children's Hospital Belgrade, 11000, Belgrade, Serbia

[20]Faculty of Medicine, University of Belgrade, 11000, Belgrade, Serbia

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[21]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6500 HB Nijmegen, Netherlands

[22]Department of Neuropediatrics, Jena University Hospital, 07747 Jena, Germany

[23]Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, 6008, Australia

[24]Clinical Institute for Genomic Medicine, University Medical Center Ljubljana, Zaloska cesta 2, 1000 Ljubljana, Slovenia

[25]University Children's Hospital, Department of Neuropaediatrics, Developmental Neurology, Social Paediatrics, University of Tübingen, Hoppe-Seyler Str. 1, 72076 Tübingen, Germany

[26]Naval Medical Center Portsmouth, Department of Pediatrics, 23322 Portsmouth, VA, United States

[27]Broad Institute Center for Mendelian Genomics, Broad Institute of MIT and Harvard, 02142, Cambridge, United States

[28]INSMC Alessandrescu-Rusescu, Regional Center of Medical Genetics Bucharest, 020395 Bucharest, Romania

[29]Genetic Health Queensland, Royal Brisbane & Women's hospital, Herston, QLD, 4006, Australia

[30]School of Medicine, Griffith University, Southport, QLD, 4215, Australia

[31]Institute of Human Genetics, Heidelberg University, 69120, Heidelberg, Germany

[32]Department of Medical Genetics, University Medical Center Utrecht, 3508 AB, Utrecht, Netherlands

[33]Praxis für Humangenetik-Bremen, 28209, Bremen, Germany

[34]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, United States

[35]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 GL, Nijmegen, Netherlands

[36]Centre of Excellence for Neuropsychiatry, Vincent van Gogh Institute for Psychiatry, 5803 AC, Venray, Netherlands

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[37]These authors contributed equally to this work

*Correspondence: Tjitske.Kleefstra@radboudumc.nl

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Summary (250 words)

Interpretation of next-generation sequencing data of individuals with an apparent sporadic neurodevelopmental disorder (NDD) often focusses on pathogenic variants in genes associated with NDD, assuming full clinical penetrance with limited variable expressivity. Consequently, inherited variants in genes associated with dominant disorders may be overlooked when the transmitting parent is clinically unaffected. While *de novo* variants explain a substantial proportion of cases with NDDs, a significant number remains undiagnosed possibly explained by coding variants associated with reduced penetrance and variable expressivity. We characterized twenty families with inherited heterozygous missense or protein-truncating variants (PTVs) in *CHD3*, a gene in which *de novo* variants cause Snijders Blok-Campeau syndrome, characterized by intellectual disability, speech delay and recognizable facial features (SNIBCPS). Notably, the majority of the inherited *CHD3* variants were maternally transmitted. Computational facial and human phenotype ontology-based comparisons demonstrated that the phenotypic features of probands with inherited *CHD3* variants overlap with the phenotype previously associated with *de novo* variants in the gene, while carrier parents are mildly or not affected, suggesting variable expressivity. Additionally, similarly reduced expression levels of CHD3 protein in cells of an affected proband and of related healthy carriers with a *CHD3* PTV, suggested that compensation of expression from the wildtype allele is unlikely to be an underlying mechanism. Our results point to a significant role of inherited variation in SNIBCPS, a finding that is critical for correct variant interpretation and genetic counseling and warrants further investigation towards understanding the broader contributions of such variation to the landscape of human disease.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Main text (2297 words)

The availability of whole exome sequencing (WES) in clinical practice has greatly improved the yield of genetic diagnostics for individuals with neurodevelopmental disorders (NDDs). In particular, sequencing of proband-parent trios, followed by filtering for *de novo* [1-3] or bi-allelic variants [4; 5], has proven a powerful tool to identify causal variants in individuals with sporadic dominant and recessive NDDs. However, while *de novo* and bi-allelic variants explain a substantial proportion of cases with NDDs [1; 4; 5], the majority remains undiagnosed [6]. Various factors may explain the difficulties to diagnose these individuals, including variation in genes not yet associated to disease, polygenic inheritance or variation in non-coding regions [7]. Also coding variants associated with reduced penetrance and variable expressivity may underlie unexplained NDD cases [6; 8]. Common diagnostic strategies to analyze next-generation sequencing data are not optimized to identify the contributions of these factors to disease. While penetrance indicates the proportion of carriers of a particular variant with a phenotype, expressivity describes the variability in severity of the phenotype between carriers of this variant [9]. Variable expressivity can cause highly variable symptoms, even in severe disorders that are caused by variants with a large effect [9; 10].

In the present study we show variable expressivity for variation in *CHD3.* CHD3 is an ATP-dependent chromatin remodeling protein that serves as core member of the NuRD complex [11]. Heterozygous variants in *CHD3* have recently been shown to cause a neurodevelopmental syndrome with a variable phenotype, ranging from mildly to more severely affected cases (MIM #618205, Snijders Blok-Campeau syndrome: SNIBCPS) [12; 13]. *CHD3* is extremely intolerant for both loss-of-function (LoF) and missense variation (pLI = 1, o/e = 0.09 (0.05 - 0.15); Z = 6.15, o/e = 0.5 (0.46 – 0.53)), suggesting haploinsufficiency as a possible disease mechanism. However, the large majority of cases diagnosed with SNIBCPS carry confirmed *de novo* missense variants or single amino acid in-frame deletion variants (51/55, 93% of cases) [12-14], clustering in the ATPase-Helicase domain of the encoded protein, and affecting its ATPase activity and/or

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

chromatin remodeling functions, which could be consistent with a dominant-negative mechanism [12].

Here, we identified twenty families with SNIBCPS, each initially identified through a proband diagnosed with a syndromic NDD carrying a rare inherited *CHD3* missense variant (n = 12) or PTV (n = 8) (NM_001005273.2/ENST00000330494.7; Figure 1). Based on clinical observations, all probands had phenotypes overlapping with the SNIBCPS-phenotype associated with *de novo* variants in *CHD3* (Figure 2; Supplemental Notes 1; Table 1, S1 and S2). Computational facial analysis also confirmed the presence of a SNIBCPS facial gestalt in probands (Figure S3; Table S2), and composite images showed similarities in facial features between probands with *de novo* and inherited *CHD3* variants (squared face, deep set eyes, pointed chin; Figure 2B).

All of the carrier parents for whom phenotypic information was available about at least development and dysmorphisms (n = 6) had at least one feature of SNIBCPS (Table S2), although in five parents (31%) this was limited to only one (family 1 and 10) or two phenotypic features (family 9, 15 and 20; Table S2). Whereas the majority of carrier parents (15/16, 94%) presented with a single (e.g. prominent forehead or deep-set eyes) or several facial features known in SNIBCPS and 53% had macrocephaly (8/15) (Table 1 and S2), the parents had either mild/borderline intellectual disability (n = 4, 22%) or no history of intellectual disability (n = 14) (Table 1, S1 and S2; Supplemental Notes 1). Taken together, these observations suggest a combination of both variable expressivity and reduced penetrance for these rare genetic variations in *CHD3.*

We more objectively compared the phenotypes of probands with *de novo* and inherited *CHD3* variants based on Human Phenotype Ontology (HPO) terminology [15], using a Partitioning Around Medoids clustering algorithm [16]. While this computational analysis did not identify a phenotypic difference between probands with *de novo* and inherited variants (31 of 54 individuals clustered correctly, p = 0.31; Figure 2C and S4), it confirmed a phenotypic difference between

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

probands with inherited *CHD3* variants and their carrier parents (32 of 38 individuals clustered correctly, p = 0.00003; Figure 2C and S4). It is important to note that, although younger generations seem more severely affected than previous generations this may be due to ascertainment bias [17]. Our finding of variable expressivity of inherited *CHD3* variants is consistent with the lack of a correlation between mutation location and phenotype [13], and the variable severity of phenotypes already described for (recurrent) *de novo CHD3* variants [12; 13]. Additional evidence for variable expressivity for *CHD3* variation is provided by the recently identified association of 19 rare *CHD3* missense variants with Chiari I malformations in individuals without features of SNIBCPS [18].

