De novo heterozygous variants in *POLR2A* cause a neurodevelopmental syndrome with profound infantile-onset hypotonia

Hanneke A. Haijes,[1,2,34] Maria J.E. Koster,[3,34] Holger Rehmann,[4,5] Dong Li,[6] Hakon Hakonarson,[6,7,8] Gerarda Cappuccio,[9,10] Miroslava Hancarova,[11] Daphne Lehalle,[12] Willie Reardon,[13] G. Bradley Schaefer,[14] Anna Lehman,[15] Ingrid M.B.H. van de Laar,[16] Coranne D. Tesselaar,[17] Clesson Turner,[18] Alice Goldenberg,[19] Sophie Patrier,[20] Julien Thevenon,[21] Michele Pinelli,[9,10] Nicola Brunetti-Pierri,[9,10] Darina Prchalová,[11] Markéta Havlovicová,[11] Markéta Vlckova,[11] Zdeněk Sedláček,[11] Elena Lopez,[15] Vassilis Ragoussis,[22] Alistair T. Pagnamenta,[22] Usha Kini,[23] Harmjan R. Vos,[5] Robert M. van Es,[5] Richard F.M.A. van Schaik,[5] Ton A.J. van Essen,[24,36] Maria Kibaek,[25] Jenny C. Taylor,[22] Jennifer Sullivan,[26] Vandana Shashi,[26] Slave Petrovski,[26,27,28] Christina Fagerberg,[29] Donna M. Martin,[30] Koen L.I. van Gassen,[2] Rolph Pfundt,[31] Marni J. Falk,[7,32] Elizabeth M. McCormick,[32] H.T. Marc Timmers,[3,33,35] Peter M. van Hasselt[1,35]*

1. Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, 3584 EA Utrecht, the Netherlands
2. Department of Biomedical Genetics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht University, 3584 EA Utrecht, the Netherlands
3. Regenerative Medicine Center and Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, 3584 CT Utrecht, the Netherlands
4. Expertise Center for Structural Biology, University Medical Center Utrecht, Utrecht University, 3584 CT Utrecht, the Netherlands
5. Molecular Cancer Research, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Oncode Institute, 3584 CT Utrecht, the Netherlands
6. Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, United States of America
7. Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, United States of America

8. Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States of America
9. Department of Translational Medicine, Federico II University, 80126, Naples, Italy
10. Telethon Institute of Genetics and Medicine, Pozzuoli, 80126, Naples, Italy
11. Department of Biology and Medical Genetics, Charles University Second Faculty of Medicine and University Hospital Motol, 150 06 Prague, Czech Republic
12. Department of Genetics, Centre Hospitalier Universitaire de Dijon, 21000 Dijon, France
13. Department of Clinical and Medical Genetics, Our Lady's Hospital for Sick Children, D12 N512, Dublin, Ireland
14. Department of Pediatrics, Section of Genetics and Metabolism, University of Arkansas for Medical Sciences, Little Rock, Arkansas, AR 72223, United States of America
15. Department of Medical Genetics, BC Children's Hospital Research Institute, University of British Columbia, BC V6H 3N1, Vancouver, Canada
16. Department of Clinical Genetics, Erasmus Medical University Center Rotterdam, 3000 CA Rotterdam, the Netherlands
17. Department of Pediatrics, Amphia Hospital Breda, 4818 CK Breda, the Netherlands
18. Department of Clinical Genetics and Pediatrics, Walter Reed National Military Medical Center, Bethesda, Maryland, MD 20814, United States of America
19. Department of Genetics, Rouen University Hospital, Centre de référence Anomalies du développement, Normandy Centre for Genomic and Personalized Medicine, 76000 Rouen, France
20. Department of Pathology, Rouen University Hospital, Centre de référence Anomalies du développement, 76000 Rouen, France
21. Department of Genetics and Reproduction, Centre Hospitalier Universitaire de Grenoble, 38700 Grenoble, France
22. Department of Genomic Medicine, Oxford Biomedical Research Centre and Welcome Centre for Human Genetics, University of Oxford, OX3 7BN Oxford, United Kingdom
23. Department of Genomic Medicine, Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS Foundation Trust, OX3 7LE Oxford, United Kingdom
24. Department of Clinical Genetics, University Medical Center Groningen, 9713 GZ Groningen, the Netherlands

25. H.C. Andersen Children Hospital, Odense University Hospital, 5000 Odense, Denmark
26. Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina, NC 27710, United States of America
27. AstraZeneca Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, CB4 0WG Cambridge, United Kingdom
28. Department of Medicine, The University of Melbourne, VIC 3010, Melbourne, Australia
29. Department of Clinical Genetics, Odense University Hospital, 5000 Odense, Denmark
30. Departments of Pediatrics and Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, MI 48109, United States of America
31. Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center Nijmegen, 6525 HR Nijmegen, the Netherlands
32. Mitochondrial Medicine Frontier Program, Division of Human Genetics, The Children's Hospital of Philadelphia, PA 19104, Philadelphia, United States of America
33. Department of Urology, Medical Center-University of Freiburg, Freiburg and German Cancer Consortium (DKTK) standort Freiburg and German Cancer Research Center (DKFZ), 79106 Heidelberg, Germany
34. These authors contributed equally to this work
35. These authors contributed equally to this work
36. This author passed away during the preparation of this manuscript

* *Corresponding author:*

P.M. van Hasselt

p.vanhasselt@umcutrecht.nl

Wilhelmina Children's Hospital

University Medical Center Utrecht

Lundlaan 6

3584 EA Utrecht

ABSTRACT

The RNA polymerase II complex (pol II) is responsible for transcription of all ~21,000 human protein-encoding genes. Here, we describe sixteen individuals harboring *de novo* heterozygous variants in *POLR2A,* encoding RPB1, the largest subunit of pol II. An iterative approach, combining structural evaluation and mass spectrometry analyses, the use of *S. cerevisiae* as a model system and the assessment of cell viability in HeLa cells, allowed us to classify eleven variants as probably disease-causing and four variants as possibly disease-causing. Of one variant the significance remains unresolved. By quantification of phenotypic severity, we could distinguish mild and severe phenotypic consequences of the disease-causing variants. Missense variants expected to exert only mild structural effects, led to malfunctioning pol II enzyme, thereby inducing a dominant negative effect on gene transcription. Intriguingly, individuals carrying these variants presented with a severe phenotype dominated by profound infantile onset hypotonia and developmental delay. Conversely, individuals carrying variants expected to result in complete loss-of-function, thus reduced levels of RPB1 but functional pol II, exhibited the mildest phenotypes. We conclude that subtle variants central in functionally important domains of *POLR2A* cause a neurodevelopmental syndrome characterized by profound infantile onset hypotonia and developmental delay through a dominant-negative effect on pol II-mediated transcription of DNA.

INTRODUCTION

Human *POLR2A* (MIM #180660) encodes the highly-conserved RPB1 protein, which is the largest of twelve subunits of the essential RNA polymerase II (pol II) enzyme[1,2]. This protein complex is responsible for transcription of all protein-encoding genes as well as several long and short non-coding RNA genes[3]. Due to its central role in gene expression, pol II and the regulation of its activity have been studied in depth since its discovery fifty years ago[4]. Biochemical and structural studies have shown that transcription initiation by pol II requires a set of basal (or general) transcription factors. After assembly of the pol II initiation complex at the promoter region, these factors are exchanged for transcription elongation factors including TFIIS, DSIF and P-TEFb to allow processive RNA synthesis by pol II[5]. Detailed structural analysis of pol II of *S. cerevisiae*[6] revealed different aspects of RNA synthesis, including initiation[7-10], nucleotide binding[11], chain elongation[12-16], error correction and back tracking[17-20]. For example, during the elongation cycle the incorporation of the incoming nucleotide involves movement of the so-called trigger loop of RPB1, which induces the forward movement of pol II over the DNA template strand coincident with nucleotide selection. *In vivo* dynamics of pol II during initiation and promotor pausing were studied in GFP-RPB1 knock-in cells, demonstrating that the continuous release and reinitiation of promotor-bound pol II is important for transcriptional regulation[21]. Mutational studies focusing on *Saccharomyces cerevisiae (S. cerevisiae)* and human RPB1 indicated that residues in the trigger loop region are controlling the rate of RNA synthesis[22,23]. Experiments in *S. cerevisiae* indicated that the phenotypic consequences of *rpb1* variants in yeast may be concealed by transcript buffering[24] and may only become apparent under conditions of environmental or nutrient stress[25]. Experiments in *Arabidopsis* mutants that encode truncated RPB1 with a different sizes of shortened C-terminal domain (CTD) show disturbed cell cycling control, demonstrating its importance in transcriptional regulation of cell cycle genes[26]. Despite its critical role in transcription, *POLR2A* has not been implicated in human disease thus far.

Here, we describe sixteen individuals harboring *de novo* heterozygous genetic variants in *POLR2A*. Acknowledging that, while occurrence of *de novo* variants provides an important clue towards variant pathogenicity[27], this is not decisive[28], since large-scale sequencing of individuals without severe pediatric disease has revealed that protein changing genetic variants are present in all ~21,000 protein-encoding genes[29]. Similarly, as individuals carrying a variant in a certain gene are increasingly found through gene-matching initiatives[30], phenotypic overlap may merely reflect general characteristics

of individuals undergoing whole-exome or whole-genome sequencing (WES, WGS). In addition, even the presence nor the absence of evidence from functional analyses can be used to prove or refute pathogenicity. This transition of genetics from a dichotomous field to an exciting but puzzling world full of shades of grey, also offers new avenues. Deep phenotyping analyses may help to accurately expose the extent of the phenotypic overlap to support pathogenicity, as do severity metrics from population genetics. We here apply an iterative process wherein we combine several lines of evidence, including a detailed phenotypic analysis, assessment of variant severity metrics, structural biology and functional analyses in both *S. cerevisiae* and HeLa cells, to assess pathogenicity of the sixteen identified variants in *POLR2A* in detail.

