Check for updates

Received: 11 March 2022 | Revised: 23 May 2022 | Accepted: 9 July 2022

DOI: 10.1002/humu.24438

RESEARCH ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Deep intronic *NIPBL de novo* mutations and differential diagnoses revealed by whole genome and RNA sequencing in Cornelia de Lange syndrome patients

Juliette Coursimault[1] | Kévin Cassinari[1] | François Lecoquierre[1] | Olivier Quenez[1] | Sophie Coutant[1] | Céline Derambure[1] | Myriam Vezain[1] | Nathalie Drouot[1] | Gabriella Vera[1] | Elise Schaefer[2] | Anaïs Philippe[2] | Bérénice Doray[3] | Laëtitia Lambert[4] | Jamal Ghoumid[5] | Thomas Smol[6] | Mélanie Rama[7] | Marine Legendre[8] | Didier Lacombe[9] | Patricia Fergelot[9] | Robert Olaso[10] | Anne Boland[10] | Jean-François Deleuze[10] | Alice Goldenberg[1] | Pascale Saugier-Veber[1] | Gaël Nicolas[1]

[1]Normandie Univ, UNIROUEN, Inserm U1245 and CHU Rouen, Department of Genetics and reference center for developmental disorders, FHU-G4 Génomique, F-76000, Rouen, France

[2]Service de Génétique Médicale, Institut de Génétique Médicale d'Alsace (IGMA), Hôpitaux Universitaires de Strasbourg, Strasbourg, France

[3]Service de Génétique Médicale, Centre Hospitalier Universitaire Félix Guyon, Bellepierre Saint Denis, France

[4]Service de Génétique Clinique, CHRU NANCY, F-54000 France, UMR INSERM U 1256 N-GERE, F-54000, Nancy, France

[5]Université de Lille, ULR7364 RADEME, CHU Lille, Clinique de Génétique « Guy Fontaine », and FHU-G4 Génomique, F-59000, Lille, France

[6]Université de Lille, ULR7364 RADEME, CHU Lille, Institut de Génétique Médicale, and FHU-G4 Génomique, F-59000, Lille, France

[7]Institut de Génétique Médicale, CHU de Lille, France

[8]Service de Génétique Médicale, CHU de Bordeaux, Bordeaux, France

[9]INSERM U1211, Université de Bordeaux; Génétique Médicale, CHU de Bordeaux, Bordeaux, France

[10]Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057, Evry, France

Correspondence
Gaël Nicolas, Inserm U1245, UFR Santé, 22, boulevard Gambetta, 76183 Rouen cedex 1, Rouen, France.
Email: gaelnicolas@hotmail.com

Abstract

Cornelia de Lange syndrome (CdLS; MIM# 122470) is a rare developmental disorder. Pathogenic variants in 5 genes explain approximately 50% cases, leaving the other 50% unsolved. We performed whole genome sequencing (WGS) ± RNA sequencing (RNA-seq) in 5 unsolved trios fulfilling the following criteria: (i) clinical diagnosis of classic CdLS, (ii) negative gene panel sequencing from blood and saliva-isolated DNA, (iii) unaffected parents' DNA samples available and (iv) proband's blood-isolated RNA available. A pathogenic *de novo* mutation (DNM) was observed in a CdLS differential diagnosis gene in 3/5 patients, namely *POU3F3*, *SPEN*, and *TAF1*. In the other two, we identified two distinct deep intronic DNM in *NIPBL* predicted to create a novel splice site. RT-PCRs and RNA-Seq showed aberrant transcripts leading to the creation of a novel frameshift exon. Our findings suggest the relevance of WGS in unsolved suspected CdLS cases and that deep intronic variants may account for a proportion of them.

 © 2022 Wiley Periodicals LLC.

KEYWORDS
clustered mutations, Cornelia de Lange syndrome, kataegis, neurodevelopmental disorder, NIPBL, noncoding sequence, whole genome sequencing

1 | INTRODUCTION

Cornelia de Lange syndrome (CdLS) is a rare malformative monogenic syndrome. Five cohesin complex genes have been implicated in CdLS (*NIPBL*, *SMC1A*, *SMC3*, *RAD21*, *HDCA8*) (Deardorff et al., 2007, 2012, 2016; Kline et al., 2018; Krantz et al., 2004; Selicorni et al., 2021; Tonkin et al., 2004). *NIPBL* is the major gene as approximately 70% of suspected CdLS patients with a genetic diagnosis harbor a (likely) pathogenic variant. The other four genes together account for approximately 15% of cases, while the remaining 15% of patients exhibit pathogenic variants in differential diagnosis genes (Kline et al., 2018). *NIPBL*, *RAD21*, *SMC3* are inherited in an autosomal dominant manner whereas *HDAC8* and *SMC1A* are inherited in an X-linked manner. All types of variants have been reported, including truncating variants, missense but also splice variants and larger deletions. The diagnostic yield is about 50% overall, leaving approximately 50% patients unsolved (Gillis et al., 2004; Piché et al., 2019; Selicorni et al., 2007). Heterozygous variants in two additional genes have been reported more recently in a small number of patients with either a CdLS diagnosis or CdLS-like features, namely *MAU2* and *BRD4*, encoding cohesin complex genes (Alesi et al., 2019; Jouret et al., 2022; Olley et al., 2018; Parenti et al., 2020).

Patients with CdLS show a clinically recognizable phenotype characterized by developmental delay (DD) and/or intellectual disability (ID), growth retardation, microcephaly, limb abnormalities and dysmorphic features (Kline et al., 2018). Due to its specific phenotype, the genetic strategy in case of a suspicion of CdLS generally consists in a targeted screening first, by sequencing of a gene panel, ideally including CdLS genes along with genes associated with related disorders. Among them, so-called transcriptomopathies share both pathophysiological and clinical features with CdLS, which itself is considered as a transcriptomopathy and, more precisely, belongs to the group of cohesinopathies. Indeed, all 5 CdLS-causing genes encode critical cohesin complex components involved in chromatin structure maintenance and transcription regulation, in addition to *BRD4* and *MAU2*. In cases with CdLS clinical suspicion, the screening of genes associated with KBG (*ANKRD11*), Rubinstein Taybi (*EP300*), and CHOPS (*AFF4*) syndromes is also recommended, as they are classically considered as putative differential diagnoses because of phenotypic features overlapping with CdLS.

Since the advent of pangenomic sequencing techniques, it has become clear that a first or second-line access to exome sequencing (ES) or whole genome sequencing (WGS) is associated with the highest diagnostic yields for developmental disorders (Wright et al., 2015), because of a huge genetic heterogeneity. For CdLS, however, gene panel sequencing may still be considered in a first line, because (i) the phenotype is recognizable in most cases and (ii) there is a relatively high proportion of mosaic *NIPBL* mutations (~23%), which are not detectable in blood and may be missed by classic 30–40x WGS or 60–120x ES performed from blood samples (Huisman et al., 2013; Nizon et al., 2016). Thus, it is important, before proposing ES or WGS for a classic-CdLS patient, to sequence DNA isolated from other tissues than blood (e.g., saliva or skin) with average depths allowing the identification of mosaics with 10%–20% allelic ratios (ARs).