We noticed that the majority of variants in our cohort were maternally inherited (14/20, 70%, p = 0.0577; Figure 1B, 1C and S5A). For single nucleotide variants with a LoF effect, 6/7 (86%, p = 0.0625) variants were maternally inherited (Figure 1B and S5A). Notably, the only father transmitting a LoF single nucleotide variant was affected (mild intellectual disability). This observation could hint at a female-protective effect for genetic variation in *CHD3*. Previous studies have repeatedly demonstrated a male bias in NDDs, a higher pathogenic variant burden in females and a maternal transmission bias in rare inherited variants [6; 19-26], suggesting that female gender protects against genetic variation in disease. This phenomenon might contribute to the variable expressivity observed for the inherited *CHD3* variants. However, we did not observe a sex-bias in the affected probands (12/20 female, $p > 0.9999$), or more severe intellectual disability in male compared to female *de novo* or inherited cases [12; 13]. To further explore the hypothesis of a female protective effect at population level, we analyzed all *CHD3* LoF variants in gnomAD (n = 15 in 198,800 individuals) and found that females had a significantly higher carrier rate than males (12/15, p = 0.0173; Figure S5B).

Few cases with SNIBCPS have been described with confirmed *de novo CHD3* PTVs (4/55, 7.3% cases) [12; 13] including one which is predicted to escape nonsense-mediated decay (NMD; NP_001005273.1:p.Phe1935GlufsTer108). However, in our study we identified seven

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

families with inherited single nucleotide PTVs and one with an intragenic deletion with a predicted LoF effect (8/20, 40%; Figure 1A). None of the inherited PTVs were predicted to escape NMD. We functionally confirmed this in family 1 (Figure 3A), for which we treated lymphoblastoid cell lines from the proband (individual III-2), carrier mother (II-2) and grandmother (I-2) and non-carrier healthy sibling of the proband (III-1) with cycloheximide to inhibit NMD, followed by direct amplification and Sanger sequencing of the *CHD3* transcript. We found that treatment with cycloheximide increased the expression of mutant allele, showing that the NM_001005273.2:c.3473G>A variant was targeted by NMD in all samples, as expected (Figure 3B).

An explanation for variable expressivity of PTVs could be compensation of expression by the wildtype allele to maintain normal expression levels [27]. To test if such compensation plays a role in variable expressivity of *CHD3* PTVs, we evaluated the expression of the *CHD3* variant in family 1 (c.3473G>A, p.W1158*) on a transcript and protein level. We found that this variant resulted in lower levels of *CHD3* transcript and CHD3 protein levels, when compared to lymphoblastoid cells from the non-carrier healthy sibling (Figure 3C and 3D). These findings confirm the LoF effect of the stop-gain variant in this family, and makes compensation by the wildtype allele as an underlying mechanism for the milder phenotype in the carrier mother and grandmother unlikely. It remains, however, to be determined, whether such LoF variance can have a tissue-specific, temporal expression specific, and/or transcript specific effect. It is unclear whether results from blood-derived cells can be extrapolated to neuronal cell types, which would be more relevant considering the NDD phenotypes in our cohort, especially given that neuron-specific alternative splicing has previously been described for *CHD3* [28]. Other explanations for the clinical variable expressivity of inherited *CHD3* variants include the presence of a second-hit on the other allele by either rare or common variation, a genome wide higher mutational burden of high-penetrant variants, or common variants in promoter/enhancers regions or in other genes, inherited from the non-carrier parent [9; 29; 30]. Such a compound inheritance mechanism, has, for

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

example, been described for thrombocytopenia with absent radii (TAR) syndrome, where the inheritance of a rare null allele together with one of two low-frequency SNPs in regulatory regions causes disease [31]. In four probands with an inherited *CHD3* variant a copy number variant (CNV) was also reported, including one 22q11.2 duplication, which has been described with highly variable features (MIM # 608363) and three CNVs of unknown significance (Table S1). Proband 7 had other (likely) pathogenic variants contributing to the phenotype (Table S1; Supplemental Notes 1). A comparison with the prevalence of additional genetics finding in individuals with *de novo CHD3* variants could not be made due to lack of reporting on additional genetic findings [12; 13].

In addition to the seven inherited *CHD3* single nucleotide PTVs and the intragenic deletion, we identified 12 families with inherited missense variants. One of the identified inherited missense variants also present in an unaffected carrier parent was identical to a variant previously reported as a *de novo* variant in an individual with SNIBCPS (p.R1342Q; individual 32 in [12]). Based on the phenotypes observed in the probands with inherited *CHD3* missense variants, the conservation of affected positions (Figure S1), and *in silico* predictions of pathogenicity (Figure S6; Table S1), we considered these inherited *CHD3* missense variants as likely pathogenic with variable expressivity in the parents. Clinically, probands carrying a *CHD3* missense variant did not seem to be more severely affected than individuals with PTVs (Table S3). The individuals with *de novo* missense variants published to date were mostly (although not entirely) localized to the ATPase-Helicase domain [12; 13]. No clustering to the ATPase-Helicase domain or elsewhere was observed among the inherited missense variants of our cohort (Figure 1A). It has been speculated that the *de novo* missense variants clustering in the ATPase-Helicase domain are unlikely to lead to a sole LoF effect [12], and may potentially act in a dominant-negative way. The identification of eight families with an inherited LoF variant and the lack of clustering of the inherited missense variants may suggest a LoF effect as the main mechanism for inherited cases, which may underlie the variable expressivity. However, our cell-based analyses of chromatin binding (for p.S477F)

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and GATAD2B-binding (for p.R1342Q, p.E1837K and p.Q1888R) did not find evidence of LoF for the protein functions that we tested (Figure S7). This does not exclude that these variants have an effect on other biological functions of CHD3. Based on 3D-protein modeling, the prior published *de novo* missense variants within the ATPase-Helicase domain localize more closely to the ATP-binding site than the inherited missense variants of our cohort (Supplemental Notes 2). Interestingly, the p.I983V (family 13) variant was found to be closer to published *de novo* variants (Supplemental Notes 2) and the carrier parent with this missense variant did have a neurodevelopmental phenotype which was more pronounced than in other carrier parents (Figure 1C and 2; Table S2; Supplemental Notes 1).