MATERIALS AND METHODS

Phenotypes of affected individuals

The cohort was assembled using the GeneMatcher initiative[30]. Inclusion criteria were detection of a *de novo* heterozygous variant in *POLR2A* (NM_000937.4), as confirmed either by clinical or research trio WES or WGS, or when the presence of the heterozygous variant in both parents could be excluded. Physicians first provided a detailed phenotypic description of the individual. In a second round, questions regarding specific clinical features of the individual were asked to further define the clinical overlap among individuals. Additionally, a thorough evaluation of the individuals' attained milestones using items of the Denver Scale[31] was performed by doctors and parents together. These data were used to assess the rate of development over time and developmental domain specific delays. To assess these focus points, Z-scores were calculated as follows: the Denver Scales provides information on normal development for p50 and p90 in months. The ninetieth percentile corresponds to 1.28 standard deviation (SD). SD was calculated by (p90 – p50)/1.28. Z-scores were calculated as (x – p50)/SD, wherein x is the month the individual attained the developmental milestone. Statistical analysis was performed in R programming language, using $\log_{10}$ transformation. All procedures followed were in accordance with the Helsinki Declaration of 1975, as revised in 2000. The legal guardians for all individuals agreed to participation of their child in the study and signed the appropriate consent forms, in agreement with institutional and national legislation of the different centers. Permission for publication of pictures was given separately.

Variant severity metrics

pLI scores for loss of function variation, and Z-scores for synonymous and missense variants, were retrieved from The Genome Aggregation Database (gnomAD) for *POLR2A* as well as for the genes encoding other pol II subunits. High pLI scores are indicative of intolerance of loss of function variation, and Z-scores are indicative of intolerance of synonymous or missense variation[29]. Additionally, ratios of observed and expected variants were retrieved, as a reflection of the degree to which variants within a gene are underreported, indicative of a survival disadvantage.

The CADD score is a tool for scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome[32]. It integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated variants. CADD scores were obtained for all missense variants in gnomAD as well as for the missense and nonsense variants of the included individuals. A CADD score > 20 suggests pathogenicity, as it indicates that the variant is predicted to be among the 1% most deleterious substitutions that can be done to the human genome.

The distribution of genetic variants throughout *POLR2A* was assessed, based on all *POLR2A* variants from gnomAD. The interval length (number of amino acid positions) between two sequential missense variants was calculated. A larger interval size signifies a stretch devoid of missense variants and was considered a possible index of pathogenicity for variants found within such an interval. The mean and SD were calculated. The "desert Z-score", indicating to which degree stretches within the gene are devoid of missense variants, was calculated by the formula (stretch length – mean stretch length)/ SD of stretch length.

Structural evaluation of POLR2A variants in pol II

Pol II of *S. cerevisiae* is extensively studied by x-ray crystallography and cryo-electron microscopy[7-11,13-14,16-20]. Sequence conservation between human *POLR2A* and *S. cerevisiae rpb1* is high, and thus the available structures can serve as a model for the human protein. To evaluate the putative consequences of *POLR2A* variants, the structures of RNA polymerase from Protein Data Bank entries 1i6h, 1r5u, 1r9s, 1twf, 1y1v, 1y1w, 1y1y, 2e2h, 2nvq, 2nvz, 3gtj, 3gtm, 3how, 3hoy, 3po2, 3po3, 4a3l, 4a93, 4bbr, 4gwp, 4v1m, 4v1n, 4v1o, 5sva, 5xog, and 5xon were superimposed by overlaying the coordinates for *rpb1*

with the algorithm provided by BRAGI[33]. Variants were mapped onto the structures and evaluated for their putative impact on local protein folding and enzymatic activity. Figures were prepared in the program Molscript[34] and raster3D[35].

Functional evaluation of variants in pol II: POLR2A variants expressed in S. cerevisiae

Six *POLR2A* variants were matched through GeneMatcher at the time the experiment in *S. cerevisiae* was initiated, therefore six *POLR2A* variants were functionally evaluated in *S. cerevisiae*. Plate phenotyping of *rpb1* mutants was performed as described before[22]. All *S. cerevisiae* strains used in this study are listed in Table S1. They were derived from CKY283, CKY718 and CKY721 (kind gifts of C. Kaplan). Shortly, to generate mutant *rpb1*, point mutations in *S. cerevisiae rpb1* were introduced by PCR mutagenesis into the *CEN LEU2* plasmid pRS315H3alt-RPB1* *Xma*I 1122-1123 T69 corrected (a kind gift of C. Kaplan). The *CEN LEU2* plasmids, containing mutant *rpb1,* were transformed into an appropriate Leu^- strain with corresponding endogenous *rpb1* deletion and complemented with a *CEN URA3* WT *rpb1* subunit gene, expressing wildtype RPB1. Leu^+ transformants were patched on solid medium lacking leucine and replica-plated to medium lacking leucine but containing 5-fluoroorotic acid (Thermo Scientific) to select against cells maintaining RPB1 wildtype *URA3* plasmids. Δ*sub1* or *Δdst1 + rpb1* mutant strains for direct testing of double mutant phenotypes were constructed based on CKY721 or CKY718 respectively, and analyzed in the same fashion as *rpb1* single mutants.

Assays were performed in biological triplicates in three different backgrounds: wildtype background, absence of TFIIS (Δ*dst1* background) and absence of the SUB1 transcription elongation factor (Δ*sub1* background). Transcriptional activity and genetic interactions were determined. p.Asn531Ser (p.Asn517Ser$_{yeast}$) and p.Thr736Met (p.Ser713Met$_{yeast}$) in wildtype background were performed in sextuplicates to increase consistency of the results. p.Ala315Asp (p.Ala301Asp$_{yeast}$) was used as a positive control for reduced transcriptional activity[36,37] and p.Glu1230Lys$_{yeast}$ was used as a positive control for reduced genetic interaction with TFIIS[38].

For spot assays, overnight yeast peptone dextrose (YPD) cultures from single colonies grown at 30°C were diluted to an OD_{600} of 0.15. Five-fold serial dilutions were prepared and spotted on indicated plates and grown for 3 to 6 days at 30°C or 37°C when indicated. YPD medium contained yeast extract (1% w/v final, (BD)), peptone (2% w/v final, (BD)), dextrose

(2% w/v final) and bacto agar (2% w/v final, (BD)). Alternative carbon source YP media were YP raffinose (2% final w/v) and YP raffinose (2% w/v final) plus galactose (1% w/v final). Synthetic complete medium lacking leucine (SC-Leu) contained: 2gr/l drop out mix minus leucine (US Biological) and 6.71 gr/l Yeast Nitrogen Base without AA, Carbohydrate & w/AS (YNB) (US Biological) with 2% dextrose. Mycophenolic acid (MPA, (Sigma-Aldrich)) was added to SC-Leu at 20 µg/ml final concentration from a 10 mg/ml stock in ethanol.

Functional evaluation of variants in pol II: POLR2A variants expressed in HeLa cells

Nine *POLR2A* variants were matched through GeneMatcher at the time the experiment in HeLa cells was initiated, therefore nine *POLR2A* variants could be functionally evaluated in HeLa cells. The open reading frame of human RPB1 was amplified by PCR using a plasmid expressing RPB1 fusion with a B10 epitope, EGFP, hRPB1 and six His residues[36] (a kind gift of P. Cook) and introduced into the pDONR201 cloning vector. The *POLR2A* coding sequence contained a point mutation (AAC -> GAC), which resulted in the replacement of asparagine 792 by aspartate and resistance to α-amanitin[39]. The observed point mutations in the *POLR2A* included individuals were introduced using the Quickchange protocol (Stratagene) and verified by DNA sequencing. p.Lys812* was used as a representative of p.Gln700* and p.Gln735*. These two variants could not be designed, since amino acid residues Gln700 and Gln735 are localized in front of the built-in resistance to α-amanitin (residue 792). p.Lys812* is expected to result in a similar truncated version of the protein as p.Gln700* and p.Gln735*. All RPB1 mutant proteins were tagged by GFP at the N-terminus. Stable doxycycline inducible cell lines were created by transfecting pCDNA5/FRQT/TO and pOG44 into HeLa FRT cells carrying the TET repressor using polyethyleneimine (PEI) followed by antibiotic selection. HeLa cells were maintained in DMEM containing 4.5 g/l glucose (Gibco) supplemented with 10% v/v heat inactivated Fetal Bovine Serum (FBS) (Sigma-Aldrich) and 10 mM L-Glutamine (Sigma-Aldrich) under blasticidin (5 µg/ml) (InvivoGen) and hygromycin B (400 µg/ml) selection (Roche Diagnostics).