In classic-CdLS patients negatively screened for the known genes, proposing second-line WGS appears as a promising strategy, with four main hypotheses: (i) a differential diagnosis, i.e., a pathogenic variant in a related disorder or a syndromic developmental disorder with overlapping phenotypic features (ii) a pathogenic variant in a novel gene causing CdLS, (iii) a pathogenic noncoding variant in a known CdLS-causing genes or (iv) a structural variant. Although ES may allow the assessment of the first two hypotheses, WGS offers the opportunity to identify noncoding variants as well as structural variants beyond copy number variants (CNV), and increased sensitivity for CNV detection (Hehir-Kwa et al., 2015). The contribution of deleterious variants in noncoding regions in rare diseases is of recent discovery and remains little explored. Disease-causing variants can occur in all regions outside the coding region, including promoters or enhancers (Cassinari et al., 2020; Shen et al., 2021; Zuin et al., 2017), 5′-untranslated regions (UTRs) (Borck et al., 2006; Labrouche-Colomer et al., 2020; Wright et al., 2021), 3′-UTRs (Dusl et al., 2015), intronic, near-splice regions and deep intronic regions. Some of the latter variants can create neo-exons and destabilize the protein and/or lead to a frameshift (Kim et al., 2020). In CdLS, a few examples of noncoding pathogenic variants have been reported, including four 5′-UTR variants—one altering RNA stability and three predicted to create upstream open reading frames (Borck et al., 2006; Selicorni et al., 2007; Coursimault et al., 2022). WGS also offers the opportunity to study structural variants, allowing for example the identification of complex rearrangement disrupting *NIPBL* (Plesser Duvdevani et al., 2020). Overall, it appears that WGS remains scarcely used in CdLS and could contribute to more diagnoses.

In addition to WGS, RNA-seq allows a genome-wide view of transcripts from the studied tissue that could be well complementary to WGS. Differences in messenger ribonucleic acid (mRNA) relative expression as well as splicing defects and aberrant transcripts can thus be detected. However, the input of RNA-seq in diagnostic procedures remains of rather recent assessment (Cummings et al., 2017; Lee et al., 2020; Rentas et al., 2020; Saeidian et al., 2020).

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

By trio-based WGS, we assessed 5 patients with a clinical diagnosis of classic-CdLS, following negative gene panel sequencing. We performed RNA-seq in two of them, following the identification of strong candidate variants in deep intronic regions. Overall, we managed to identify the likely cause of the syndromic developmental disorder in all five patients.

2 | METHODS

2.1 | Patients enrollment

In this study, we considered probands referred to the Rouen University Hospital molecular genetics department for gene panel sequencing in the context of a clinical suspicion of CdLS, among patients without a (likely) pathogenic variant. The genetics laboratory of the Rouen University Hospital proposes gene panel sequencing with a national multicentric recruitment in France. Medical charts of all patients referred to our centre with a clinical suspicion of CdLS are assessed by an expert clinician and diagnoses are subsequently classified as classical or nonclassical phenotypes before the genetic molecular analysis according to Kline et al. criteria (Kline et al., 2018). Gene panel sequencing targets the coding sequence of all 5 CdLS genes and 17 differential diagnosis genes (*ESCO2*, *CREBBP*, *EP300*, *ANKRD11*, *AFF4*, *KMT2A*, *TAF6*, *SRCAP*, *ARID1A*, *ARID1B*, *SMARCB1*, *SMARCA4*, *SMARCE1*, *SMARCA2*, *SOX11*, *PHF6*, *SETD5*) following custom Agilent QXT capture and Illumina MiSeq sequencing with a approximately 700x average depth of coverage. Patients DNAs are isolated either from blood or from saliva, and sequencing is performed from either or both tissues. RNA sample is not usually required as RNA studies are not part of first-line routine diagnostics. Data are processed following standard procedures for single nucleotide variants (SNVs) and indels, and copy number variants are called using a CANOES-based workflow (Quenez et al., 2021).

In September 2020, among 184 patients referred for a CdLS diagnosis suspicion, the overall diagnostic yield was about 45%. To identify the molecular defects underlying CdLS in patients negatively screened by the above-mentioned procedures, we performed trio-based WGS in patients fulfilling the following restrictive criteria: (i) a typical (or classic) CdLS diagnosis after clinical expertise, (ii) a negative family history, (iii) absence of known molecular cause after analysis of our gene panel, performed from blood and saliva samples with a high sequencing depth (~700x) as described above, (iv) unaffected parents, with blood/DNA samples available in sufficient quantity and quality, (v) RNA of index cases available (blood sample on Paxgene tube), for possible further investigation. All patients gave informed written consent for genetic analyses. From an initial list of 102 CdLS patients without a coding pathogenic variant, 10 patients presented with a diagnosis of sporadic typical CdLS after exclusion of fetuses, and 5 patients fulfilled the above-mentioned criteria. Some other patients were lost to follow-up, while it was not possible to get DNA samples from parents or a saliva sample for the proband (when initial screening was based on blood), or an RNA sample for the other patients.

2.2 | WGS and variant detection

Whole-genome sequencing was performed by the Centre National de Recherche en Génomique Humaine (CNRGH, Institut de Biologie François Jacob, CEA, Evry, France). After a complete quality control, genomic DNA (1 µg) was used to prepare a library for WGS, using the Illumina TruSeq DNA PCR-Free Library Preparation Kit (Illumina Inc.), according to the manufacturer's instructions. After quality control and normalization, qualified libraries were sequenced on a NovaSeq 6000 platform from Illumina (Illumina Inc.), as paired-end 150 bp reads. Samples were pooled on a NovaSeq 6000 S4 flowcell to target a minimal average sequencing depth of 30x.

Sequence quality parameters were assessed throughout the sequencing run and FASTQ file were generated for each sample. FastQ sequences were aligned on human genome hg19 using the BWA-mem program (v0.7.17). The GATK tools (v4.0.6.0) were then used for the postprocessing of the bam files (BQSR and deduplication). SNVs and short insertions and deletions (indels) were called using the GATK HaplotypeCaller tool and annotated using SNPEff and SNPSift. Structural variant analysis was performed using CANVAS (v.1.39.0) (Roller et al., 2016) and MANTA (v.1.6.0) (Chen et al., 2016) for detection of CNVs, translocations, insertions, inversions. Mobile element insertions were detected using MELT (v.2.1.5) (Gardner et al., 2017). All SVs variants detected by CANVAS, MANTA and MELT were then annotated using AnnotSV (v2.2) (Geoffroy et al., 2018). Candidate variants were confirmed by Sanger sequencing in all probands and their parents.

To accurately detect *de novo* SNVs and indels (*de novo* mutations [DNM]) in each proband, we applied complementary methods and filtration steps. We used Deepvariant with WGS model to call SNVs and indels in probands and both parents (Poplin et al., 2018). Individual GVCF were merged via glnexus with DeepVariantWGS preset to produce a multisample VCF. *De novo* candidates were obtained after filtration on GT (genotype), DP (depth), GQ (genotype quality) and VAF (variant allele fraction) fields. Filters used to call *de novo* candidates from multi sample VCF (5 trios) were: Ref genotype in parents, DP >10 and GQ >29 in proband and both parents, VAF >0.3 in proband, VAF(proband)/VAF(parents) >4, and VAF(proband)/VAF(controls) >5. For a given proband, controls corresponded to the four other probands and their parents (12 individuals). *De novo* candidate variants were manually reviewed via a custom Integrative Genomics Viewer (IGV)-based filtration interface (https://github.com/francois-lecoquierre/genomic_shortcuts/). As a quality control and to detect potential outliers regarding the count of *de novo* variants, and because the paternal age at conception is the main determinant of this biological phenotype (Jónsson et al., 2017), we plotted the number of *de novo* variants in each proband against paternal age at conception.

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.3 | Noncoding variants annotation

Variations in the 5′-UTR regions were annotated with the 5utr ['suter'] tool, allowing the search for uORF creations (https://github.com/leklab/5utr). Splicing was assessed using the SpliceAI tool which is a deep neural network that predicts splice junctions from a pre-mRNA transcript sequence (Jaganathan et al., 2019) (lastly assessed, june 2021). MaxEntScan, NNSPLICE, GeneSplicer and SpliceSiteFinder-like, as provided in the Alamut Visual software, were used to predict whether selected variants affected splice sites (Pertea et al., 2001; Reese et al., 1997; Yeo & Burge, 2004; Zhang, 1998).