The presence of rare, likely pathogenic *CHD3* variants in healthy individuals prompted us to study possible effects of variation in this gene at a population level, using data from the UK Biobank resource [32-37]. For a detailed description of these analyses, see Supplemental Notes 3. We found no associations between rare missense variation at mutation-intolerant locations in *CHD3* (minor allele frequency ≤ 1%, located in functional domains, damaging in PolyPhen or SIFT and with a CADD-PHRED score > 25) and fluid intelligence (N = 77,998), educational qualification (N = 120,596) or intracranial volume (N = 18,254). However, we identified a group difference in intracranial volume between rare *CHD3* putative LoF variant carriers and non-carriers, with carriers having an increased volume compared to controls (n = 4, $t = 2.37$, $p = 0.018$). This seems consistent with the observation of macrocephaly in 44-53% of probands with a (likely) pathogenic *CHD3* variant and in 53% of carrier parents (Table 1), and the link of rare *CHD* variants with abnormal brain growth [18]. To test possible relationships between *CHD3* common genetic variation and head circumference and/or intracranial volume, we performed gene-level analyses using previously published SNP-wise association summary statistics for these traits [38; 39], but no tests survived multiple testing correction (Supplemental Notes 3).

With the identification and characterization of inherited *CHD3* variants with variable expressivity in twenty families, we showed that, in addition to highly penetrant *de novo* variants,

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

rare predicted likely pathogenic inherited variants in *CHD3* should be considered as possibly pathogenic depending on mutation characteristics in cases with phenotypic concordance to SNIBCPS. Interestingly, variable penetrance and expressivity has been noted in numerous families with another dominant NDD, KBG syndrome, caused by LoF variants in *ANKRD11* (MIM #148050) [40; 41]. So this phenomenon is likely more common for dominant NDDs, with important implications for clinical genetic counseling, in the context of recurrence risk, prenatal diagnostics, prognosis and variant interpretation.

Clinically, we recommend that it can be helpful to evaluate the parents of children with *CHD3* variants for subtle SNIBCPS features. In particular macrocephaly and facial dysmorphisms including a prominent forehead and pointed chin could be recognized in a substantial number of carrier parents (53% and 94% respectively; Figure 2A; Table S2). Taken together, our results illustrate the continuum of causality for NDDs with genetic origins [17; 42] and significantly underline the hypothesis that variable expressivity and reduced penetrance likely explain a large portion of as yet unexplained NDD cases. Overall, we show that even for genes already known to be implicated in a NDD inherited variation and variable expressivity can play a major role and are thus important to consider in genetic counseling.

Data availability

Supplementary Notes 1, Table S1, facial photographs (Fig. 2A and Fig. S2) and all datasets generated and analyzed during the current study are available from the corresponding author on request.

Acknowledgements

We are extremely grateful to all families participating in this study. In addition, we wish to thank the members of the Cell culture facility, Department of Human Genetics, RadboudUMC, Nijmegen for culture of cell lines. This work was financially supported by the Dutch Research Council grant

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

to TK (015.014.036 and 1160.18.320) and LELMV (015014066) and Netherlands Organization for Health Research and Development to TK (91718310) and LELMV (843002608, 846002003), by Donders Junior Researcher Grant 2019 to TK and LELMV and the Max Planck Society (JdH, DS, CF, SEF). 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-HP-FPA ERN-01-2016/739516]. The aims of this study contribute to the Solve-RD project (CG, HGB, LELMV and TK) which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779257. For families 10,13,14 and 19 sequencing and analysis were provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and were funded by the National Human Genome Research Institute, the National Eye Institute, and the National Heart, Lung and Blood Institute grant UM1 HG008900 and in part by National Human Genome Research Institute grant R01 HG009141. For family 15 exome sequencing was performed in the framework of the German project "TRANSLATE NAMSE", an initiative from the National Action League for People with Rare Diseases (Nationales Aktionsbündnis für Menschen mit Seltenen Erkrankungen, NAMSE) facilitating genetic diagnostics for individuals with suggested rare diseases. Part of this research has been conducted using the UK Biobank Resource under application number 16066, with Clyde Francks as the principal applicant. Our study made use of imaging-derived phenotypes generated by an image-processing pipeline developed and run on behalf of UK Biobank.

Materials and Methods

Individuals and consent

The cohort presented in this study was assembled from hospitals and laboratories across the Netherlands, Germany, United States of America, Slovenia, Australia and Canada. Informed consent for the use and publication of medical data and biological material was obtained from all

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

patients or their legal representative by the involved clinician. Consent for publication of photographs was obtained separately. Genetic testing and research were performed in accordance with protocols approved by the local Institutional Review Boards.

Next-generation-sequencing

CHD3 variants in all probands were identified using whole exome sequencing (WES) or whole genome sequencing (WGS; family 4 and 12), with filtering as previously described [3; 43-52]. Due to inheritance from seemingly healthy/mildly affected parents the *CHD3* variants were initially classified as variants of unknown significance. Inheritance of variants was confirmed either as part of trio WES or using targeted Sanger sequencing after identification in singleton exon analysis. Similarly, if applicable, other family members were tested using targeted sequencing.

Pathogenicity of missense variants was further evaluated using CADD-PHRED v1.6 [53], PolyPhen-2 [54] and SIFT [55] scores. Allele frequencies of all variants in gnomAD were based on ENST00000330494.7 [14].

Facial analysis

We established a 2D hybrid facial model which combines the analysis of the 'Clinical Face Phenotype Space' pipeline with the facial recognition system of the 'OpenFace' pipeline [56; 57]. First, we generated a 468-dimensional feature vector of the facial features of 30 individuals with *de novo CHD3* variants. After extraction of the hybrid features for each of the individuals, we calculated whether the individuals with *de novo CHD3* variants cluster together when compared to a group of matched controls based on the nearest neighbor principle (Euclidean distance) – these matched controls were individuals with ID and are age-, ethnicity- and gender matched. The Mann-Withney U test was used to determine whether the clustering of individuals with *de novo CHD3* variants was significantly higher than expected based on random chance. A *p*-value smaller than 0.05 was considered significant.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Furthermore, a classifier was built using a logistic regression model trained on the 468-dimensional feature vector of the 30 individuals. The performance was evaluated performing leave-one-out cross validation and the classifier was shown to have a sensitivity of 0.91, a specificity of 0.83 and an overall area under the ROC-curve of 0.91. Finally, using the trained classifier, we determined for each inherited case whether that individual clusters within the *de novo CHD3* group or the control group (Figure S3).

Construction of composite face

For 12 individuals with an inherited *CHD3* variant and 30 with a *de novo CHD3* variant, facial 2D-photographs were available for generating a composite face. As previously described, average faces were generated while allowing for asymmetry preservation and equal representation by individuals [58].

Human Phenotype Ontology (HPO)-based phenotype clustering analysis

HPO-based clustering analysis was performed as described elsewhere [59]. We included 35 individuals with *de novo CHD3* variants [12], 19 of 20 probands with an inherited *CHD3* variant, and 19 of 20 carrier parents in the analysis: the proband and carrier mother of family 6 were excluded because no clinical data were available, and the mother is mosaic for the *CHD3* variant (~37%). The Wang score (a measure of semantic similarity) between all terms was calculated using the HPO Sim package [60; 61]. The terms were divided in groups, based on the similarity score: a new feature – the sum of the terms in the group - was created as a replacement for the terms in that specific group (Figure S4A; Table S4). HPO terms that could not be added to a group feature were added as a separate term. To quantify and visualize possible differences in our cohort, we used Partitioning Around Medoids (PAM) clustering on these grouped features. We compared probands with a *de novo* and inherited variant and probands with inherited variants and their carrier parents in a second analysis. To assess statistical significance, a permutations test (*n* =

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

100,000) was used with relabeling based on variant types, while keeping the original distribution of variant types into account.