Expression of GFP-tagged RPB1 was induced by treatment with 1 µg/ml doxycycline for 24 hours. Cells were spun down for 5 minutes at 400g and washed with PBS. Next, cells were lysed in whole cell lysate (WCL) buffer (50 mM Tris-HCl pH 8.0, 420 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, 0.1% NP40, 0.5 mM DTT, containing protease inhibitor cocktail) and lysates were subjected to centrifugation at 400g for 10 minutes. The supernatant was harvested and stored at -80˚C.

For the MTT (3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazoliumbromide, Sigma-Aldrich) assay, doxycycline-inducible GFP-RPB1$^{amanitin\ resistant}$ expressing mutant HeLa FRT cells were cultured 3 to 5 days in the presence of 1 µg/ml doxycycline to induce expression of the mutants. Cells were seeded at 3,000 cells per well in 100 µl normal DMEM medium in the presence of doxycycline in 96-well plates. After a few hours, medium was replaced by DMEM with doxycycline and when indicated, 2.5 µg/ml α-amanitin (Boehringer Mannheim GmbH) was added. After 72 hours of incubation, MTT was added to a final concentration of 0.5 mg/ml[40]. Cells were incubated for 4 hours under normal culturing conditions, followed by careful removal of medium and 100 µl DMSO was added to dissolve the formazan crystals. The absorbance was measured at 570- and 630-nm wavelengths. Relative growth was calculated using equation $(A_{570,sample} - A_{630,sample})/ (A_{570,untreated} - A_{630,untreated})$.

GST pull-down

A GST-TFIIS recombinant expression construct[41] was transformed in BL21DE3, single colonies were picked, grown in LB + ampicillin and induced for 3 hours with 1 mM IPTG when cells reached an OD_{600} of 0.6. Cells were pelleted by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 7.0, 300 mM KCl, 2 mM EDTA, 20% Sucrose, 0.1% Triton X-100, 0.5 mM PMSF, 1 mM DTT) containing 50 µl Lysozyme (25 mg/ml) and incubated for 10 minutes on ice. The suspension was freeze-thawed 3 times and sonicated 3 times for 20 seconds. Cells were spun down at 25,000 rpm for 45 minutes at 4˚C and the supernatant was harvested and stored at -80˚C.

Glutathione-Agarose (GA) beads (50% slurry) were washed three times with binding buffer (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM EDTA, 10% glycerol, 0.1 mM $ZnCl_2$, 1 mM DTT and 0.5 mM PMSF). 20 µl GA beads (Sigma-Aldrich) per reaction were coated with 25 µl GST-TFIIS lysate per reaction for 1 hour at 4˚C. After this, GA-beads were washed three times with binding buffer and incubated with 300 µg HeLa FRT WCLs (unless otherwise stated) in a final volume of 800 µl (50 mM NaCl final concentration) for 2 hours at 4˚C while rotating. Beads were washed three times and proteins were eluted in 15 µl sample buffer by a 5 minutes incubation at 95˚C.

Samples were lysed in sample buffer (160 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 0.05% bromophenol blue), run on an 8% SDS-PAGE and transferred onto a PVDF membrane. The membrane was developed with the appropriate antibodies and ECL. Antibodies were used from the following sources: α-Tubulin (DM1A, CP06, Calbiochem) and GFP (#632381, JL-8 Clontech).

Interactome analysis using mass spectrometry

For interactome analysis, HeLa cells were grown and nuclear extracts were prepared as described[42]. Protein concentration was determined using Bradford assay. GFP affinity purification was essentially performed as described[42]. Protein extracts were incubated with GFP-trap beads (Chromotek) or with blocked agarose beads (Chromotek) as a negative control on a rotating wheel at 4°C using 1 mg input protein. Peptides were eluted from the beads by 2 h trypsin incubation in elution buffer (100 mM Tris-HCl pH 7.5, 2 M urea, 10 mM DTT). Eluate was collected and beads were eluted for a second time. Eluates were combined and trypsin digested overnight. Tryptic digests were desalted using Stage-tips[43].
Samples were cleaned up with in-house-made stage tips. Peptides were separated on a 30-cm pico-tip column (50 µm ID, New Objective) in-house packed with 3 µm aquapur gold C-18 material (Dr. Maisch) by applying a gradient (7–80% ACN 0.1% FA, 140 min), delivered by an easy-nLC 1000 system (LC120, Thermo Scientific), and electro-sprayed directly into an LTQ Orbitrap Mass Spectrometer (Velos, Thermo Scientific). Raw files were analyzed with the MaxQuant software version 1.5.1.0. with oxidation of methionine set as variable and carbamidomethylation of cysteine as fixed modification. The human protein database of UniProt was searched with peptide and protein false discovery rate set to 1%.

Assessing phenotype - genotype correlation

To assess phenotype – genotype correlation, we evaluated the *POLR2A* variants for which support for pathogenicity was obtained from positional and functional analyses, and the variants for which support for pathogenicity was obtained solely from positional analysis. These variants were considered “probable” disease causing, indicating that we are strongly convinced that the individuals’ genotype is causing the individuals’ phenotype. The phenotypic features of the individuals

harboring these variants were used to delineate the phenotypic spectrum. For the remaining variants, pathogenicity was assessed by determining the degree of overlap. Phenotypic similarity was defined as presence of at least three phenotypic features of the five most prevalent phenotypic features in individuals harboring "probable" disease causing variants in *POLR2A*. If overlap in phenotypic features was sufficient, these variants were considered "possibly" disease causing. If phenotypic overlap was not sufficient or could not be determined the variant was classified as "unknown" whether this variant is disease causing.

To correlate phenotypic severity to the predicted consequences of the *POLR2A* variant and to find an explanation for the wide phenotypic spectrum, an ad hoc severity score was conceived[44], that was calculated as follows: each tabulated item that was present (+) in the individual scores 1 point, absent (-) scores 0 points (Table 1). The items "sit without support", "walk well" and "brain magnetic resonance imaging (MRI) abnormalities" were calculated as follows: "sit without support" <12 months scores 0, 13-18 months scores 1, 19-30 months scores 2 and >30 months scores 3 points; "walk well" <24 months scores 0, 25-30 months scores 1, 31-48 months scores 2 and >48 months scores 3 points; no brain abnormalities scores 0 points, wide ventricles and/or delayed myelination scores 1 point and white matter abnormalities scores 2 points. Severity score < 12 points indicates a mild phenotype, 12-17 moderate, 18-23 severe and >23 points indicates a profound phenotype.

RESULTS

Phenotypes of affected individuals

Sixteen individuals, all harboring ultra-rare *de novo* heterozygous variants in *POLR2A*, were captured via GeneMatcher[30]. Other (genetic) causes of the phenotypes of the included individuals were thoroughly excluded. The cohort included three individuals with truncating variants (two stop codons and one frame-shift), three individuals with in-frame deletions (IF deletion) and ten individuals with missense variants. Portraits are shown in Figure 1A, clinical characteristics of all individuals are summarized in Table 1 and extended case-reports can be found in the Supplemental Data. In summary, eleven individuals were born after a generally uneventful pregnancy and delivery, at term with a normal birth weight and normal perinatal events. Decreased fetal movements were noted in three pregnancies and one female was born preterm at

28 weeks. One pregnancy carrying a male fetus (Individual 3, p.Asn531Ser) was terminated because of corpus callosum agenesis, frontonasal dysplasia and a cleft lip. The most frequent early and usually striking phenotypic feature was hypotonia. It was noted in fourteen individuals, being profound in nine (Figure S1), as evidenced by decreased muscle tone, a frog-like posture in infancy and reduced spontaneous movements but normal tendon reflexes. Muscles were considered atrophic in four individuals. Muscle biopsy, on suspicion of a myopathy, was performed in four individuals, but results were inconclusive. In addition, there were symptoms and signs commonly associated with hypotonia, including brachy-plagiocephaly (in 5), high arched palate (in 5), pectus excavatum (in 3), recurrent respiratory tract infections (in 8) and inguinal hernia (in 4). The majority of individuals had feeding difficulties (in 10), with gastro-esophageal reflux (in 6), resulting in failure to thrive (in 3). Ocular signs included strabismus (in 11) and delayed visual maturation (in 6), resulting in decreased vision (in 5). Moderate to severe sensorineural hearing loss was reported in 2, and conductive hearing loss due to recurrent respiratory tract infections was reported in 1 individual. Dysmorphic features were generally mild and non-distinctive, and included a high forehead (in 7), hypertelorism (in 6) and teeth misalignment (in 5), which is an abnormal spacing of the teeth, without missing any teeth. Brain MRI performed in eleven individuals revealed white matter abnormalities (in 10), ranging from delayed myelination (in 5) to wide lateral ventricles, putatively due to white matter loss (in 4). A thin corpus callosum, a feature consistent with loss of white matter was noted in two individuals. Additionally, three individuals had cerebellar abnormalities. Severe epilepsy was reported in three individuals. Eight individuals had sleeping difficulties. Autistic (in 4) and aggressive behavior (in 4) was reported, including extreme behavior such as pica (eating dung). Hypotonia tended to improve over time, albeit slowly. Later in life, endurance was diminished (in 10). A delayed development involving all domains was noted in all individuals and ranged from mild to severe. Severity of the delay was evaluated by calculating Z-scores based on acquisition of developmental milestones, with higher Z-scores indicating later acquisition of milestones. Developmental Z-scores of the individuals were stable over time, consistent with a gradual development, without developmental catch-up or decline. The Z-scores between domains were similar, arguing against developmental domain specific delays (Figure 1B). The degree of delay appeared to correlate with the degree of hypotonia. Temporary loss of milestones was reported in four individuals, usually following an infection. Additional metabolic investigations ruled out known inborn errors of metabolism.