2.4 | Detection of candidate disease-causing variants

WGS was conducted on trios composed of the affected child and both unaffected parents. Coding regions were analyzed first. Variants were selected under several filtering scenarios and were interpreted according to ACMG-AMP recommendations (Richards et al., 2015). We analyzed variants regarding the following filtration scenarios (i) *de novo* variants, (ii) variants present and/or pathogenic in patient-derived databases (Clinvar, Denovo-db) (lastly assessed, june 2021), (iii) variants that are very rare in the general population (minor allele frequency <0.001) segregating according to autosomal recessive or X-linked inheritance (Karczewski et al., 2020). Remaining gene variants underwent further prioritization and manual interpretation. Secondarily, the noncoding variants were analyzed by focusing on *de novo* candidate variants in introns and 5′-UTR/3′-UTR regions and then extended the analysis to other inheritance hypotheses.

For structural variants, we filtered out those with more than one occurrence in the Database of Genomic Variant gold standard (MacDonald et al., 2014) or gnomAD-SV database (Collins et al., 2020) (both assessed in May, 2021).

2.5 | RNA-seq

Total RNAs from whole blood were extracted with PAXgene blood RNA kit according to the manufacturer's recommendations (Qiagen PreAnalytiX GmbH) and stored at −80°C until use. The quality and quantity of RNA were assessed using the 4200 TapeStation (Agilent Technologies) and the Qubit 3.0 device (Thermo Scientific). Only RNA samples with a minimal RNA integrity number of 7 were used for subsequent experiments. Libraries were prepared using the NEBNext Ultra II Directional RNA Library Kit for Illumina (New England Biolabs) kit and High-throughput sequencing of the libraries was performed on an Illumina NextSeq 500 (Illumina) using 2*75 bp sequencing to generate 60M read pairs on average per sample. Bioinformatics analysis was carried out using nf-core/RNA-seq v3.1 analysis pipeline to generate multi quality control report that uses the STAR v2.6.1d and SALMON v1.4.0 tools for alignment (Ewels et al., 2020). Visual exploration of the BAM files was performed with the IGV tool from the Broad Institute.

2.6 | Targeted RNA analyses

Complementary DNA (cDNA) was synthesized from total RNA (collected as described above), using a high-capacity cDNA Reverse transcription kit (Applied Biosystems) with RiboLock RNase inhibitor (Thermo Fisher scientific) which was further amplified to obtain PCR products using specific primers. ThermoPrime Taq DNA polymerase from Thermo scientific was used for PCR amplification. We used the following respective primers (Figure S1). In the case of patient 4: specific primers on either side of the suspected aberrant exon were designed (PCR 1: exon 31 forward primer: 5′-CTCCAACTCCACA CAATGACA-3′ and exon 34 reverse primer: 5′- GCTGGGGTC TTATTTTGCTGA-3′). Primers were also picked within the aberrant exon (PCR 2: exon 31 forward primer: 5′-CTCCAACTCCACA CAATGACA-3′ and exon 32 reverse primer 5′-TTGGGAGGCTGA GGAAAGAG-3′); PCR 3: exon 32 forward primer 5′-TCTTTCCTCAG CCTCCCAAG-3′ and exon 34 reverse primer 5′-GCTGGGGTCT TATTTTGCTGA-3′). In the case of patient 5: specific primers on either side of the suspected aberrant exon were designed (PCR 1: exon 7 forward primer: 5′-AGACATGGTTCAAGTGAGGACT-3′ and exon 9 reverse primer: 5′-ACATTGCCGCTTTCTCACTC-3′). Primers were also picked within the aberrant exon (PCR 2: exon 7 forward primer: 5′-AGACATGGTTCAAGTGAGGACT-3′ and exon 8 reverse primer 5′-TGTGGTCTTCTCTTTCTCCCT-3′; PCR 3: exon 8 forward primer 5′-AGGGAGAAAGAGAAGACCACA-3′ and exon 9 reverse primer 5′- ACATTGCCGCTTTCTCACTC-3′). PCR products were ultimately separated on a 2.5% agarose gel and validated by Sanger sequencing.

3 | RESULTS

We included five patients with a clinical diagnosis of classic CdLS assessed by a clinical expert, and negative panel sequencing performed on DNA isolated from both blood and saliva. In addition, three of them also had negative gene panel sequencing from a bulk skin biopsy sample. We performed trio-based WGS on DNA isolated from blood. Patients were aged from 5 to 18 years. Summary phenotypic data are presented in Table 1 and further described below.

The average depth of coverage was of 44x. SNVs, short insertions and deletions (indels) as well as structural variants were analyzed with a focus on known Mendelian genes (all inheritance patterns, OMIM database and home-made curated extension) and at the genome-wide level for *de novo* mutations.

We identified a likely pathogenic/pathogenic (LP/P) variant in all five patients. All were *de novo* heterozygous SNV or indels. Three patients exhibited a *de novo* LP/P coding variant in a gene causing another developmental syndrome, namely *TAF1*, *SPEN*, and *POU3F3*

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Clinical description of the five patients with typical CdLS analysed by WGS

Phenotype	Patient 1 *SPEN* *c.4345G>T; p.(Glu1449*)*	Patient 2 *POU3F3* *c.1084C>A; p.(Arg362Ser)*	Patient 3 *TAF1* *c.4748A>G; p.(Tyr1583Cys)*	Patient 4 *NIPBL* *c.5862+3487C>T*	Patient 5 *NIPBL* *c.869-640G>C*
Age/sex	18 years; female	10 years; female	5 years; male	15 years; female	13 years; male
Antenatal and neonatal period	Hypotrophy	Birth weight (11th P); Height (2nd P)	Severe IUGR and then hypotrophy	Hypotrophy	Severe IUGR and then hypotrophy
Psychomotor development	Global delay, mild to moderate ID (no formal diagnosis)	Global delay (no language, not walking at the age of 4 years); hypotonia	Global delay, coordination problems	Global delay (walked at 28 months, no language acquisition at the age of 15 years); anxiety, OCD	Moderate global delay; attention deficit disorder, walked at the age of 21 months, first words at the age of 3 years, has improved lexical fields with short sentences; adapted school
Organ malformations	Surgical treatment of aortic coarctation; strabismus	None	Scrotum in shawl	None	Heart defect (atrial septal defect)
Limb abnormalities	Brachymetacarpia of the 1st ray, Clinodactyly of the 5th fingers; Walking with walker; Hypertonia of the extremities; severe hip dysplasia	None	Relative brachymetacarpia of the 1st ray of the feet; wide thumbs	Clinodactyly of the 5th fingers; Brachymetacarpia of the 1st ray; III–V Brachymetatarsia	Micromelia with brachymetacarpia of the 1st ray
Head circumference	Microcephaly	Microcephaly	Microcephaly	Microcephaly	Microcephaly
	(−4 SD)		(−4.5 SD)	(−4.2 SD)	(−4 SD)
Growth/eating disorders	Growth retardation	Growth retardation	Eating difficulties (especially; solid food)	Growth retardation	Early eating disorders, GERD, enteral feeding tube
Dysmorphic features	Typical, ptosis, blepharophimosis, pectus excavatum; Stretch marks and hypertrophic scars	Typical; cupped ears	Typical	Typical	Typical

Abbreviations: CdLS, Cornelia de Lange syndrome; GERD, gastroesophageal reflux disease; ID, intellectual disability; IUGR, intrauterine growth restriction; OCD, obsessive-compulsive disorder.

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(Figure S2), while two patients showed a *de novo* deep intronic *NIPBL* variant predicted to create a frameshift neo-exon and thus resulting in a likely loss of function. We further confirmed the consequences of intronic *NIPBL* DNM by targeted RNA analyses and RNA-seq.