Three-dimensional protein modeling

We modeled the protein structure of the ATPase-Helicase domain of CHD3 in interaction with the DNA using the homology modeling script in the WHAT IF [62] & YASARA [63] Twinset with standard parameters. As a template, we used PDB file 6RYR which contains the human Nucleosome-CHD4 complex structure of a single copy of CHD4 [64]. The PHD2 variant (p.S477F) was modeled in the PHD2 domain of CHD4 (PDB 2L75, 89% sequence identity with CHD3) [65].

DNA expression constructs and site-directed mutagenesis

The cloning of *CHD3* (NM_001005273.2/ENST00000330494.7) has been described previously [12]. The coding DNA sequence of *GATAD2B* (NM_020699.3/ENST00000368655.4) and a C-terminal region of CHD3-encoding residues 1246-1944 (NM_001005273.2) were amplified using primers listed in Table S6. Variants in full-length CHD3 or the C-terminal CHD3 construct were generated using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). The primers used for site-directed mutagenesis are listed in Table S7. cDNAs were subcloned using BamHI/HpaI (full-length CHD3), BamHI/XbaI (GATAD2B) or HindIII/BamHI (C-terminal CHD3 construct) into pYFP, pHisV5 and pRluc, created by modification of the pEGFP-C2 vector (Clontech) as described before [66]. All constructs were verified by Sanger sequencing.

Cell culture

Lymphoblastoid cell lines (LCLs) were established by Epstein-Barr virus transformation of peripheral lymphocytes from blood samples collected in heparin tubes, and maintained in RPMI medium (Sigma) supplemented with 15% fetal bovine serum and 5% HEPES (both Invitrogen). HEK293T/17 cells (CRL-11268, ATCC) were grown in DMEM supplemented with 10% fetal

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

bovine serum and 1x penicillin-streptomycin (all Invitrogen) at 37°C with 5% CO2. Transfections were performed using GeneJuice (Millipore) following the manufacturer's protocol.

Testing for nonsense mediated decay of truncating variants

LCLs of members of family 1 and controls were grown overnight with 100 µg/ml cycloheximide (Sigma) to block NMD. After treatment, cell pellets were collected, and RNA and protein were extracted using the RNeasy Mini Kit (Qiagen) or with 1x RIPA buffer supplemented with 1% PMSF and 1x PIC, respectively. RT-PCR was performed using SuperScript III Reverse Transcriptase (ThermoFisher) with random primers, and regions of interest were amplified from cDNA using primers listed in Table S5. Sanger trace peak sizes of the wildtype and variant allele were measured using the 'Area' option in ImageJ and proportion of the variant allele was calculated: peak area variant allele / (peak area variant + wildtype allele).

Direct fluorescent imaging

HEK293T/17 cells were grown on coverslips coated with poly-D-lysine (Sigma). Forty-eight hours after transfection with the YFP-tagged C-terminal CHD3 construct and HisV5-tagged GATAD2B, cells were fixed with 4% paraformaldehyde (PFA, Electron Microscopy Sciences). Nuclei were stained with Hoechst 33342 (Invitrogen). Fluorescence images were acquired with a Zeiss LSM880 confocal microscope and Airyscan unit using ZEN Image Software (Zeiss).

FRAP assays

HEK293T/17 cells were transfected in clear-bottomed black 96-well plates with YFP-tagged full-length CHD3 or p.S477F. After 48 h, medium was replaced with phenol red-free DMEM supplemented with 10% fetal bovine serum (both Invitrogen), and cells were moved to a temperature-controlled incubation chamber at 37°C. Fluorescent recordings were acquired using a Zeiss LSM880 and Zen Black Image Software, with an alpha Plan-Apochromat 100x/1.46 Oil

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

DIC M27 objective (Zeiss). FRAP experiments were performed by photobleaching an area of 0.98 µm x 0.98 µm within a single nucleus with 488-nm light at 100% laser power for three iterations with a pixel dwell time of 32.97 µs, followed by collection of times series of 150 images with a 2.5 zoom factor and an optical section thickness of 1.4 µm (2.0 Airy units). Individual recovery curves were background subtracted and normalized to the pre-bleach values, and mean recovery curves were calculated using EasyFRAP software [67]. Curve fitting was done with the FrapBot application using direct normalization and a single component exponential model, to calculate the half-time and maximum recovery [68].

Immunoblotting

Whole-cell lysates were collected in 1x RIPA buffer (ThermoFisher) supplemented with 1x PIC (Roche) and 1% PMSF (Sigma). Cells were lysed for 20 min at 4 °C followed by centrifugation for 20 min at 12,000 rpm. Samples were loaded on 4–15% Mini-PROTEAN TGX Precast Gels (Bio-Rad) and transferred onto polyvinylidene fluoride membranes. Membranes were blocked in 5% milk for 1 h at room temperature and then probed with rabbit-anti-CHD3 antibody (1:1000; Abcam, ab109195) or mouse-anti-GFP (1:8000; Clontech, 632380) overnight at 4°C. Next, membranes were incubated with HRP-conjugated goat-anti-rabbit or goat-anti-mouse antibody (1:10,000; Jackson ImmunoResearch) for 1.5 h at room temperature. Bands were visualized with the SuperSignal West Femto Maximum Sensitivity Substrate Reagent Kit (CHD3; ThermoFisher) or the Novex ECL Chemiluminescent Substrate Reagent Kit (YFP-fusion proteins; Invitrogen) using a ChemiDoc XRS + System (Bio-Rad).

Co-immunoprecipitation

HEK293T/17 cells were transfected with the YFP-tagged C-terminal region of CHD3 and Rluc-tagged GATAD2B. After 48h, whole-cell lysates were collected in Pierce IP Lysis Buffer (25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40 and 5% glycerol; ThermoFisher)

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

supplemented with 1x PIC (Roche) and 1% PMSF (Sigma). Cells were lysed for 20 min at 4°C followed by centrifugation for 20 min at 12,000 rpm. YFP-fusion proteins were immobilized on GFP-trap magnetic agarose beads (Chromotek) overnight at 4°C. Deactivated beads (Chromotek) were used as a negative control. The elutions and 5% of the input were resolved on 4–15% Mini-PROTEAN TGX Precast Gels (Bio-Rad) and transferred onto polyvinylidene fluoride membranes. Membranes were blocked in 5% milk for 1 h at room temperature and then probed with rabbit-anti-Rluc antibody (1:2000; GeneTex) overnight at 4°C. Next, membranes were incubated with HRP-conjugated goat-anti-rabbit antibody (1:10,000; Jackson ImmunoResearch) for 1.5 h at room temperature. Bands were visualized with the SuperSignal West Femto Maximum Sensitivity Substrate Reagent Kit (ThermoFisher) using a ChemiDoc XRS + System (Bio-Rad).