Variant severity metrics

Large scale genetic data from individuals without severe pediatric disease retrieved from gnomAD indicate that *POLR2A* is intolerant of deleterious, heterozygous, protein-changing variants. This is evidenced by a maximal pLI score (1.0) and a very low observed/expected ratio (0.08), indicating intolerance of loss of function variants (Table S2). Moreover, the Z-score for missense variants is one of the highest (7.13) of all human protein coding genes, suggesting that subtle heterozygous changes can also cause a survival disadvantage (Table S2).

Assessment of conservation across species (Figure 2A) revealed that nine out of ten missense variants affect highly conserved amino acid residues (Figure 2B). However, this variant property fails to be discriminatory as the overall degree of conservation of *POLR2A* is extremely high (Figure 2A). In the pol II core, roughly 50% of residues are identical and an additional 20% are highly similar between human and yeast. Similarly, CADD scores of individuals' variants were all above the arbitrary cut-off of 20, which indicates that these variants are predicted to be the 1% most deleterious substitutions that can be done to the human genome, thereby suggesting pathogenicity[32], but so were most of the CADD scores of the variants observed in the gnomAD cohort of individuals without severe pediatric disease (Figure S2). Thus, these indices of pathogenicity at the amino-acid level lack discriminative power.

Next, we evaluated whether pathogenicity was related to the position within *POLR2A*. The observation that the individuals' variants were found throughout the gene argued against a domain-specific deleteriousness. In line with Lelieveld *et al*. and Havrilla *et al.* [45,46] we hypothesized that the distribution of protein changing variants within *POLR2A* observed in the gnomAD cohort of individuals without severe pediatric disease – which are apparently tolerated well enough and therefore more likely to be found at positions tolerant to change – could conversely unveil areas intolerant to change. *S. cerevisiae* tolerates truncations of the CTD region comprising up to 50% of its 26 heptad repeats in *Rpb1*, the homolog of *POLR2A*[47,48]. This strongly suggests that the CTD region of human *POLR2A*, consisting of 52 repeats, may be equally tolerant to loss of a proportion of the heptad repeats. Indeed, the majority of the truncating variants reported in gnomAD – and none of the truncations found in these individuals – are positioned in the distal portion of the CTD region (Figure 2A). In addition, the density of *POLR2A* missense variants reported in gnomAD is also highest in the CTD region, not only

in the distal part but throughout the whole repeat region, indicating that a putative functional loss of a single heptad repeat is tolerated, regardless of its position.

Within the remaining part of *POLR2A*, missense variants are strikingly unevenly distributed when compared with synonymous variants (Figure 2A). In line with our hypothesis, while the median distance between variants is only two amino acid residues, several stretches devoid of missense variants are observed, the longest of which spans about 80 amino acid residues (Figure 2A). These stretches devoid of missense variants correspond to a number of known functionally important regions of the protein. Nine out of ten missense variants from our individuals, two out of three in-frame deletions and two out of three truncating variants were positioned within these stretches with desert Z-scores between 2.9 and 11.9, indicating intolerance to change in these regions and thereby supporting pathogenicity (Figure 2A, Table 2).

Structural evaluation of POLR2A variants in pol II

Most of the amino residues affected in individuals have identical or similar residues in the *S. cerevisiae* counterpart (Figure 2A-B and 2J), for which high resolution structural information is available. This allows for structural evaluation of the variants in pol II. Variants p.Gln700* and p.Gln735* result in truncation, missing more than half of the protein. If translation is not already prevented by nonsense-mediated decay, the truncated protein is still not expected to form a stable fold with the second largest and core subunit of pol II, RPB2, and it lacks binding sites for several other subunits (Table 2). It is therefore most likely that effects observed with p.Gln700* and p.Gln735* result from RPB1 haploinsufficiency. In support, mass spectrometry analyses showed that the variant p.Lys812* (used as a representative of p.Gln700* and p.Gln735* since p.Lys812* is similarly truncated, see Materials and Methods) could only associate with RPABC3 (*POLR2H)* but not with any of the other pol II subunits (Figure 3C). Variant p.Pro1767fs is located in the proximal half of the CTD and results in the loss of the last 200 amino acids, corresponding to about 50% of the heptad repeats. Folding of the polymerase core is not expected to be affected by this truncation[49,50]. Therefore, p.Pro1767fs is expected to form pol II, but since the CTD region of RPB1 is too short, pol II function might be affected.

The IF deletions p.Tyr669del (p.Phe646$_{yeast}$) and p.Ser755del (p.Leu732$_{yeast}$) are localized in close proximity to the quay, whilst the IF deletion p.Lys1125del (p.Lys1102$_{yeast}$) is localized in the trigger loop (Figure 2C-F). The deletions are predicted to alter the local protein fold and to have a strong impact on catalytic activity. It is difficult to predict to what extent the overall protein fold and the ability to assemble into pol II is impaired, but such impairment would shift the properties from severe dominant negative to mild haploinsufficiency. In case of p.Ser755del, mass spectrometry data showed that the interaction with other pol II subunits is largely unaltered, showing that pol II complex formation is not affected (Figure 3C).

Interestingly, most missense variants are centered around the catalytic site (Figure 2C-F, Table 2): p.Tyr1109His (p.Phe1086$_{yeast}$) and p.Leu1124Pro (p.Leu1101$_{yeast}$) are positioned in the trigger loop, p.Thr736Met (p.Ser713$_{yeast}$) in the opposing quay, p.Ile457Thr (p.Leu443$_{yeast}$) and p.Met769Thr (p.Met746$_{yeast}$) in the nucleotide binding site, p.Pro371Leu (p.Pro357$_{yeast}$) in the funnel (Figure 2G) and p.Ile848Thr (p.Ile825$_{yeast}$) in the bridge helix (Figure 2C-F, Table 2). These *POLR2A* variants are likely to reduce the transcriptional pausing, elongation and/or back-tracking activities of pol II, without affecting transcription initiation. Indeed, all tested missense RPB1 variants were found to interact with other pol II subunits, indicating that complex formation is largely unaltered (Figure 3C).

The missense mutation p.Asn1251Ser (p.Asn1232$_{yeast}$) is located distantly from the catalytic site, but is localized in the interaction surface for transcription elongation factor TFIIS, which is required for back-tracking of pol II due to elongation blocks or nucleotide mis-incorporation (Figure 2I). Pol II mutant for TFIIS interactions would result in both an increased error rate and reduced elongation rates. The consequences of the missense variants p.Asn531Ser (p.Asn517$_{yeast}$) (Figure 2H) and p.Arg1603His are more difficult to predict. p.Asn531 is not directly associated with a structural element involved in catalysis (Figure 2H). p.Arg1603His is localized in the region that connects the RPB1 core to the heptad repeats (Figure 2A). No structural information is available for this region and it is only partially conserved between human and *S. cerevisiae* RPB1 (Figure 2A).

Altogether, structural evaluation of *POLR2A* variants resulted in expected haploinsufficiency for p.Gln700* and p.Gln735*, putative haploinsufficiency for the IF deletions and an expected dominant negative effect for the missense variants p.Ile457Thr, p.Thr736Met, p.Met769Thr, p.Tyr1109His and p.Leu1124Pro (Table 2).

Functional evaluation of POLR2A variants in pol II

The consequences of *POLR2A* variants on pol II function were investigated in *S. cerevisiae* and HeLa cells for six and nine of the individuals' variants, respectively, representing the whole clinical spectrum (Table 2).

In *S. cerevisiae*, in a wildtype genetic background, normal growth was observed in all but one variant (p.Leu1124Pro (p.Leu1101Pro$_{yeast}$) under optimal conditions as well as under temperature stress conditions (YPD at 37ºC) (Table 2, Figure S3). In genetic backgrounds lacking transcription factors Dst1 (yeast homolog of TFIIS) or Sub1 (yeast homolog of PC4) aberrant growth surfaced in four out of six variants (p.Leu1124Pro (p.Leu1101Pro$_{yeast}$), p.Ile457Thr (p.Leu443Thr$_{yeast}$), p.Thr736Met (p.Ser713Met$_{yeast}$) and p.Ser755del (p.Leu732del$_{yeast}$) when compared with wildtype and the positive controls for Δ*dst1* (p.Glu1230Lys$_{yeast}$) (Table 2, Figure S3) and Δ*sub1* (p.Ala301Asp$_{yeast}$) (Figure 3A, Table 2). A read-through transcription assay, in which GAL7 production is inhibited in WT (*gal10Δ56*) cells, causing galactose sensitivity on YPRaf/Gal[23], also shows aberrant growth for these four variants (Figure 3A, Table 2, Figure S3). Although milder, these four variants also exhibited aberrant growth when exposed to the nucleotide synthesis inhibitor mycophenolic acid (MPA) in one or both of these genetic backgrounds (Figure 3A, Table 2, Figure S3). Disturbances in both genetic backgrounds (Δ*dst1* and Δ*sub1*) are suggestive of reduced transcriptional fidelity.