3.1 | Patient 1. Chr1(GRCh37):g.16257080G>T; NM_015001.2(SPEN):c.4345G>T, p.(Glu1449*)

3.1.1 | Clinical summary

Patient 1 is an 18-year-old girl with mild-to-moderate ID. She was born prematurely at 26 weeks of gestation in the context of maternal fever with a weight of 675 g (19.37th percentile), a length of 31 cm (12.37th percentile) and an OFC of 22.5 cm (17.91th percentile). She benefited from surgical treatment of aortic coarctation. Sucking, swallowing and digestive outcome after the neonatal period were normal. She exhibited growth retardation. At 5½ years, she weighed 12.8 kg (−2.9 SD) for a height of 99 cm (−2.5 SD). At the age of 12½ years upon last visit, height was 144 cm (−1.1 SD), weight 36 kg (−1 SD), body mass index (BMI) 17.4, even though she underwent growth hormone treatment between 3 and 9 years of age. She presented microcephaly (51 cm; −4 SD). She also had global delay. Brain magnetic resonance imaging (MRI) showed leukomalacia. She presented a pyramidal syndrome, sharp and polykinetic reflexes, and spasticity. She underwent surgery for severe hip dysplasia and multiple tenotomies associated with botulinum toxin injections. At last visit, she could walk with walker assistance because of hypertonia of extremities. There was hypotonia of the oral sphere with difficulties of elocution, mastication and drooling. She had been treated by physiotherapy and speech therapy. She was in a regular school in a class adapted for children with special needs. She showed brachymetacarpus of the 1st ray and had typical CdLS dysmorphism with ptosis, strabismus, blepharophimosis, arched eyebrows, short nose with anteverted nostrils, thin upper lip, flat and prominent philtrum and downturned corners of the mouth (Figure 1). Hearing was normal. We also noted the presence of stretch marks with thin skin. Kline consensus clinical score was 14. CdLS gene panel sequencing was negative on blood, saliva and skin biopsy.

3.1.2 | WGS analysis

Interpretation of coding variants from the WGS data revealed three heterozygous DNM in the *SPEN* gene: a heterozygous (AR = 42.6%) truncating variant (NM_015001.2:c.4345G>T; p.(Glu1449*), ClinVar submission SUB10575763), a synonymous variant (c.3642C>T, p.(Pro1214Pro), AR = 45%), and a missense variant (c.3656C>T, p.(Thr1219Ile), AR = 41.3%). The truncating variant was prioritized. This variant is absent from the gnomAD database. The probability of loss-of-function intolerance (pLI) of this gene is 1 in the gnomAD browser. *SPEN* is reported to be enriched in *de novo* and truncating

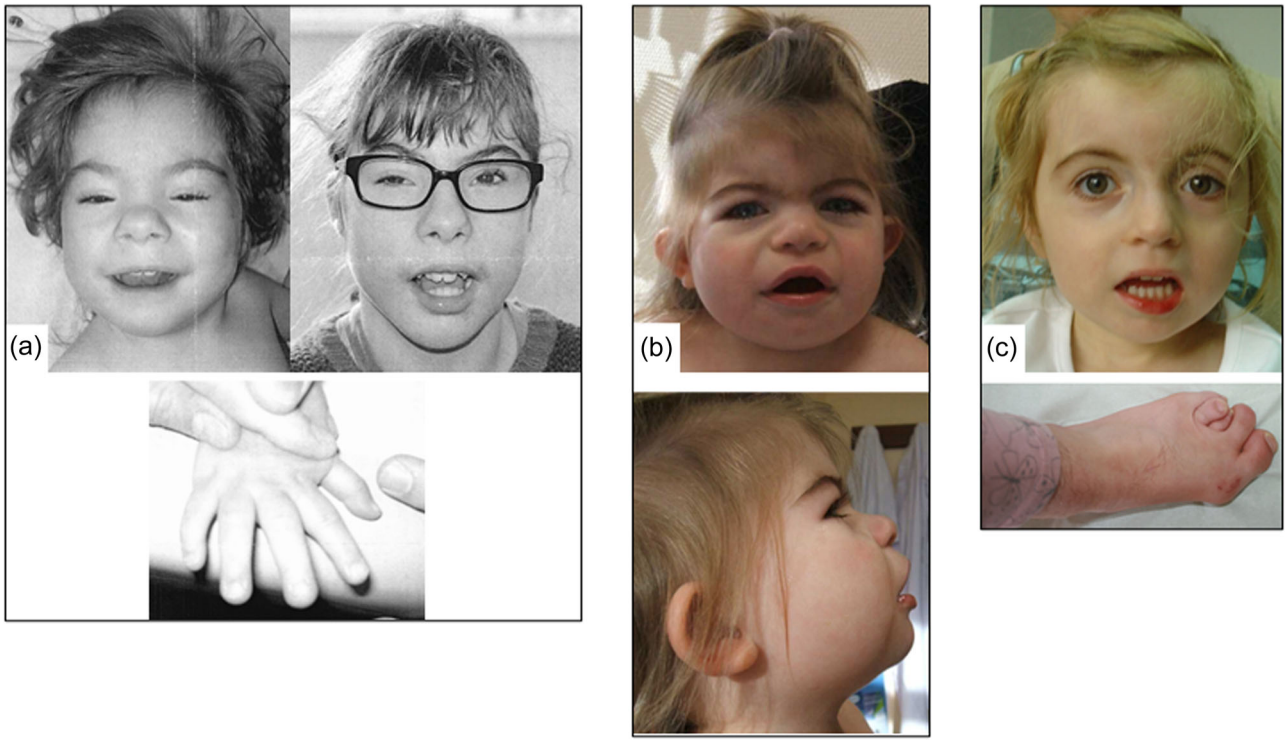

FIGURE 1 Clinical photographs of 3 individuals with clinically suspected CdLS. (a) Patient 1, front view at 1 and 11 years of age and photograph of her right hand with brachymetacarpus of the 1st ray; (b) Patient 2, front and lateral view at 2 years of age; (c) Patient 4, frontal view at the age of 5 years and photograph of the left foot

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

variants in patients with a developmental disorder (Radio et al., 2021; Wang et al., 2020). It encodes a transcriptional repressor with a major role in the initiation of X-chromosome inactivation (Dossin et al., 2020). Pathogenic *SPEN* variants have been very recently associated with a neurodevelopmental disease in two clinical cohorts (MIM# 619312) (Radio et al., 2021; Wang et al., 2020). The patient's phenotype appeared to be consistent with the descriptions in the literature: DD, ID, hypotonia, abnormal pyramidal signs, central nervous system (CNS) abnormalities, cardiac malformation, ophthalmologic involvement, and dysmorphism. Taken together, the c.4345G>T; p.(Glu1449*) *SPEN* variant was classified as pathogenic (Class 5) according to the ACMG-AMP guidelines (Richards et al., 2015). This allowed us to revise the diagnosis and to conclude to a neurodevelopmental disease linked to the haploinsufficiency of the *SPEN* gene in this patient. The role of the other two *SPEN* variants remains unclear. BAM files showed that they occurred on the same allele, but it was not possible to phase them with the truncating variant. The synonymous variant is not predicted to affect splicing. The missense variant was predicted deleterious by 10/18 bioinformatics software as assessed on the Varsome website in June 2021. They both remain of unknown significance. Of note, no other candidate variant was prioritized among SNV/indels and structural variants.

3.2 | Patient 2. Chr2(GRCh37): g.105473052C>A; NM_006236.3(POU3F3):c.1084C>A, p.(Arg362Ser)

3.2.1 | Clinical summary

Patient 2 is 5-year-old girl who was born at 38 weeks of gestation. Birth length (44.5 cm; 2.24th percentile) and head circumference (31.5 cm; 5th percentile) were in the lower range, and her weight was 2560 g (11.56th percentile). Her development was delayed and associated to global hypotonia. She sat at 12 months and still did not walk at 4 years of age upon last visit. She had neither gastro-esophageal reflux disease (GERD) in infancy nor eating disorder. She showed growth retardation and microcephaly (87 cm, −3.5 SD). She had no language at the age of 4. The malformative assessment was normal. She had no limb abnormalities. Dysmorphic features included hirsutism, a low frontal hairline, arched eyebrows with synophris, long eyelashes, short nose with anteverted nares, flat and prominent philtrum, downturned corners of the mouth and cupped ears (Figure 1). Kline consensus clinical score was 13. CdLS gene panel sequencing was negative on blood and saliva.