Population-based analysis of the association of CHD3 variation with intelligence, educational qualification and intracranial volume/head circumference

Using WES data of 200,000 individuals from the UKB Exome Sequencing Consortium ([33; 34], and https://www.ukbiobank.ac.uk/media/cfulxh52/uk-biobank-exome-release-faq_v9-december-2020.pdf) we studied the association of *CHD3* missense and putative LoF variants with 'Fluid intelligence score' (data field ID 3533), 'Qualifications' (data field ID 6138) and 'Volume of EstimatedTotalIntraCranial' (data field ID 7054). Additionally, we used genome-wide association meta-analysis summary statistics of head circumference (N ≤ 18,881), and HC combined with intracranial volume (N ≤ 45,458) in child- and adulthood [38], and infant head circumference (N ≤ 10,768) [39] to calculate gene-level p-values reflecting the common variant associations of *CHD3* with these traits using MAGMA [69]. For detailed description of the methods, see Supplemental Notes 3.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1

Table 1: Summary of phenotypes seen in individuals with *CHD3* variants

	Probands with *de novo* variant[1]	Probands with inherited variant	Carrier parents
Development			
Developmental delay	100% (55/55)	100% (20/20)	18% (3/17)
Intellectual disability	98% (46/47)	79% (11/14)	22% (4/18)
borderline/borderline-mild	6% (3/47)	14% (2/14)	11% (2/18)
mild/mild-moderate	30% (14/47)	29% (4/14)	6% (1/18)
moderate/moderate-severe	36% (17/47)	14% (2/14)	0% (0/18)
severe	23% (11/47)	0% (0/14)	0% (0/18)
level unknown	2% (1/47)	21% (3/14)	6% (1/18)
Speech delay/disorder	100% (53/53)	100% (19/19)	25% (4/16)
Autism or autism-like features	35% (18/51)	56% (10/18)	19% (3/16)
Neurology			
Hypotonia	81% (39/48)	89% (16/18)	18% (2/11)
Macrocephaly	53% (28/53)	44% (8/18)	53% (8/15)
CNS abnormalities	50% (24/48)	62% (8/13)	25% (1/4)
Neonatal feeding problems	31% (10/32)	17% (3/18)	7% (1/15)
Facial dysmorphisms			
High/broad/prominent forehead	85% (28/33)	84% (16/19)	56% (9/16)
Thin upper lip	79% (15/19)	58% (11/19)	44% (7/16)
Widely spaced eyes	69% (35/51)	74% (14/19)	25% (4/16)
Broad nasal bridge	75% (15/20)	79% (15/19)	25% (4/16)
Full cheeks	58% (11/19)	68% (13/19)	7% (1/15)
Pointed chin	60% (12/20)	50% (9/18)	38% (6/16)
Deep-set eyes	55% (11/20)	44% (8/18)	47% (7/15)
Other			
Joint laxity (generalized and/or local)	36% (18/50)	42% (8/19)	31% (4/13)
Vision problems	72% (38/53)	26% (5/19)	57% (8/14)
Male genital abnormalities	32% (8/25)	25% (2/8)	0% (0/4)
Hernia (umbilical, inguinal, hiatal)	13% (6/48)	11% (2/19)	0% (0/13)

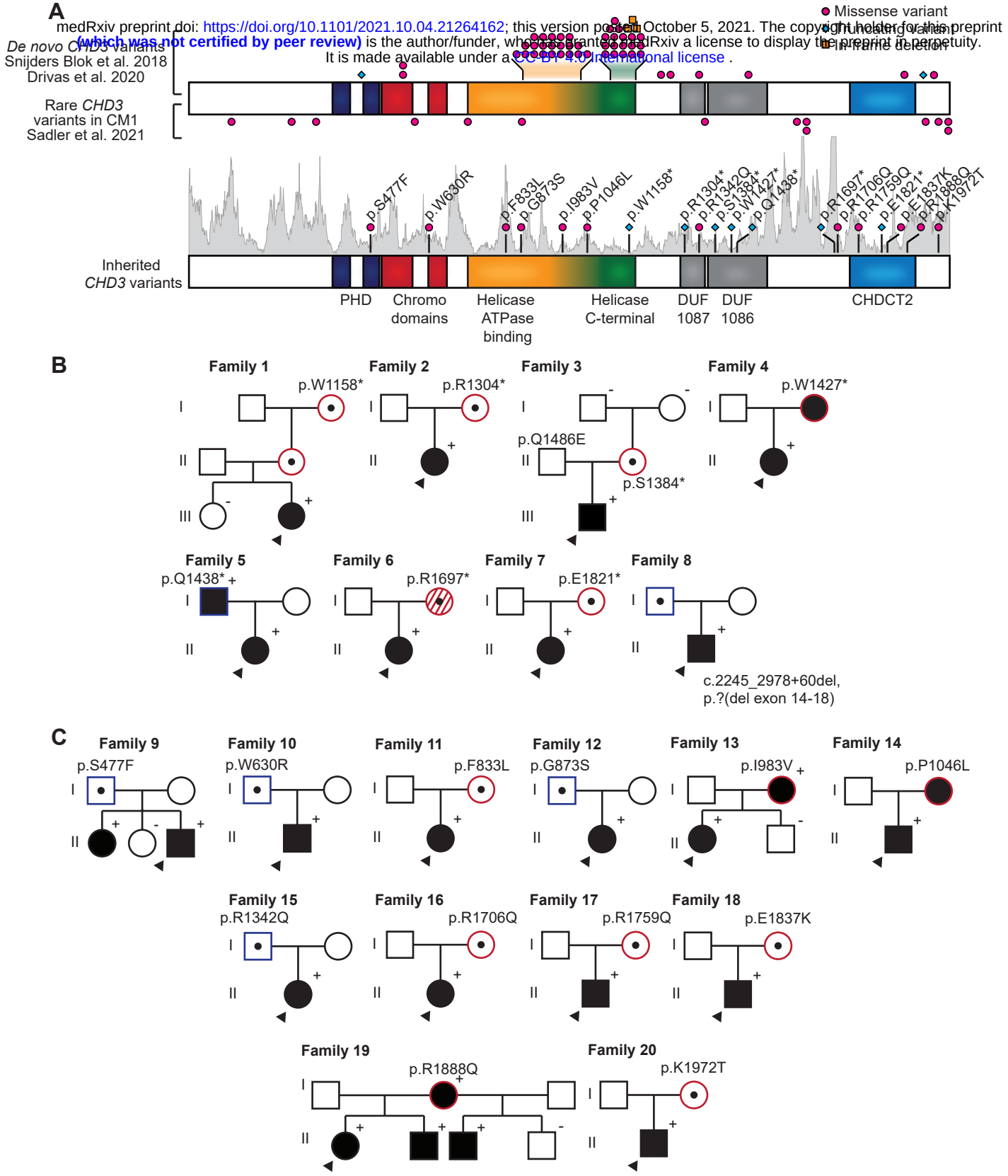
[1]Combined cases from Snijders Blok et al. 2018 and Drivas et al. 2020 (confirmed *de novo* only)

CNS: central nervous system

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Figures

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 1. Twenty families with inherited *CHD3* variants.