Cell viability was assessed with an MTT (3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazoliumbromide) assay in doxycycline-inducible GFP-RPB1$^{amanitin\ resistant}$ expressing mutant HeLa FRT cells, in which the endogenous wildtype RPB1 was annihilated using the pol II-specific drug α-amanitin. Cell viability was grossly reduced in mutants expressing variants p.Thr736Met, p.Ser755del, p.Lys812* and p.Leu1124Pro (Figure 3B), which supports the pathogenicity of these variants. No significant effects were seen for the variants p.Ile457Thr, p.Asn531Ser, p.Asn1251Ser and p.Arg1603His (Figure 3B). For all mutants the amounts of RPB1 protein were similar to the protein amount of wildtype RPB1 (Figure S4). Binding affinity towards TFIIS, a key transcription factor, was unaffected for all mutants. This included the p.Asn1251Ser mutant, despite its

location in the TFIIS binding site (Figure S4). Taken together, characterization in *S. cerevisiae* and MTT cell viability in HeLa cells provide evidence for pathogenicity of variants located close to the catalytic core, but not for p.Asn531Ser, p.Asn1251Ser and p.Arg1603His (Table 2).

Assessing phenotype – genotype correlation

To delineate the phenotypic spectrum that could be attributed to *POLR2A* variants, we focused on variants considered to be probable disease causing. Positional and functional evaluation of the *POLR2A* variants on pol II function supported pathogenicity of the missense variants p.Ile457Thr, p.Thr736Met and p.Leu1124Pro, as well of the IF deletion p.Ser755del and the truncated variants p.Gln700* and p.Gln735* (Table 2) and solely positional evaluation supported pathogenicity for the missense variants p.Met769Thr, p.Ile848Thr, p.Tyr1109His and p.Asn1251Ser and for the IF deletion p.Lys1125del (Table 2). The phenotypic features of the eleven individuals harboring these variants were used to delineate the *POLR2A* phenotypic spectrum (Table 2). These phenotypic features included profound general hypotonia (in 10), as evidenced by a frog position in infancy (in 8), strabismus (in 9), decreased endurance (in 8) and feeding difficulties (in 7) (Table 1). Pathogenicity of the remaining variants was assessed by determining the degree of phenotypic overlap. Indiviual 1 (p.Pro371Leu), individual 4 (p.Tyr669del), individual 15 (p.Arg1603His) and individual 16 (p.Pro1767fs) presented with clinical phenotypes clearly fitting the phenotypes of the other *POLR2A* individuals since all had at least three of the five most prevalent *POLR2A* symptoms (Table 1) and therefore these variants were considered possibly disease causing (Table 2). For individual 3 (p.Asn531Ser) phenotypic overlap could not be assessed, since pregnancy had been terminated and it is thus unknown whether this variant is disease causing (Table 2).

We noted that the degree of developmental delay correlated with the predicted consequences of the variant: in the individuals with variants predicted to cause haploinsufficiency relatively low Z-scores (mean Z-score 1.74 $\pm$ 3.13) were noted, but variants expected to cause a dominant negative effect were associated with appreciably higher Z-scores (median Z-score 4.23 $\pm$ 3.96, p = <0.0001) and thus more severe delays (Figure 1B). To assess this further, a severity score was calculated which summarizes the severity of the clinical phenotype for *POLR2A*. Severity scores of individuals harboring variants with expected or putative haploinsufficiency (mean severity score 10, range 6-14 points) were lower

than in individuals harboring variants with expected dominant negative effects (mean severity score 17, range 6-25 points, Table 1). This suggests that a missense mutation exerting a dominant negative effect is more likely to result in a severe phenotype (severity score > 15) than a mutation inducing haploinsufficiency.

DISCUSSION

In this study we applied an iterative approach to assess pathogenicity of identified variants in *POLR2A*. The main argument supporting pathogenicity of the identified variants, is that all variants occurred *de novo*[27]. In addition, *POLR2A* is, overall, very intolerant of deleterious, heterozygous, protein-changing variants. Furthermore, nine out of ten missense variants affected highly conserved amino acid residues that are localized at important functional regions of the gene. Functional analyses, although we note that both yeast and HeLa cells are model systems that may not fully recapitulate the complex developmental disease of the individuals reported here, support pathogenicity of the variants p.Ile457Thr, p.Thr736Met, p.Gln700*, p.Gln735*, p.Ser755del and p.Leu1124Pro. Although we did not demonstrate support from functional analyses for the variants p.Asn531Ser and p.Asn1251Ser, pathogenicity of p.Asn1251Ser is strongly supported by the conservation of the amino acid residue and the importance of the region for interaction with TFIIS. Altogether, based on the results of both predictive and functional analyses, eleven variants were classified as probably disease-causing (Table 2).

Deep phenotyping of the individuals, including quantification of phenotypic severity allowed us to delineate the phenotypic spectrum that could be ascribed to the pathogenic *POLR2A* variants. The severe end of the spectrum is characterized by profound infantile onset hypotonia and developmental delay, and is further accompanied by strabismus, decreased endurance and feeding difficulties. Determining the degree of clinical overlap between the individuals harboring probable disease-causing variants and the remaining individuals, allowed us to classify four additional variants as possibly disease-causing. Additional support for pathogenicity of the variant p.Asn531Ser, other than *de novo* occurrence, conservation across species and a severe phenotype, could not be determined, so this variant was classified as of unknown significance.

Thus, the iterative process described here allowed us to delineate pathogenicity for the majority – but not all – individuals' variants. With the current knowledge, evidence that either confirms or excludes pathogenicity of four possibly disease-

causing variants and the one variant of unknown significance, will be hard to obtain. This is illustrated by the variant p.Arg1603His. Support for pathogenicity is obtained from its *de novo* occurrence and from the phenotypic overlap of individual 15 with individuals harboring probable disease-causing variants in *POLR2A,* such as extreme hypotonia with frog-like positioning of the legs, developmental delay, strabismus, recurrent respiratory tract infections and wide ventricles, thin corpus callosum and bilateral white matter loss on brain MRI. However, arguments against pathogenicity are that the variant has been reported three times in gnomAD and that it is part of a poorly conserved region with no known functional importance. We stress the importance of including these unresolved issues when describing novel disease causing genes, as this aids subsequent resolution of variant pathogenicity when more individuals are described.

In our search for lines of evidence to support pathogenicity, we found an extra line of evidence to support pathogenicity of the variants in *POLR2A* in the design of a desert Z-score. While the CADD score[32], which is usually seen as an estimate of the likelihood of a variant being pathogenic, did not differentiate between variants reported in gnomAD and the missense variants reported here, the variants that we report appeared to cluster within *POLR2A* in large regions devoid of apparently harmless variants reported in gnomAD. We calculated the desert Z-score to reflect this property for all variants within *POLR2A*. This score confirmed not only that the size of these regions exceeded the size that would be expected by chance (the largest region had a Z-score of 11.9), but also the clustering of individuals' variants within these regions. The relevance of this finding was supported by subsequent structural analyses which unveiled the functional importance of these 'desert regions'. We therefore anticipate in line with Lelieveld *et al.* and Havrilla *et al.,*[45,46] that this property may be helpful to identify within-gene regions that are relatively intolerant of genetic variants, and are likely to be of functional importance.

The quantification of phenotypic severity unveiled that individuals harboring heterozygous truncating variants that are incapable of proper pol II formation presented with a mild phenotype, while individuals harboring heterozygous missense variants that allow formation of pol II exhibited the most severe phenotype. This observation implies that the presence of malfunctioning species of pol II is more detrimental than reduced availability of pol II alone. We propose that the aberrant

pol II enzymes are capable of proper assembly of the pol II machinery at the transcription start sites, but subsequent elongation of the nascent RNA occurs at reduced rates and possibly with an increased error rate, blocking access and progression of wildtype pol II on the same DNA strand. Altogether, this could greatly influence the dynamics of pol II initiation and release which are essential components of transcriptional regulation[21]. The results reported here direct further mechanistic studies to investigate this defective elongation hypothesis.

Malfunctioning pol II has deleterious consequences on transcription. To date, none of the eleven other pol II subunits have been implicated in human disease. Interestingly, we noticed some phenotypic overlap between individuals harboring variants in *POLR2A* and individuals harboring variants in subunits of pol III. Mutations in *POLR1C* (MIM #610060, encoding a subunit that is part of both pol III and pol I), *POLR3A* (MIM #614258) and *POLR3B* (MIM #614366) have been described to cause hypomyelinating leukodystrophy 11 (HLD11, MIM #616494), 7 (HLD7, MIM #607694) and 8 (HLD8, MIM #614381), respectively. These three diseases are summarized under the name POLR3-related leukodystrophy[51]. Recently, two individuals with HLD were reported to harbor mutations in *POLR3K*[52], thereby expanding the list of potential genetic defects underlying POLR3-related leukodystrophy. In addition to hypomyelination on brain imaging, the clinical phenotype of POLR3-related leukodystrophy is characterized by progressive cerebellar dysfunction and cognitive dysfunction. Non-neurological features are abnormal dentition and hypogonadotropic hypogonadism. Phenotypic features of POLR3-related leukodystrophy that are also noted in *POLR2A* individuals are delayed myelination (5/15), white matter loss (6/15) and teeth misalignment (5/15). However, the cellular functions of the three nuclear RNA polymerases are different, as pol III and pol II are responsible for tRNA and mRNA synthesis, respectively. This can be reconciled by the proposal that mutations in both pol II and pol III subunits can affect protein synthesis via different mechanisms, and that this can result in similar clinical phenotypes.