3.2.2 | WGS analysis

Interpretation of coding variants from the WGS data revealed a heterozygous missense DNM in the *POU3F3* gene (NM_006236.3:c.1084C>A; p.(Arg362Ser), AR = 53%) (ClinVar Submission SUB10575785), a gene associated with Snijders Blok-Fisher syndrome (MIM# 618604). The variant is absent from the gnomAD database and is predicted to be deleterious by 19/22 bioinformatics tools as assessed on the Varsome website in June 2021. The mutation was not previously reported in the ClinVar database but two other patients harbored distinct missense substitutions at the same codon c.1085G>T; p.(Arg362Leu) (Snijders Blok et al., 2019). This variant is located in one of the two known functional domains of POU3F3: the POU-specific (POU-S) domain where a clustering of missense variants has been reported. *POU3F3* encodes a transcription factor belonging to the POU family. It is involved in the regulation of many key processes in CNS development, including cortical neuron migration, specification and production of upper layers and neurogenesis. Consistent with the previous studies, Patient 2 also had hypotonia, DD, ID, morphological features with atypical cup ears, smooth philtrum, and open gendarme hat mouth. Taken together, the c.1084C>A; p.(Arg362Ser) *POU3F3* variant was classified as pathogenic (Class 5) according to the ACMG-AMP guidelines (Richards et al., 2015). The identification of this variant allowed us to make a diagnosis of Snijders Blok-Fisher syndrome in this patient, which also led to a differential diagnosis of CdLS. Of note, no other candidate variant was prioritized among SNV/indels and structural variants.

3.3 | Patient 3. ChrX(GRCh37):g.70644083A>G; NM_004606.5(TAF1):c.4748A>G, p.(Tyr1583Cys)

3.3.1 | Clinical summary

Patient 3 is a 5-year-old boy who was born at 35 weeks and 6 days of gestation with congenital torticollis. Pregnancy was marked by IUGR with hyperechogenic small intestine. Prenatal array CGH and *CFTR* gene screening revealed no abnormalities. Birth measurements were as follows: weight 1290 g (<1st percentile), length 38 cm (<1st percentile) and OFC 28 cm (<1st percentile). At the age of 20 months, his weight was 9.1 kg (1st percentile), his height was 77 cm (−2.5 SD) and his OFC was 43 cm (−4.5 SD). He showed global delay, clumsiness and hearing loss. He had a horseshoe-shaped kidney. Cardiac ultrasound was normal. He exhibited a shawl scrotum. He showed brachymetacarpia of the 1st ray of the feet and also wide thumbs. At the last examination at the age of 4 years 1 month, his weight was 12.6kg (BMI 0.4 SD) and his height 87.5 cm (−3.65 SD). OFC was 45 cm (−4.95 SD). He just started to sit unaided, tried to stand, and to walk with a walker. He could say only 3 words (mum, dad, dog) even with hearing aid. Gastroesophageal reflux remained a problem as well as constipation. He presented difficulties to eat solid food. Brain MRI at the age of 23 months showed microcephaly with suboptimal white matter myelination, a small corpus callosum and small basal ganglia as well as some degree of simplification of the cortical gyration and passive ventriculomegaly. Dysmorphic features associated plagiocephaly, arched eyebrows with synophris, long eyelashes, short nose with anteverted nares, prominent philtrum, thin upper lip with downturned corners of the mouth and large ears. He

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

had an achromic and a café-au-lait spot. Kline consensus clinical score was 13. CdlS gene panel sequencing was negative on blood and saliva.

3.3.2 | WGS analysis

Interpretation of coding variants from the WGS data revealed a heterozygous NM_004606.4:c.4748A>G;p.(Tyr1583Cys) (AR = 100%) DNM in the *TAF1* gene (ClinVar Submission SUB10575808). This variant is absent from the gnomAD database and predicted to be deleterious by 16/18 bioinformatics software as assessed on the Varsome website in June 2021. Many *TAF1* missense variants have been reported to cause an X-linked neurodevelopmental disease (MIM# 300966) (Cheng et al., 2020; Hurst et al., 2018; O'Rawe et al., 2015). More than 30 families, including both male and female patients, have been described. Of them, two patients had been given first a clinical diagnosis of CdLS (Cheng et al., 2020). The phenotype associated with this disease seems to be compatible with the phenotype of Patient 3 including hypotonia, DD predominantly in language, ID, autism spectrum disorders, clumsiness, IUGR, postnatal growth retardation, feeding difficulties, microcephaly, as well as dysmorphic features.

A population-scale study ranked *TAF1* 53rd among the top 1003 constrained human genes (Samocha et al., 2014) with a maximal pLI (pLI = 1 in gnomAD). *TAF1* has recently been reported as a neurodevelopmental gene enriched in *de novo* variations (Martin et al., 2021). This highly conserved gene plays a major role in the establishment of protein complexes associated with RNA Pol 2 transcription. Missense variants, distributed all along the gene, have been exclusively identified in this disease, suggesting a loss-of-function mechanism for these missense variants. Based on standards and guidelines by the ACMG-AMP, the variant was classified as likely pathogenic (Class 4). We concluded that Patient's 3 disease was likely attributable to this *de novo* missense c.4748A>G;p.(Tyr1583Cys) variant of the *TAF1* gene, again leading to a differential diagnosis of CdLS. Of note, no other candidate variant was prioritized among SNV/indels and structural variants.

3.4 | Patient 4. NC_000005.9:g.37031001C>T, c.5862+3487C>T NIPBL variant

3.4.1 | Clinical summary

Patient 4 is a 15-year-old girl with ID, anxiety and obsessive-compulsive disorder. She benefited from a prenatal karyotype because of nuchal translucency showing a normal 46,XX results. She was born at 36 weeks of gestation. Birth parameters were 2770 g (27th percentile) for weight and 48 cm for height (43rd percentile), suggesting neonatal hypotrophy. OFC was 32 cm (10th percentile). Early motor milestones were delayed: she sat at 12 months and walked first unaided at 28 months. Currently at the age of 15 years, she has no oral language, she communicates poorly with sign language. She shows growth delay with a weight of 35 kg (−2.4 SD) and height of 143 cm (−3.2 SD) and presents severe microcephaly (49.5 cm, −4.2 SD). In addition, she has limb abnormalities, with clinodactyly of the 5th fingers, brachymetacarpia of the 1st ray and III–V brachymetatarsia (Figure 1). Cardiac and abdominal ultrasound were normal. She has no hearing loss. She has dysmorphic features including a low frontal hairline, arched eyebrows, long and prominent philtrum, downturned corners of the mouth, thin upper lip. Kline consensus clinical score was 14. CdLS gene panels sequencing was negative on blood, saliva and skin biopsy. Screening for *MECP2*, *FOXG1* and array CGH were normal.