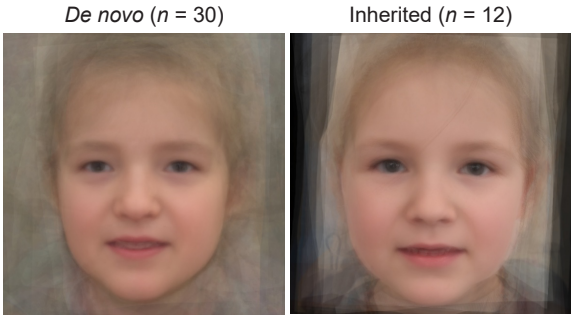
A) Schematic representation of the CHD3 protein (NM_001005273.2/NP_001005273.2), including functional domains, with PTVs labeled as cyan diamonds, in-frame deletions as orange squares, and missense variants as magenta circles. The intolerance landscape visualized using MetaDome [70] and computed based on single-nucleotide variants in the GnomAD database showing per amino acid position the missense over synonymous ratio, is shaded in gray. The top schematic shows the cases with *de novo CHD3* variants identified in NDD reported in [12] and [13] and rare variants associated with Chiari I malformations (CM1) reported in [18]. The bottom schematic presents cases with inherited *CHD3* variants described in this study. **B-C**) Pedigrees of families identified with inherited *CHD3* variants, with in (B) families with predicted LoF variants and in (C) families with missense variants. The arrow head indicates the proband, filled symbols represent affected individuals (defined as individuals with developmental delay and/or intellectual disability), open symbols with a central dot represent confirmed carriers without developmental delay/intellectual disability, '+' is used for a confirmed familial *CHD3* variant and '-' for confirmed non-carriers. Symbols with red contours represent female carriers, symbols with blue contours represent male carriers. Dashed symbol for family 6 represents mosaic state of the variant in the mother. In pedigrees, only genetically tested siblings of the proband are shown.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

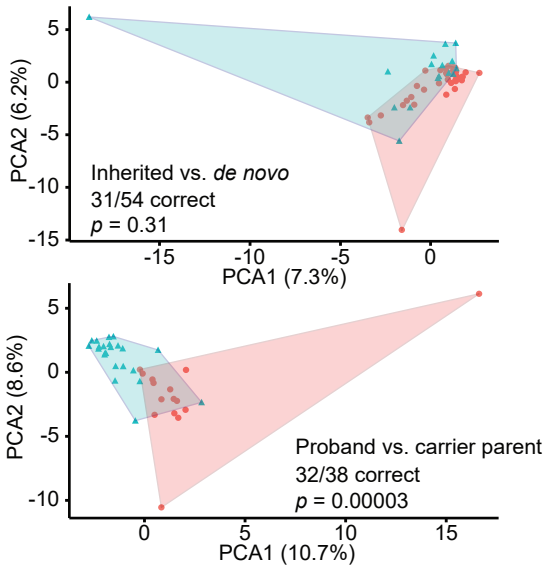
A

Facial photographs not available in this preprint

B



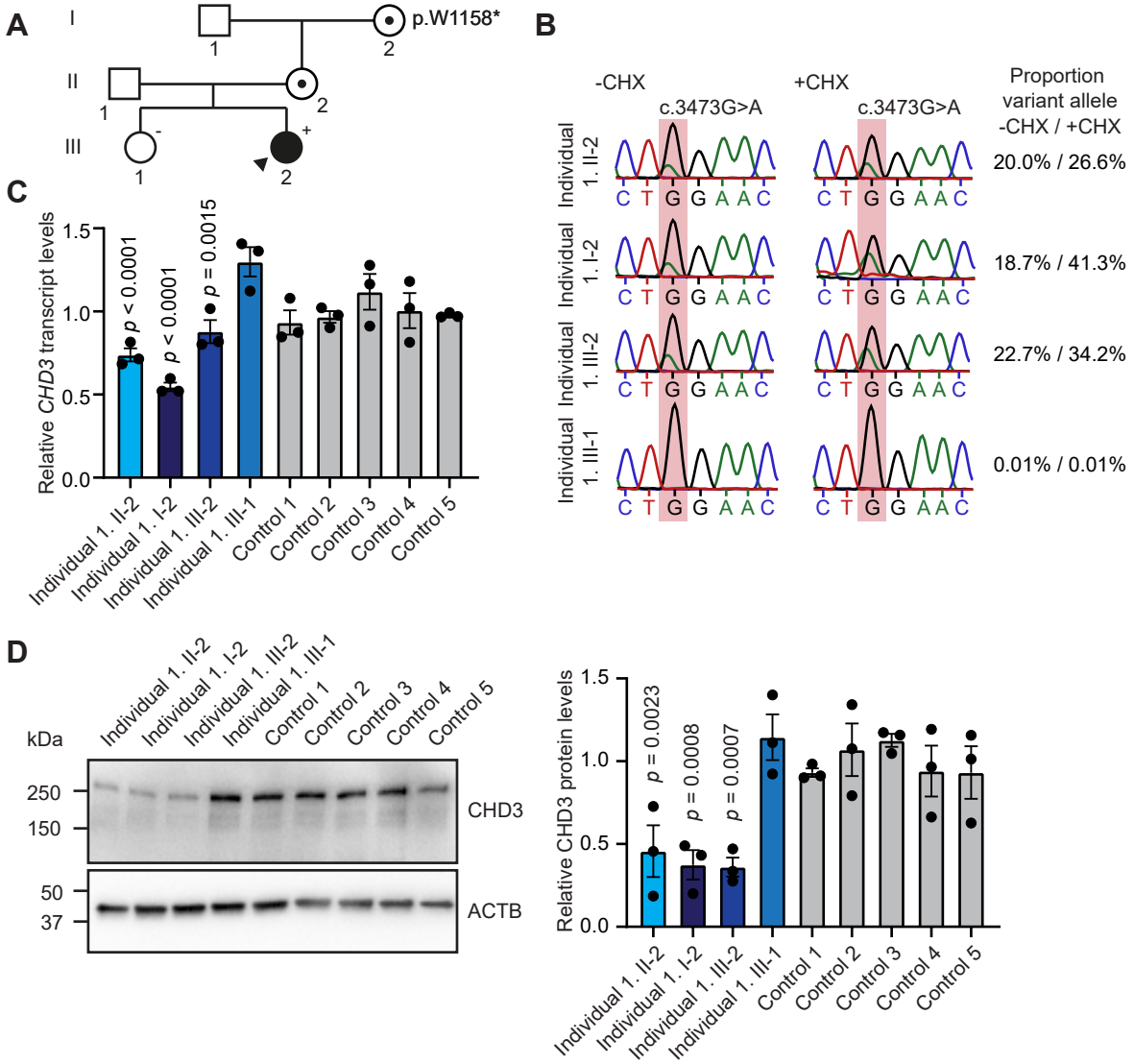
C



medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 2. Facial features and clinical evaluation of individuals with inherited *CHD3* variants. **A**) Facial photographs of individuals with inherited *CHD3* variants. Individuals demonstrate features also observed in individuals with *de novo CHD3* variants, including a squared appearance of the face, prominent forehead, widely spaced eyes, thin upper lip, pointed chin and deep-set eyes. These characteristics are also present in carrier parents. As observed previously, facial gestalt changes with age [13]. For example, a prominent nose is especially seen in adult individuals. For childhood pictures of carrier parents, see Figure S2. **B**) Computational average of facial photographs of 30 individuals with *de novo CHD3* variants (left) and 12 probands with inherited *CHD3* variants (right). **C**) Partitioning Around Medoids analyses of clustered HPO-standardized clinical data from 35 individuals with *de novo CHD3* variants, 19 affected probands with an inherited variant, and 19 carrier parents. The analyses do not show a significant distinction between the clusters of probands with *de novo* and probands with inherited variants (upper graph; $p = 0.31$). There is, however, a significant difference between the clusters of affected probands with inherited variants and carrier parents (bottom graph; $p = 0.00003$).