In addition to POLR3-related leukodystrophy, *POLR1A* has been associated to both acrofacial dysostosis[53] (MIM #616462) and to severe neurodegenerative disease with ataxia, psychomotor retardation, cerebellar and cerebral atrophy and leukodystrophy[54] and *POLR3A* is in addition to HLD7 also associated with Wiedemann-Rautenstrauch syndrome, a

neonatal progeroid syndrome[55] (MIM #264090) and discussed as potential cause of hereditary ataxia and spastic paraparesis[56,57], implying a broader phenotypic range for *POLR3A* mutations.

In summary, we here report that heterozygous *de novo* variants in *POLR2A*, which is indispensable for the synthesis of mRNA and several non-coding RNAs as it encodes the RPB1 subunit of pol II, can result in a neurodevelopmental syndrome characterized by profound infantile onset hypotonia and developmental delay. We conclude that the clinical consequences of probable disease-causing variants in *POLR2A* are dependent on their effect on pol II-mediated transcription, as *POLR2A* variants predicted to result in loss of RPB1 protein are better tolerated than missense variants, which we propose can cause a dominant negative effect on pol II-dependent transcription.

SUPPLEMENTAL DATA DESCRIPTION

The supplement contains extended case reports, 4 figures and 2 tables.

DECLARATION OF INTERESTS

S.P. discloses that he is an employee of AstraZeneca. The other authors declare no competing interests.

ACKNOWLEDGEMENTS

We thank all included individuals and their families for their cooperation. We are indebted to Craig Kaplan for providing yeast strains, plasmids and insightful discussions. The study was supported by the following grants: for individuals p.Pro371Leu and p.Met769Thr 17-29423A from the Czech Ministry of Health, for individual p.Gln735* from the Oxford NIHR Biomedical Research Centre and the Health Innovation Challenge Fund (R6-388/WT100127, a parallel funding partnership between the Wellcome Trust and the Department of Health), for individual p.Ile848Thr the Duke Genome Sequencing Clinic supported by the Duke University Health System, for individual p.Tyr1109His Telethon Foundation, Telethon Undiagnosed Diseases Program (TUDP, GSP15001) and for individual p.Lys1125del Mining for Miracles (BCCH Foundation) and Genome British Columbia in the CAUSES study (Clinical Assessment of the Utility of Sequencing as a

Service); investigators include Shelin Adam, Christèle du Souich, Alison M. Elliott, Anna Lehman, Jill Mwenifumbo, Tanya N. Nelson, Clara van Karnebeek, and Jan M. Friedman, with bioinformatics support from the lab of Wyeth Wasserman. The work of HTMT is supported by the SFB850 and SFB992 programs of the Deutsche Forschungsgesellschaft (DFG). The mass spectrometry analysis was supported by the "Proteins at Work" program of The Netherlands Organization for Scientific Research (NWO, 184.032.201). The work of H.A. Haijes is supported by the personal Alexandre Suerman Stipend of the University Medical Centre Utrecht.

WEB RESOURCES

GeneMatcher	http://www.genematcher.org
gnomAD	http://gnomad.broadinstitute.org
OMIM	http://www.omim.org
Protein Data Bank	http://www.rcsb.org
UniProt	https://www.uniprot.org

REFERENCES

1. Wintzerith M, Acker J, Vicaire S, Vigneron M, Kedinger C. (1992). Complete sequence of the human RNA polymerase II largest subunit. Nucleic Acids Res *20*:910.
2. Mita K, Tsuji H, Morimyo M, Takahashi E, Nenoi M, Ichimura S, Yamauchi M, Hongo E, Hayashi A. (1995). The human gene encoding the largest subunit of RNA polymerase II. Gene *159*:285-286.
3. Kornberg, R.D. (1999). Eukaryotic transcriptional control. Trends Cell Biol *9*(12):46-9.
4. Roeder, R.G., Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*:234-237.
5. Thomas, M.C., Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*(3):105-178
6. Sainsbury S, Bernecky C, Cramer P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol *16*:129-143.

7. Bushnell, D.A., Westover, K.D., Davis, R.E., Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science *303*(5660):983-988.
8. Sainsbury, S., Niesser, J., Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. Nature *493*(7432):437-440.
9. Plaschka, C., Lariviere, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F. et. al. (2015). Architecture of the RNA polymerase II-mediator core initiation complex. Nature *518*(7539):376-380.
10. Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.J., Burlingame, A.L., Kornberg, R.D. (2016). Structure of a complete mediator-RNA polymerase II pre-initiation complex. Cell *166*(6):1411-1422.
11. Westover, K.D., Bushnell, D.A., Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*(4):481-489.
12. Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. Curr Op Genet Dev *14*(2):218-226.
13. Cheung, A.C., Sainsbury, S., Cramer, P. (2011) Structural basis of initial RNA polymerase II transcription. EMBO J *30*(23):4755-4763.
14. Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., Kornberg, R.D. (2011). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science *292*(5523):1876-1882.
15. Jonkers I, Lis JT. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*:167-177.
16. Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., Sekine, S.I. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. Science *357*(6354):921-924.
17. Kettenberger, H., Armache, K.J., Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol Cell *16*(6):955-965.

18. Sydow, J.F., Brueckner, F., Cheung, A.C. Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol Cell *34*(6):710-721.
19. Cheung, A.C., Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*(7337):249-253.
20. Walmacq, C., Cheung, A.C., Kireeva, M.L., Lubkowska, L., Ye, C., Gotte, D., Strathern, J.N., Carell, T., Cramer, P., Kashlev, M. (2012). Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. Mol Cell *46*(1):18-29.
21. Steurer B, Janssens RC, Geverts B, Geijer ME, Wienholz ME, Wienholz F, Theil AF, Chang J, Dealy S, Pothof J. et. al. (2018). Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promotor-paused RNA polymerase II. Procl Natl Acad Sci U S A *115*(19):E4368-E4376.
22. Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., Kornberg, R.D. (2006). Structural basis of transcription: the role of the trigger loop in substrate specificity and catalysis. Cell *127*(5):941-954.
23. Kaplan, C.D., Jin, H., Zhang, I.L., Belyanin, A. (2012). Dissection of pol II trigger loop function and pol II activity-dependent control of start site selection *in vivo*. PLoS Genetics *8*(4):e1002627.
24. Timmers, H.T.M., Tora, L. (2018). Transcription buffering: a balancing act between mRNA synthesis and mRNA degradation. Mol Cell *72*(1):10-17.
25. Sugaya, K. (2003). Amino acid substitution of the largest subunit of yeast RNA polymerase II: effect of a temperature-sensitive mutation related to G1 cell cycle arrest. Curr Microbiol *47*(2):159-162.
26. Zhang QQ, Li F, Fu ZY, Liu XB, Yuan K, Fang Y, Liu Y, Li G, Zhang XS, Chong K et. al. (2018). Intact Arabidopsis RPB1 functions in stem cell niches maintenance and cell cycling control.
27. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, 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(5):405-424.

28. Lelieveld, S.H., Reijnders, M.R.F., 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. Nature Neuroscience *19*(9):1194-1196.
29. Lek, M., Karczewski, K.J., Minikel, E.V. Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*(7616):285-291.
30. Sobreira, N., Schiettecatte, F., Valle, D., Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mut *36*(10):928-930.
31. Gray, O.P. (1972). The Denver Scale. Dev. Med. Child Neurol. *14*(5):666-667.
32. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. *46*(3):310-317.
33. Schomburg, D., Reichelt, J. (1988). BRAGI – A comprehensive protein modeling program system. Journal of Molecular Graphics *6*(3):161-165.
34. Kraulis, P.J. (1991). Molscript – a program to produce both detailed and schematic plots of protein structures. Journal of Applied Crystallography *24*:946-950.
35. Merritt, E.A., Murphy, M.E.P. (1994). Raster3D version 2.0 – A program for photorealistic molecular graphics. Acta Crystallographica Section D-Biological Crystallography *50*:869-873.
36. Sugaya, K., Vigneron, M., Cook, P.R. (2000). Mammalian cell lines expressing functional RNA polymerase II tagged with the green fluorescent protein. J Cell Sci *113*(15):2679-2683.
37. Sugaya, K., Sasanuma, S., Cook, P.R., Mita, K. (2001). A mutation in the largest (catalytic) subunit of RNA polymerase II and its relation to the arrest of the cell cycle in G(1) phase. Gene *274*(1-2):77-81.
38. Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., Strathern, J.N. (2006). Mutations in the *Saccharomyces cerevisiae* RPB1 gene conferring hypersensitivity to 6-azauracil. Genetics *172*(4):2201-2209.
39. Nguyen, V.T., Giannoni, F., Dubois, M.F., Seo, S.J., Vigneron, M., Kédinger, C., Bensaude, O. (1996). *In vivo* degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. Nucleic Acids Res *24*(15):2924-9.