3.4.2 | WGS analysis

Neither any candidate single nucleotide/indel variant, nor any candidate structural variant was identified following the analysis of coding regions. Analysis of the noncoding DNMs highlighted a heterozygous deep intronic DNM in the *NIPBL* gene (NM_133433.3:c.5862+3487C>T, intron 32, AR = 51.4%) (ClinVar Submission SUB10575836). This variant is absent from the gnomAD database. The variant is predicted to affect splicing by creating a novel splice donor site according to MaxEntScan, NNSPLICE, GeneSplicer and SpliceSiteFinder-like (Figure 2). The SpliceAI tool predicted a donor gain with a Δ score of 0.19 (Δ scores range from 0 to 1 and a detailed characterization is provided for 0.2, 0.5, 0.8 cutoffs). In the vicinity of this variant, several putative cryptic acceptor sites are also predicted in both wild-type and mutant contexts, of which at least one could be used, putatively leading to the creation of a novel exon between natural exons 32 and 33 (Figure 2). We performed RNA-seq from proband's whole blood collected in a PAXgene tube. Following RNA-seq, inspection of NIPBL alignments showed the presence of abnormal splice products mapping to intron 32 at the expected positions, showing the inclusion of the predicted 118-bp novel exon (Figure 2), with aberrant junctions to exon 32 and exon 33, respectively (chr5: 37,030,882_37,030,999). The inclusion of this neo-exon was further confirmed by RT-PCR and sequencing (Figure 2 and Figure S3.A). This insertion is out of frame and results in a premature stop codon (r.5862_5863ins118;p.Asn1954fs*50). In RNA-seq data, 20 junctions supported the existence of the neo-exon (14 between exon 32 and the neo-exon and 6 between neo-exon and exon 33), compared to 79 normal exon 32—exon 33 junctions, suggesting some degradation of neo-exon-containing transcripts by nonsense-mediated decay (NMD), and/or partial use of the newly created splicing site. In the coding sequence of *NIPBL*, one heterozygous common SNV, in exon 10 (rs3822471), was available to assess allele-specific expression. The AR was 52% at this position, suggesting that the inclusion of this neo-exon is not associated with strong degradation of transcripts containing this neo-exon by NMD. Thus, we conclude that the c.5862+3487C>T variant is associated with a significant but partial inclusion of a frameshift neo-exon, leading to a partial loss-of-function allele.

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

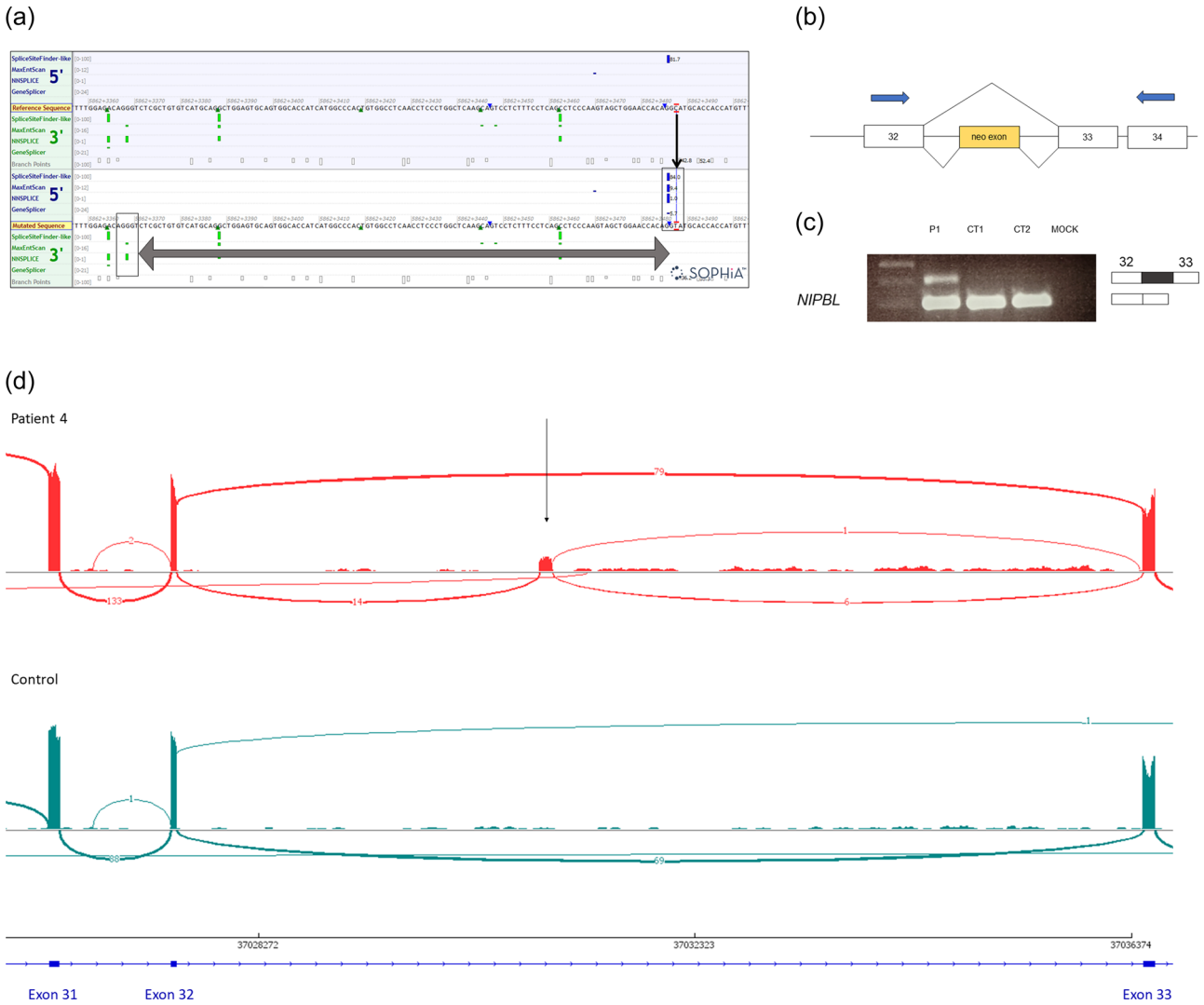

FIGURE 2 Assessment of *de novo* deep intronic *NIPBL* mutation in patient 4. (a) Splicing predictions. According to MaxEntScan, NNSPLICE, GeneSplicer and SpliceSiteFinder-like tools available in the Alamut visual software, the variant is predicted to affect splicing by creating a novel splice-donor site. The black arrow shows the mutation. The splice acceptor and donor site used are surrounded by a black square. The grey double arrow symbolizes the new exon potentially synthesized. (b) Schematic representation of the splicing alteration responsible for the creation of a new aberrant exon between exons 32 and 33. (c) Migration of RT-PCR products on agarose gel showing the presence of an extra band in the patient, compared to two controls (CT1 and CT2). This band migrates at approximately 400 bp. (d) Sashimi plots generated following the alignment of RNA-seq reads on the Integrated Genome Viewer (IGV) showing abnormal *NIPBL* splicing in patient 4 relative to a control

3.5 | Patient 5. NC_000005.9:g.36975238G>C, c.869-640G>C NIPBL variant

3.5.1 | Clinical summary

Patient 5 is a 13-year-old boy, born at 37 weeks of gestation after a pregnancy marked by IUGR and hydramnios. Birth measurements were abnormal with a weight of 1920 g (<1st percentile), height of 45 cm (5.44th percentile) and OFC of 29 cm (0.04th percentile). He presented in the perinatal period with eating disorder and GERD, improved by enteral feeding tube. Hearing was normal. Upon last visit at 13 years, he showed microcephaly (OFC = 50 cm, −4.5 SD), presented moderate delay and ADHD. He also had micromelia with brachymetacarpia of the 1st ray. Cardiac ultrasound revealed an atrial septal defect. He presented facial dysmorphism including arched eyebrows with synophris, blepharophimosis, short nose with anteverted nares, flat and prominent philtrum, thin upper lip and micrognatism. Kline consensus clinical score was 13. CdLS gene panel sequencing was negative on blood, skin biopsy and saliva.