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure 3. Functional consequences of the CHD3 p.W1158* PTV in subject-derived lymphoblastoid cell lines. A) Pedigree of family identified with an inherited *CHD3* c.3473G>A/p.W1158* variant **B**) Sanger sequencing chromatographs of EBV-immortalized lymphoblastoid cell lines derived from members of family 1. Individuals I-2, II-2 and III-2 carried the c.3473G>A/p.W1158* variant and individual III-1 was a healthy non-carrier sibling. Cells were treated with (+CHX) or without cycloheximide (-CHX) to test for NMD. The mutated position is shaded in red. The transcript carrying the variant allele is present at lower levels than the wild-type allele, and increases after CHX treatment (proportion variant allele calculated as: peak area variant allele / (peak area variant + wildtype allele), showing that this variant is targeted by NMD. **C**) qPCR of EBV-immortalized lymphoblastoid cell lines of family 1 (shades of blue) and five unrelated controls (gray) for *CHD3* transcript levels (NM_001005273.2). Values are normalized to expression of *PPIA* and *TBP* and shown relative to unrelated controls. Bars represent the mean with individual data points plotted (n = 3; p-values compared to individual III-1 (healthy non-carrier sibling), one-way ANOVA and post hoc Bonferroni test). **D**) Left, a representative immunoblot of protein lysates prepared from lymphoblastoid cell lines for CHD3 (expected molecular weight: ~227 kDa). The blot was probed for ACTB, to ensure equal protein loading. Right, a graph showing the quantification of immunoblots with bars presenting the mean and individual data points plotted (n = 3; p-values compared to individual III-1 (healthy non-carrier sibling), one-way ANOVA and post hoc Bonferroni test). Controls are shaded in gray and samples from family 1 are shaded in blue. **C-D**) The cell lines carrying c.3473G>A/p.W1158* show lower *CHD3* transcript/protein levels compared to the control samples.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature 542, 433-438.

2. Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al. (2010). A de novo paradigm for mental retardation. Nat Genet 42, 1109-1112.

3. de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 367, 1921-1929.

4. Akawi, N., McRae, J., Ansari, M., Balasubramanian, M., Blyth, M., Brady, A.F., Clayton, S., Cole, T., Deshpande, C., Fitzgerald, T.W., et al. (2015). Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nat Genet 47, 1363-1369.

5. Martin, H.C., Jones, W.D., McIntyre, R., Sanchez-Andrade, G., Sanderson, M., Stephenson, J.D., Jones, C.P., Handsaker, J., Gallone, G., Bruntraeger, M., et al. (2018). Quantifying the contribution of recessive coding variation to developmental disorders. Science 362, 1161-1164.

6. Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y., Gallone, G., Lelieveld, S.H., Martin, H.C., McRae, J.F., et al. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature 586, 757-762.

7. Perenthaler, E., Yousefi, S., Niggl, E., and Barakat, T.S. (2019). Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. Front Cell Neurosci 13, 352.

8. Kosmicki, J.A., Samocha, K.E., Howrigan, D.P., Sanders, S.J., Slowikowski, K., Lek, M., Karczewski, K.J., Cutler, D.J., Devlin, B., Roeder, K., et al. (2017). Refining the role of

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet 49, 504-510.

9. Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., Guigo, R., Iossifov, I., Vasileva, A., and Lappalainen, T. (2018). Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. Nat Genet 50, 1327-1334.

10. Chen, R., Shi, L., Hakenberg, J., Naughton, B., Sklar, P., Zhang, J., Zhou, H., Tian, L., Prakash, O., Lemire, M., et al. (2016). Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nat Biotechnol 34, 531-538.

11. Torchy, M.P., Hamiche, A., and Klaholz, B.P. (2015). Structure and function insights into the NuRD chromatin remodeling complex. Cell Mol Life Sci 72, 2491-2507.

12. Snijders Blok, L., Rousseau, J., Twist, J., Ehresmann, S., Takaku, M., Venselaar, H., Rodan, L.H., Nowak, C.B., Douglas, J., Swoboda, K.J., et al. (2018). CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. Nat Commun 9, 4619.

13. Drivas, T.G., Li, D., Nair, D., Alaimo, J.T., Alders, M., Altmuller, J., Barakat, T.S., Bebin, E.M., Bertsch, N.L., Blackburn, P.R., et al. (2020). A second cohort of CHD3 patients expands the molecular mechanisms known to cause Snijders Blok-Campeau syndrome. Eur J Hum Genet 28, 1422-1431.

14. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434-443.

15. Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gourdine, J.P., Gargano, M., Harris, N.L., Matentzoglu, N., McMurry, J.A., et al. (2019). Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res 47, D1018-d1027.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

16. Kaufmann, L., and Rousseeuw, P. (1987). Clustering by Means of Medoids. Data Analysis based on the L1-Norm and Related Methods, 405-416.

17. Wright, C.F., West, B., Tuke, M., Jones, S.E., Patel, K., Laver, T.W., Beaumont, R.N., Tyrrell, J., Wood, A.R., Frayling, T.M., et al. (2019). Assessing the Pathogenicity, Penetrance, and Expressivity of Putative Disease-Causing Variants in a Population Setting. Am J Hum Genet 104, 275-286.

18. Sadler, B., Wilborn, J., Antunes, L., Kuensting, T., Hale, A.T., Gannon, S.R., McCall, K., Cruchaga, C., Harms, M., Voisin, N., et al. (2021). Rare and de novo coding variants in chromodomain genes in Chiari I malformation. Am J Hum Genet 108, 100-114.

19. Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. Pediatr Res 65, 591-598.

20. Jacquemont, S., Coe, B.P., Hersch, M., Duyzend, M.H., Krumm, N., Bergmann, S., Beckmann, J.S., Rosenfeld, J.A., and Eichler, E.E. (2014). A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. Am J Hum Genet 94, 415-425.

21. Turner, T.N., Wilfert, A.B., Bakken, T.E., Bernier, R.A., Pepper, M.R., Zhang, Z., Torene, R.I., Retterer, K., and Eichler, E.E. (2019). Sex-Based Analysis of De Novo Variants in Neurodevelopmental Disorders. Am J Hum Genet 105, 1274-1285.

22. Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron 74, 285-299.

23. Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 70, 886-897.

24. Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Multiple recurrent

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70, 863-885.

25. Duyzend, M.H., Nuttle, X., Coe, B.P., Baker, C., Nickerson, D.A., Bernier, R., and Eichler, E.E. (2016). Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am J Hum Genet 98, 45-57.
26. Krumm, N., Turner, T.N., Baker, C., Vives, L., Mohajeri, K., Witherspoon, K., Raja, A., Coe, B.P., Stessman, H.A., He, Z.X., et al. (2015). Excess of rare, inherited truncating mutations in autism. Nat Genet 47, 582-588.
27. Knight, J.C. (2004). Allele-specific gene expression uncovered. Trends Genet 20, 113-116.
28. Porter, R.S., Jaamour, F., and Iwase, S. (2018). Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. Mol Cell Neurosci 87, 35-45.
29. Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet 132, 1077-1130.
30. Girirajan, S., Rosenfeld, J.A., Coe, B.P., Parikh, S., Friedman, N., Goldstein, A., Filipink, R.A., McConnell, J.S., Angle, B., Meschino, W.S., et al. (2012). Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med 367, 1321-1331.
31. Albers, C.A., Paul, D.S., Schulze, H., Freson, K., Stephens, J.C., Smethurst, P.A., Jolley, J.D., Cvejic, A., Kostadima, M., Bertone, P., et al. (2012). Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet 44, 435-439, s431-432.
32. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12, e1001779.