40. Baas, R., Lelieveld, D., van Teeffelen, H., Lijnzaad, P., Castelijns, B., van Schaik, F.M., Vermeulen, M., Egan, D.A., Timmers, H.T., de Graaf, P. (2014). A novel microscopy-based high-throughput screening method to identify proteins that regulate global histone modification levels. J Biomol Screen *19*(2):287-96.
41. Pereira, L.A., van der Knaap, J.A., van den Boom, V., van den Heuvel, F.A., Timmers, H.T. (2001). TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. Mol Cell Biol *21*(21):7523-34.
42. Baymaz, H.I., Spruijt, C.G., Vermeulen, M. (2014). Identifying nuclear protein-protein interactions using GFP affinity purification and SILAC-based quantitative mass spectrometry. Methods Mol. Biol. *1188*: 207-226
43. Rappsilber, J., Ishihama, Y., Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*(3): 663-670
44. Haijes, H.A., Jaeken, J., Foulquier, F., van Hasselt, P.M. (2018). Hypothesis: lobe A (COG1-4)-CDG causes a more severe phenotype than lobe B (COG5-8)-CDG. J Med Genet *55*(2):137-142.
45. Lelieveld, S.H., Wiel, L., Venselaar, H., Pfundt, R., Vriend, G., Veltman, J.A., Brunner, H.G., Vissers, L.E.L.M, Gilissen, C. (2017). Spatial clustering of *de novo* missense mutations identifies candidate neurodevelopmental disorder-associated genes. Am J Hum Genet *101*(3):478-484.
46. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR (2019) A map of constraint coding regions in the human genome. Nat Genet *51*(1):88-95.
47. Scafe, C., Chao, D., Lopes, J., Hirsch, J.P., Henry, S., Young, R.A. (1990). RNA polymerase II C-terminal repeat influences response to transcriptional enhancer signals. Nature *347*(6292):491-4.
48. Eick, D., Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem Rev *113*(11):8456-8490.
49. Nonet M, Sweetser D, Young RA (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell *50:*909-915
50. Rosinona E, Blencowe BJ (2004) Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3'-end cleavage. RNA *10:*581-589
51. Bernard, G., Vanderver, A. (2012). POLR3-related leukodystrophy. GeneReviews. Updated: May 11, 2017.

52. Dorboz I, Dumay-Odelot H, Boussaid K, Bouyacoub Y, Barreau P, Samaan S, Jmel H, Eymard-Pierre E, Cances C, Bar C. et. al. (2018). Mutation in *POLR3K* causes hypomyelinating leukodystrophy and abnormal ribosomal RNA regulation. Neurol Genet *4*(6):e289.
53. Weaver KN, Watt KEN, Hufnagel RB, Navajas Acedo J, Linscott LL, Sund KL, Bender PL, Konig R, Lourenco CM, Hehr U. et. al. (2015) Acrofacial dysostosis, Cincinatti type, a mandibulofacial dysostosis syndrome with limb anomalies, is caused by *POLR1A* dysfunction. Am J Hum Genet *96*:765-774.
54. Kara B, Köröglu C, Peltonen K, Steinberg RC, Maras Genc H, Hölttä-Vuori M, Güven A, Kanerva K, Kotil T, Solakoglu S. et. al. (2017). Severe neurodegenerative disease in brothers with homozygous mutation in *POLR1A*. Eur J Hum Genet *25*(3):315-323.
55. Jay AM, Conway RL, Thiffault I, Saunders C, Farrow E, Adams J, Toriello HV. (2016). Neonatal progeroid syndrome associated with biallelic truncating variants in *POLR3A*. Am J Med Genet A *170*(12):3343-3346.
56. Rydning SL, Koht J, Sheng Y, Sowa P, Hjorthaug HS, Wedding IM, Erichsen AK, Hovden IA, Backe PH, Tallaksen CME. et. al. (2019). Biallelic *POLR3A* variants confirmed as a frequent cause of hereditary ataxia and spastic paraparesis. Brain. [Epub ahead of print]
57. Minnerop M, Kurzwelly D, Wagner H, Schüle R, Ramirez A. Reply: biallelic *POLR3A* variants confirmed as a frequent cause of hereditary ataxia and spastic paraparesis. Brain. [Epub ahead of print]

FIGURE LEGENDS

Figure 1. A) Photos of thirteen of the sixteen individuals (ind.) harboring a *de novo* variant in *POLR2A*, in order of severity score. **B)** Attained and unattained developmental milestones during the first eighteen months of life, of fourteen of the sixteen individuals harboring a *de novo* variant in *POLR2A*, in order of severity score. For each milestone, Z-scores were calculated indicating the delay in attaining that milestone. The higher the Z-scores, the greater the developmental delay. The figure demonstrates that in individuals with a mild phenotype, developmental Z-scores are mostly <10, in individuals with a moderate phenotype, developmental Z-scores are <20 and in individuals with a severe phenotype, developmental Z-scores can even reach 30. In addition, among the individuals with a severe phenotype, relatively more (and more early) developmental milestones were not attained by the individual. Individual 3 (p.Asn531Ser) was not included in this figure, since the fetus was aborted.

Figure 2. A) RPB1 is represented with the core and the C-terminal heptad repeats as bars. Functional elements are highlighted in dark blue and are labelled underneath. The positions of gained stop codons, frameshifts, in-frame deletions (IF deletion), missense mutations, and synonymous mutations at the protein level as reported in the ExAC database, are indicated by vertical black strokes. The length of the stroke represents the number of different mutations observed on DNA level in the corresponding codon. Individuals' variants are indicated by red strokes and are labelled. Stretches devoid of missense mutations are highlighted by yellow background and are labelled by Roman numerals. The degree of conservation in human and *S. cerevisiae* is depicted underneath, whereby each identical amino acid is represented by a vertical black stroke. **B)** Conservation across species of amino acid residues affected in individuals with missense mutations. p.Arg1603His is not conserved and is localized in the region between the pol II core and the heptad repeats, and p.Pro1767fs is localized in the heptad repeat region. For these regions, alignments are not reliable. **C)** Space filling model of the elongating form of pol II from *S. cerevisiae* with TFIIS in ribbon representation in red. Subunits other than RPB1 are in dark grey. RPB1 is shown in light grey with positions of individuals' variants highlighted in magenta and stretches devoid of missense variants in shades of yellow labelled by Roman numerals as in A. The incoming nucleotide is shown as space filling model in orange. The template strand, coding strand, and the nascent mRNA are shown as

backbone trace in grey, black, and orange, respectively. **D)** Same as in C, but with omission of subunits other than RPB1. Only the RPB1 stretches devoid of missense variants and the positions of individuals' variants are shown and all other parts of RPB1 have been omitted. The 3D-envelope of RPB1 is projected in light grey on the background plane. **E)** Same as in D but with additional omission of the RPB1 stretches devoid of missenses variants. The Cα-atoms of amino acid residues affected in individuals' variants are shown as magenta spheres. **F)** Detailed view at the catalytic center of pol II. The coding strand in dark grey, the template strand in light grey, the nascent mRNA in orange and the incoming nucleotide in red are shown as space filling model. The bridge helix is shown as a rod and selected other parts of RPB1 as backbone trace. The trigger loop is shown in its open and closed conformation in light and dark blue, respectively. Side chains of amino acid residues affected in individuals with a missense variant and of some residues known to be of catalytic importance are shown in ball-and-stick representation. The Cα-atoms of amino acid residues affected by IF deletions or gain of stop codon are shown as spheres in magenta. **G)** Environment of residue Pro357. Elements from RPB1, subunit 2, and the pol II subunit RPB11 are depicted in light blue, green, and orange, respectively. Hydrogen bonds are indicated by dotted lines **H)** Environment of residue Asn517. Elements of RPB1 are shown in blue **I)** Environment of residue Asn1232. Elements of RPB1 and TFIIS are shown in blue and red, respectively. Putative hydrogen bonding partners for Asn1232 in TFIIS are marked by an asterisk. **J)** Table assigning amino acid residues affected in individuals to the corresponding residues of RPB1 in *S. cerevisiae*.

Figure 3. Functional evaluation of *POLR2A* variants. **A)** Growth assay of six *rpb1* mutants in *S. cerevisiae*: p.Ile457Thr (p.Leu443Thr), p.Asn531Ser (p.Asn517Ser), p.Thr736Met (p.Ser713Met), p.Ser755del (p.Leu732del), p.Leu1124Pro (p.Leu1101Pro) and p.Asn1251Ser (p.Asn1232Ser) are compared to wildtype (WT). p.Ala301Asp demonstrates a positive control for reduced transcriptional activity and p.Glu1230Lys demonstrates a positive control for reduced genetic interaction with TFIIS. YPD 30°C depicts growth under normal circumstances, YPD 37°C depicts growth under stress induced by increased temperature. YPRaf/Gal day 4 and day 6 depict growth under stress induced by increased inappropriate GAL 3'termination, on day 4 and day 6 of culture, respectively. SC-LEU + MPA day 4 and day 6 depict growth under stress induced by lack of leucine and addition of mycophenolic acid (MPA), on day 4 and day 6 of culture, respectively. **B)** HeLa cell viability. Individuals' variants were introduced in an α-amanitin resistant version of *POLR2A*

(ama^{R}) and expressed from a Doxycycline-inducible promoter as a GFP fusion construct in HeLa cells. After induction of expression, endogenous pol II was left untreated (dark grey bars) or inhibited by the application of α-amanitin (light grey bars) and cell viability was determined. p.Lys812* was used as a representative of p.Gln700* and p.Gln735* since it is similarly truncated, see Materials and Methods. **C**) Interactome analysis of wildtype and mutant versions of RPB1 using mass spectrometry. Nuclear extracts from HeLa cells expressing GFP-tagged wildtype or mutant versions of RPB1 were used for precipitations with control agarose beads (-) or GFP-affinity beads (+) in triplicate. In total 1133 proteins were identified with more than two peptides by IBAQ-based quantitative mass spectroscopy. Known members of pol II (high-lighted) and a random selection of 5% of the other proteins are shown. The summed peptide intensities of identified proteins were logarithmically transformed to the basis of ten and are represented by shades of grey. Note, intensities smaller than 10^{4} are not obtained due to technical constrains. p.Lys812* was used as a representative of p.Gln700* and p.Gln735*.