3.5.2 | WGS analysis

Neither any candidate single nucleotide/indel variant, nor any candidate structural variant was identified following the analysis of coding regions. Surprisingly, Patient 5 presented 6 distinct intronic DNM in the *NIPBL*

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

gene (NM_133433.3:c.610+287T>C [AR = 0.48%]; c.610+1339G>A [AR = 0.46%]; c.869-640G>C [AR = 0.57%]; c.3121+3010T>C [AR = 0.40%]; c.3305-479T>G [AR = 0.43%]; c.3856-1054G>C [AR = 0.54%]). These variants were distributed throughout the gene, in several introns (two variants in intron 6 and the others, respectively, mapping to introns 8, 10, 11, and 16) (Figure 3). Overall, Patient 5 harbored a total of 69 DNMs (61 SNVs and 8 indels), which was in the expected range given the father's age at conception (Figure S4). Therefore, these results were in favor of a mutational cluster specifically localized in the *NIPBL* gene. Among these 6 *NIPBL* variants, only one (NM_133433.3:c.869-640G>C,

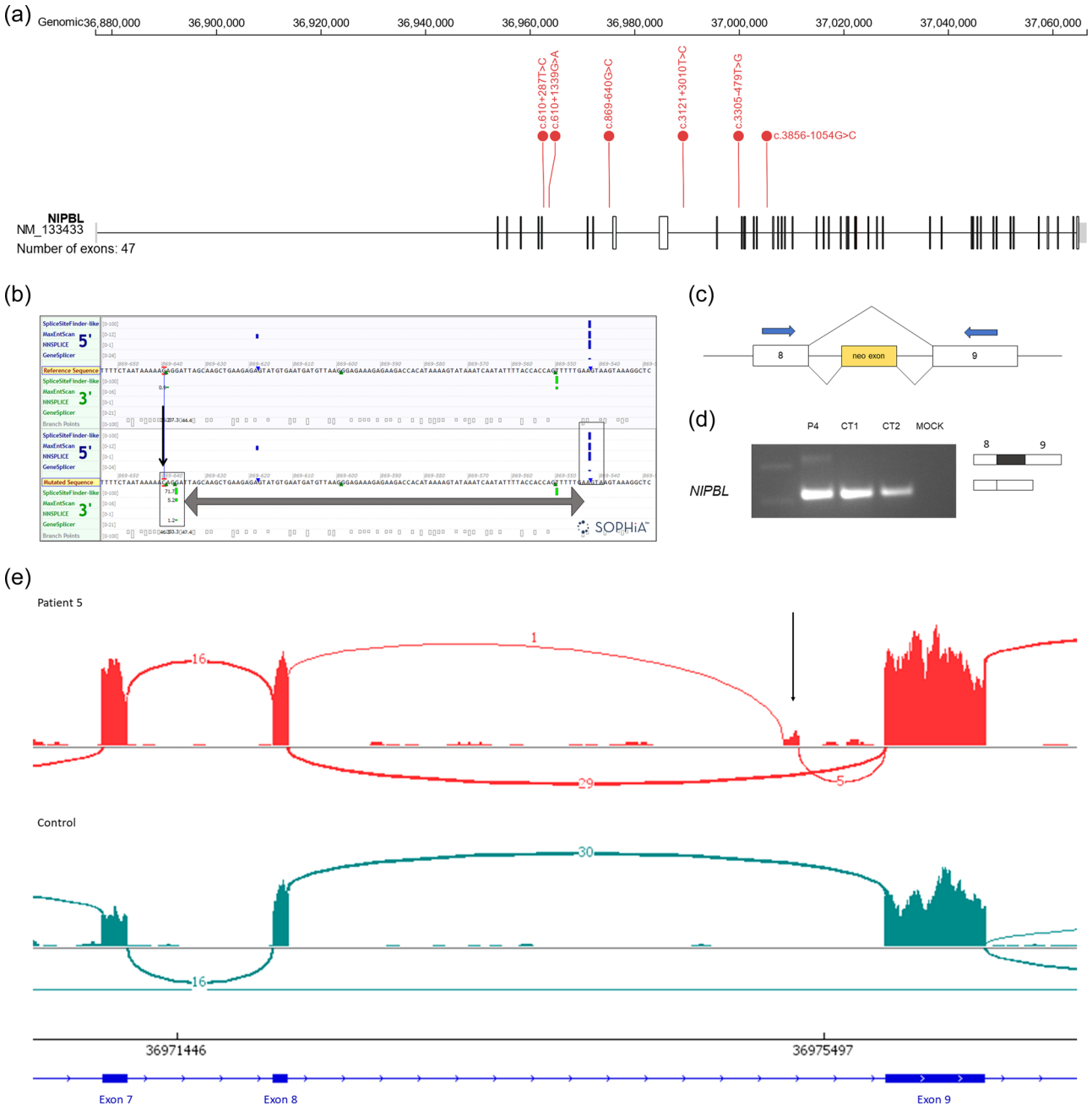

FIGURE 3 Assessment of *de novo* deep intronic *NIPBL* mutation in patient 5. (a) Distribution of the 6 intronic variants along the *NIPBL* mRNA in patient 5. GenomePaint (St. Jude Cloud; https://genomepaint.stjude.cloud/) has been used for this representation. (b) Splicing predictions. According to MaxEntScan, NNSPLICE, GeneSplicer and SpliceSiteFinder-like tools available in the Alamut visual software, the variant is predicted to affect splicing by creating a new splice acceptor site leading to a theoretical neo-exon. The black arrow shows the variant. The splice acceptor and donor site used are surrounded by a black square. The grey double arrow symbolizes the new exon potentially synthesized. (c) Schematic representation of the splicing alteration responsible for the creation of a new aberrant exon between exon 8 and exon 9. (d) Migration of RT-PCR products on agarose gel showing the presence of an extra band in the patient, compared to two controls (CT1 and CT2). This band migrates at approximately 320 bp. (e) Sashimi plots following the alignment of RNA-seq reads on IGV showing abnormal *NIPBL* splicing in patient 5 relative to a control. mRNA, messenger ribonucleic acid

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

intron 8, absent from gnomAD) showed strong predictions of effect on splicing, with a predicted creation of a new splice acceptor site (ClinVar Submission SUB10575921). The spliceAI tool predicted an acceptor gain with a Δ score of 0.83. Together with predictions of cryptic donor sites in the surrounding regions in both mutant and WT contexts, we hypothesized a possible creation of a neo-exon at position chr5:36975240_36975335 (Figure 3). We performed RNA-seq from the proband's whole blood collected in a PAXgene tube. Following RNA-seq, *NIPBL* analysis of aligned reads showed the presence of abnormal splice products mapping to intron 8 at the expected positions, showing the inclusion of the predicted 95-bp novel exon (Figure 3), with aberrant junctions to exon 8 and exon 9, respectively. In RNA-seq data, 6 junctions supported the existence of the neo-exon (1 between exon 8 and the neo-exon and 5 between neo-exon and exon 9), compared to 29 normal exon 8—exon 9 junctions and we could not observe any aberrant splice junction in the surrounding regions of the other *de novo* mutations identified in the other *NIBPL* introns. The exonized intronic sequence is out of frame and results in a premature stop codon within the exonized sequence (r.868_869ins95, p.Gly291fs*3). The neo-exon inclusion was further confirmed by RT-PCR and cDNA sequencing (Figure 3 and Figure S3.B). Unfortunately, there was no SNV in the *NIPBL* coding sequence, thus precluding the assessment of allele specific expression.

4 | DISCUSSION

Following the selection of 5 patients with a classic-CdLS presentation and negatively screened for the known genes, WGS analysis ended up with a probable genetic cause in all 5 patients. Nevertheless, these selected patients may not represent all patients negatively screened for CdLS genes and it is unlikely that WGS would provide such a high diagnostic rate using broader inclusion criteria. Follow-up analyses on additional patients, using less stringent inclusion criteria, may lead to a lower diagnostic rate and a higher number of variants of unknown significance. Interestingly, among the 5 (likely) pathogenic *de novo* variations, 3 were coding and are hence theoretically detectable by simplex or trio-based ES, and 2 were noncoding, highlighting one of the inputs of WGS as compared to ES.

The analysis of noncoding regions appears to be relevant, as an increasing number of reports shows a noncoding cause of neurodevelopmental diseases, sometimes after years of diagnostic odyssey (Cassinari et al., 2020; Labrouche-Colomer et al., 2020; Wright et al., 2021). In our study, we identified two different *de novo* intronic variants in two patients, both predicted to result in the creation of a novel splice site. We were able to confirm the pathogenicity of these variants by RNA-seq and RT-PCR followed by Sanger sequencing. RNA-seq was particularly useful for the patient with multiple deep intronic variants, allowing a global view, compared to targeted RT-PCR procedures.