33. Szustakowski, J.D., Balasubramanian, S., Sasson, A., Khalid, S., Bronson, P.G., Kvikstad, E., Wong, E., Liu, D., Davis, J.W., Haefliger, C., et al. (2020). Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank. medRxiv, 2020.2011.2002.20222232.

34. Van Hout, C.V., Tachmazidou, I., Backman, J.D., Hoffman, J.D., Liu, D., Pandey, A.K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., et al. (2020). Exome sequencing and characterization of 49,960 individuals in the UK Biobank. Nature.

35. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203-209.

36. Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.R., Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., et al. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage 166, 400-424.

37. Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 19, 1523-1536.

38. Haworth, S., Shapland, C.Y., Hayward, C., Prins, B.P., Felix, J.F., Medina-Gomez, C., Rivadeneira, F., Wang, C., Ahluwalia, T.S., Vrijheid, M., et al. (2019). Low-frequency variation in TP53 has large effects on head circumference and intracranial volume. Nature communications 10, 357.

39. Taal, H.R., Pourcain, B.S., Thiering, E., Das, S., Mook-Kanamori, D.O., Warrington, N.M., Kaakinen, M., Kreiner-Møller, E., Bradfield, J.P., Freathy, R.M., et al. (2012). Common

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

variants at 12q15 and 12q24 are associated with infant head circumference. Nat Genet 44, 532-538.

40. Maegawa, G.H., Leite, J.C., Félix, T.M., da Silveira, H.L., and da Silveira, H.E. (2004). Clinical variability in KBG syndrome: report of three unrelated families. Am J Med Genet A 131, 150-154.

41. Low, K., Ashraf, T., Canham, N., Clayton-Smith, J., Deshpande, C., Donaldson, A., Fisher, R., Flinter, F., Foulds, N., Fryer, A., et al. (2016). Clinical and genetic aspects of KBG syndrome. Am J Med Genet A 170, 2835-2846.

42. Katsanis, N. (2016). The continuum of causality in human genetic disorders. Genome Biol 17, 233.

43. Sadedin, S.P., Dashnow, H., James, P.A., Bahlo, M., Bauer, D.C., Lonie, A., Lunke, S., Macciocca, I., Ross, J.P., Siemering, K.R., et al. (2015). Cpipe: a shared variant detection pipeline designed for diagnostic settings. Genome Med 7, 68.

44. Bowling, K.M., Thompson, M.L., Amaral, M.D., Finnila, C.R., Hiatt, S.M., Engel, K.L., Cochran, J.N., Brothers, K.B., East, K.M., Gray, D.E., et al. (2017). Genomic diagnosis for children with intellectual disability and/or developmental delay. Genome Med 9, 43-43.

45. Corsten-Janssen, N., Bouman, K., Diphoorn, J.C.D., Scheper, A.J., Kinds, R., El Mecky, J., Breet, H., Verheij, J., Suijkerbuijk, R., Duin, L.K., et al. (2020). A prospective study on rapid exome sequencing as a diagnostic test for multiple congenital anomalies on fetal ultrasound. Prenat Diagn 40, 1300-1309.

46. Brunet, T., Jech, R., Brugger, M., Kovacs, R., Alhaddad, B., Leszinski, G., Riedhammer, K.M., Westphal, D.S., Mahle, I., Mayerhanser, K., et al. (2021). De novo variants in neurodevelopmental disorders-experiences from a tertiary care center. Clin Genet 100, 14-28.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

47. Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. Genet Med 18, 696-704.

48. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P., de Vries, B.B., Willemsen, M.H., Kleefstra, T., Löhner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat Neurosci 19, 1194-1196.

49. Deciphering Developmental Disorders Study. (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223-228.

50. den Dunnen, J.T., and Antonarakis, S.E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15, 7-12.

51. Plon, S.E., Eccles, D.M., Easton, D., Foulkes, W.D., Genuardi, M., Greenblatt, M.S., Hogervorst, F.B., Hoogerbrugge, N., Spurdle, A.B., and Tavtigian, S.V. (2008). Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 29, 1282-1291.

52. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). 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.

53. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47, D886-d894.

54. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

55. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat Protoc 11, 1-9.

56. Dingemans, A.J.M., Stremmelaar, D.E., van der Donk, R., Vissers, L., Koolen, D.A., Rump, P., Hehir-Kwa, J.Y., and de Vries, B.B.A. (2021). Quantitative facial phenotyping for Koolen-de Vries and 22q11.2 deletion syndrome. Eur J Hum Genet.

57. van der Donk, R., Jansen, S., Schuurs-Hoeijmakers, J.H.M., Koolen, D.A., Goltstein, L., Hoischen, A., Brunner, H.G., Kemmeren, P., Nellåker, C., Vissers, L., et al. (2019). Next-generation phenotyping using computer vision algorithms in rare genomic neurodevelopmental disorders. Genet Med 21, 1719-1725.

58. Reijnders, M.R.F., Miller, K.A., Alvi, M., Goos, J.A.C., Lees, M.M., de Burca, A., Henderson, A., Kraus, A., Mikat, B., de Vries, B.B.A., et al. (2018). De Novo and Inherited Loss-of-Function Variants in TLK2: Clinical and Genotype-Phenotype Evaluation of a Distinct Neurodevelopmental Disorder. Am J Hum Genet 102, 1195-1203.

59. den Hoed, J., de Boer, E., Voisin, N., Dingemans, A.J.M., Guex, N., Wiel, L., Nellaker, C., Amudhavalli, S.M., Banka, S., Bena, F.S., et al. (2021). Mutation-specific pathophysiological mechanisms define different neurodevelopmental disorders associated with SATB1 dysfunction. Am J Hum Genet 108, 346-356.

60. Deng, Y., Gao, L., Wang, B., and Guo, X. (2015). HPOSim: an R package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. PLoS One 10, e0115692.

61. Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S., and Chen, C.F. (2007). A new method to measure the semantic similarity of GO terms. Bioinformatics 23, 1274-1281.

62. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. J Mol Graph 8, 52-56, 29.

63. Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. Proteins 47, 393-402.

medRxiv preprint doi: https://doi.org/10.1101/2021.10.04.21264162; this version posted October 5, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

64. Farnung, L., Ochmann, M., and Cramer, P. (2020). Nucleosome-CHD4 chromatin remodeler structure maps human disease mutations. Elife 9.

65. Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., and Mackay, J.P. (2011). Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. J Biol Chem 286, 11779-11791.

66. Deriziotis, P., Graham, S.A., Estruch, S.B., and Fisher, S.E. (2014). Investigating protein-protein interactions in live cells using bioluminescence resonance energy transfer. J Vis Exp.

67. Koulouras, G., Panagopoulos, A., Rapsomaniki, M.A., Giakoumakis, N.N., Taraviras, S., and Lygerou, Z. (2018). EasyFRAP-web: a web-based tool for the analysis of fluorescence recovery after photobleaching data. Nucleic Acids Res 46, W467-w472.

68. Kohze, R., Dieteren, C.E.J., Koopman, W.J.H., Brock, R., and Schmidt, S. (2017). Frapbot: An open-source application for FRAP data. Cytometry A 91, 810-814.

69. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 11, e1004219.

70. Wiel, L., Baakman, C., Gilissen, D., Veltman, J.A., Vriend, G., and Gilissen, C. (2019). MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Hum Mutat 40, 1030-1038.