Table 1 - Characteristics of individuals harboring a heterozygous de novo variant in POLR2A

Individual Variant	**5** p.Gln700*	**6** p.Gln735*	**8** p.Ser755del	**12** p.Leu1124Pro	**13** p.Lys1125del	**1** p.Pro371Leu	**16** p.Pro1767fs
Sex	M	F	M	F	M	F	M
Gestation (weeks)	42	28	Term	38	37	41	41
Age (years)	17	13	3	4	7	7	9
General hypotonia	-	+	+	+	+	+	+
Strabismus	-	+	+	+	+	+	-
Frog position infancy	-	-	+	+	-	-	-
Decreased endurance	-	-	+	+	+	-	+
Feeding difficulties	-	+	-	+	-	+	+
Recurrent RTI	-	-	+	+	-	-	+
High forehead	+	+	+	-	+	-	+
Disturbed sleeping	+	-	-	-	-	-	-
Gastro-esoph. reflux	-	+	-	-	-	-	-
High palate	+	-	-	-	+		-
Delayed visual matur.	+	-	-	-	-	-	-
Microcephaly	-	-	-	-	-	-	-
Brachyplagiocephaly	-	-	-	-	+	-	+
Muscle atrophy	-	-	-	-	-	-	-
Hypertelorism	-	-	-	-	+	+	+
Teeth misalignment	-	+	-	+	+	-	+
Decreased vision	+	-	-	-	-	+	-
Stagnation episodes	-	-	-	-	-	-	-
Inguinal hernia	-	-	-	-	-	-	-
Decreased fetal movem.	-	-	NA	+	-	-	-
Autistic behavior	+	-	-	NA	-	-	-
Aggressive behavior	-	-	-	-	-	-	-
Failure to thrive	-	-	-	-	-	+	-
Pectus excavatum	-	-	-	-	-	-	-
Epilepsy	-	-	-	-	-	-	-
Sit, no support (months)	7	12	NA	12	18	24	24
Walk well (months)	18	30	>36	23	24	72	28
Brain MRI abnormalities	-	Cerebellar atrophy, inferior vermis, small pons, megacisterna magna, slightly small corpus callosum	Delayed myelination	Delayed myelination	Delayed myelination	-	-
Severity score	**6**	**9**	**9**	**9**	**10**	**11**	**11**
Severity class	Mild	Mild	Mild	Mild	Mild	Mild	Mild

...Table 1 continued

Individual Variant	**4** p.Tyr669del	**11** p.Tyr1109His	**10** p.Ile848Thr	**15** p.Arg1603His	**9** p.Met769Thr	**14** p.Asn1251Ser	**2** p.Ile457Thr	**7** p.Thr736Met	**3** p.Asn531Ser
Sex	F	M	F	F	M	F	M	F	M
Gestation	41	40	41	Term	41	40	40	39	30 †
Age (years)	11	6	13	7	18	6	4	9	0
General hypotonia	+	+	+	+	+	+	+	+	
Strabismus	-	+	-	+	+	+	+	+	
Frog position infancy	+	+	+	+	+	+	+	+	
Decreased endurance	+	NA	+	-	+	+	+	+	
Feeding difficulties	+	-	+	-	+	+	+	+	
Recurrent RTI*	-	+	-	+	-	+	+	+	
High forehead	-	-	-	-	+	-	+	-	-
Disturbed sleeping	+	-	-	+	+	+	+	+	
Gastro-esoph. reflux	+	-	+	-	-	+	+	+	
High palate	-	-	-	-	+	-	+	+	Cleft
Delayed visual matur.	+	-	-	+	+	-	+	+	
Microcephaly	-	+	-	+	+	+	-	+	-
Brachyplagiocephaly	-	-	+	-	-	+	+	-	-
Muscle atrophy	-	-	+	-	+	-	+	+	
Hypertelorism	-	-	-	-	+	-	+	-	+
Teeth misalignment	+	-	-	-	-	-	-	-	-
Decreased vision	-	-	-	+	+	-	-	+	
Stagnation episodes	+	+	-	-	-	-	+	+	
Inguinal hernia	+	-	-	-	+	-	+	+	
Decreased fetal movem.	-	-	-	-	-	+	+	-	-
Autistic behavior	+	-	-	+	+	-	-	-	
Aggressive behavior	+	-	+	+	-	+	-	-	
Failure to thrive	-	-	-	-	-	+	-	+	
Pectus excavatum	+	-	-	-	+	+	-	-	
Epilepsy	-	+	-	+	-	-	-	+	
Sit, no support (months)	11	24	24	Attained	14	48	23	>108	
Walk well (months)	26	>72	>156	56	27	65	>55	>108	
Brain MRI abnormalities	-	Wide ventricles, increased T2 signals in nuclei pallidi	Wide ventricles, cystic area in periventricular white matter	Wide ventricles, thin corpus callosum, bilateral loss white matter	Delayed myelination	Megacisterna magna	Delayed myelination, wide ventricles	Wide ventricles and sulci, cerebellar volume loss, abnormal globus pallidus	Corpus callosum agenesis
Severity score	**14**	**14**	**15**	**16**	**19**	**21**	**23**	**25**	**NA**
Severity class	Moderate	Moderate	Moderate	Moderate	Severe	Severe	Severe	Profound	

M: male, F: female. RTI: respiratory tract infections.

Table 2 - Lines of evidence for pathogenicity of variants in POLR2A

Individual	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Variant	p.Pro371Leu	p.Ile457Thr	p.Asn531Ser	p.Tyr669del	p.Gln700*	p.Gln735*	p.Thr736Met	p.Ser755del
cDNA change	c.1112C>T	c.1370T>C	c.1592A>G	c.2006_2008delACT	c.2098C>T	c.2203C>T	c.2207C>T	c.2262_2264delCTC
Phenotype								
Severity score	11	23		14	6	9	25	9
Severity class	Mild	Severe		Moderate	Mild	Mild	Profound	Mild
Variant severity metrics								
Yeast conservation	Identical	Similar	Identical	*NA*	*NA*	*NA*	Similar	*NA*
Amino acid stretch	366-375	448-470	523-532	664-671	689-715	689-715	717-797	717-797
Stretch length	9	22	9	7	26	26	80	80
Desert Z-score	0.8	2.9	0.8	0.5	3.5	3.5	11.9	11.9
Structural evaluation								
Position	Funnel	Catalytic site		Close to quay	Quay	Quay	Quay	Quay
Protein folding				Locally altered	Loss	Loss		Locally altered
Dominant negative		Expected					Expected	
Haploinsufficiency				Putative	Expected	Expected		Putative
Functional evaluation								
Yeast mutation		Leu443Thr	Asn517Ser				Ser713Met	Leu732del
Yeast growth (WT)		Normal	Normal				Normal	Normal
Yeast growth (Δ*dst1*; Δ*sub1*)		Aberrant	Normal				Aberrant	Aberrant
Yeast galactose sensitivity		Yes	No				Yes	Yes
Yeast MPA sensitivity		Yes	No				Yes	Yes
HeLa cell viability		=	=		↓↓↓	↓↓↓	↓↓↓	↓↓↓
Disease causing variant	Possible	**Probable**	*Unknown*	Possible	**Probable**	**Probable**	**Probable**	**Probable**

Table 2 continues on the next page …

…Table 2 continued

Individual	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Variant	p.Met769Thr	p.Ile848Thr	p.Tyr1109His	p.Leu1124Pro	p.Lys1125del	p.Asn1251Ser	p.Arg1603His	p.Pro1767fs
Mutation	c.2306T>C	c.2543T>C	c.3325T>C	c.3371T>C	c.3373_3375delAAG	c.3752A>G	c.4808G>A	c.5298dup
Phenotype								
Severity score	19	15	14	9	10	21	16	11
Severity class	Severe	Moderate	Moderate	Mild	Mild	Severe	Moderate	Mild
Variant severity metrics								
Yeast conservation	Identical	Identical	Similar	Identical	*NA*	Identical	Other	*NA*
Amino acid stretch	717-797	836-877	1086-1135	1086-1135	1086-1135	1218-1254	1603-1604	1766-1769
Stretch length	80	41	49	49	49	36	1	3
Desert Z-score	11.9	5.8	7.1	7.1	7.1	5.0	-0.4	-0.1
Structural evaluation								
Position	Catalytic site	Bridge helix	Trigger loop	Trigger loop	Trigger loop	TFIIS interact.		Hepta repeats
Protein folding					Locally altered			Core normal
Dominant negative	Expected		Expected	Expected				
Haploinsufficiency					Putative			
Functional evaluation								
Yeast mutation				Leu1101Pro		Asn1232Ser		
Yeast growth (WT)				Aberrant		Normal		
Yeast growth (Δ*dst1*; Δ*sub1*)				Aberrant		Normal		
Yeast galactose sensitivity				Yes		No		
Yeast MPA sensitivity				Yes		No		
HeLa cell viability				↓↓↓		=	=	
Disease causing variant	**Probable**	Possible	**Probable**	**Probable**	**Probable**	**Probable**	Possible	**Probable**

WT: wildtype. MPA: mycophenolic acid.