Overall, deep intronic variants creating a frameshift neo-exon are a classic disease-causing mechanism, albeit extremely rarely reported in Mendelian disorders. It has been described for example in the *DMD* gene where the inclusion of a pseudo-exon was responsible of a milder Becker's muscular dystrophy phenotype (Cummings et al., 2017). We could find only one example in CdLS (Rentas et al., 2020), where the patient was diagnosed with moderate–severe CdLS and had abnormal *NIPBL* splicing with inclusion of a novel exon within intron 21 sequence that was expected to introduce a premature stop codon. The increased access to WGS in the clinic will certainly unveil a larger number of similar situations.

Of note, RNA-seq was used here sequentially after WGS. It is still unclear whether combined use of RNA-seq plus WGS is a more efficient approach than sequential use (Cummings et al., 2017; Gonorazky et al., 2019; Kremer et al., 2018; Murdock et al., 2021). Integrating RNA-seq with WGS resulted in additional cases with clear diagnosis in a recent study, with an overall diagnostic rate going from 31% without RNA-seq to 38% with RNA-seq contribution (Lee et al., 2020). Moreover, 18% of all genetic diagnoses returned required RNA-seq to determine variant causality. The contribution of RNA-seq has also been highlighted recently in molecular diagnostics of rare genodermatoses (Saeidian et al., 2020) and rare muscle disorders with an overall diagnostic rate of 35% (Cummings et al., 2017). Our results, albeit in a small series of patients, suggest that a sequential use may provide a cost-effective strategy, although likely increasing the delay to patient report.

It is worth noting that Patient 5 displays a spatial aggregation of 6 intronic DNMs in the *NIPBL* sequence, including one that we considered as the cause of the disorder through its effect on splicing. While the mechanism associated with such an aggregation remains unclear, we hypothesize that all these variants occurred on the same parental haplotype on a single multihit event or sequence. Unfortunately, no polymorphisms could be identified nearby the variants to phase them. Long-read sequencing would be necessary to further study this hypothesis. The increase of *de novo* mutations in specific gene areas, also called clustered mutations or *kataegis*, is a recently discovered phenomenon observed in many models, both in cancers and in the germline genome, and is imperfectly elucidated at the biological level (Chan & Gordenin, 2015). It has been defined as more than five or six mutations in a range of 1000 bp of the human genome. Particular mutational patterns have been observed within these clusters, including C>G transversions and a role for CpG islands (The BRIDGES Consortium et al., 2018), however the limited number of variants in our patient precluded the identification of a specific mutational pattern. Some epigenetic marks are also associated, such as H3K36me3 nucleosome methylation or chromatin opening measured by DNase sensitivity. Different observations have incorporated a role for sex and age of the transmitting parent (Goldmann et al., 2016; Jónsson et al., 2017). Although not falling into the definition of *kataegis*, Patient 1 also harbored three DNM within the *SPEN* gene, including one pathogenic variant, without any enrichment in DNM overall either (Figure S4).

In addition to deep intronic DMN, our results also highlight three genes which were not usually considered as differential diagnoses until now, *POU3F3*, *SPEN*, and *TAF1*. They do not belong to the spectrum of cohesinopathies and are thus not included in our gene panel. Phenotypic features associated with these differential diagnoses overlapped with that of CdLS albeit with not very specific clinical signs, such as DD/ID, behavioral disorders, eating disorders

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and microcephaly. Dysmorphic features associated with these conditions include synophris, arched eyebrows, short nose with anteverted nostrils, prominent philtrum, and hirsutism, hence also overlapping with CdLS. These three diagnoses highlight the difficulty of establishing a clinically solid diagnosis based solely on the patient's phenotype, even when the phenotype appears to be very specific. Previous publications of ES in patients with suspected CdLS also revealed variants in genes not involved in the cohesin complex, e.g., *ZMYND11*, *MED13L*, and *PHIP*, responsible for so-called CdLS-like phenotypes (Aoi et al., 2019). These results underline, on one hand, the interest to propose additional genetic analyses in patients without a confirmed diagnosis after gene panel sequencing, and, on the other hand, some limits to phenotype-first approaches.

To consider the description of novel differential diagnosis genes, even in so-called classic-CdLS presentations, it should be discussed either (i) to successively add new CdLS genes and differential diagnoses to gene panels, as they are identified, and thus to resequence patients without a confirmed molecular diagnosis on the first versions of the panel, or (ii) to move towards a second-line WGS or ES strategy. This second strategy seems to be the best one, from a cost-effectiveness and clinical management point of view, given the increasing accessibility of ES/WGS. This approach also allows a reasonable amount of targeted genetic tests in a context of increasing number of ID genes described. Until recently, WGS was mainly accessible for research purposes due to its cost and the amount of data generated. Very–high-throughput genomic sequencing platforms are being made accessible in a medical setting in growing number of countries, allowing an easier access to genomic medicine. However, one should remind that, in CdLS, because of the existence of mosaic *NIPBL* mutations, gene panel sequencing remains essential as a first-tier analysis. Some of the mosaics are indeed not detectable by ES or WGS methods because of too low depth of coverage. Thus, it seems important to propose CdLS gene sequencing on salivary or skin samples as a first line to patients presenting a suggestive phenotype, especially those with a typical phenotype, before proceeding to a genome-wide analysis.

In conclusion, using trio-based WGS in highly selected patients, we have identified the genetic cause in 5/5 clinically-diagnosed CdLS patients with negative gene panel sequencing. Of them, we highlight (i) two genes, *POU3F3*, *SPEN*, the pathogenic variants of which being associated with CdLS-like phenotypic features, but also confirm that *TAF1* can be a differential diagnosis gene of CdLS (Cheng et al., 2020), and (ii) two cases of likely pathogenic deep intronic variants in *NIPBL* generating novel exons leading to a frameshift. Our data show WGS potential to diagnose unsolved patients with clinical suspicion of CdLS.

WEB RESOURCES

5utr ['suter'] tool: https://github.com/leklab/5utr

Custom IGV-based filtration interface: (https://github.com/francois-lecoquierre/genomic_shortcuts/)

SpliceAI: https://spliceailookup.broadinstitute.org/

AUTHOR CONTRIBUTIONS

Conception and design: Juliette Coursimault, Pascale Saugier-Veber, and Gaël Nicolas. *Material preparation, data collection, and analysis*: Juliette Coursimault, François Lecoquierre, Kévin Cassinari, Gabriella Vera, Nathalie Drouot, Céline Derambure, and Myriam Vezain. *Bioinformatic analysis*: Olivier Quenez, Sophie Coutant and François Lecoquierre. *First draft of the manuscript*: Juliette Coursimault. *Critical revision*: Gaël Nicolas, Kévin Cassinari, François Lecoquierre, Alice Goldenberg, Pascale Saugier-Veber. All authors contributed to data acquisition. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Several authors of this publication are members of the European Reference Network for Developmental Anomalies and Intellectual Disability (ERN-ITHACA). The authors have no conflict of interest to declare. Collaboration CEA-DRF-Jacob-CNRGH—CHU de Rouen. This work did benefit from support of the France Génomique National infrastructure, funded as part of the «Investissements d'Avenir» program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). We thank all the patients and their families as well as their referring physicians for their participation to this study. This study was co-supported by the European Union and Région Normandie in the context of Recherche Innovation Normandie (RIN2018). Europe gets involved in Normandie with the European Regional Development Fund (ERDF). This work was generated within the European Reference Network for Developmental Anomalies and Intellectual Disability. This work was perfomed in the framework of FHU-G4 Génomique.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ETHICS STATEMENT

Legal representatives of patients provided informed written consents for genetic analyses in a diagnostic setting. The retrospective report on the patients' medical and genetic results was approved by the Institutional Review Board of the Rouen University Hospital (CERDE, Comité d'Ethique pour la Recherche sur Données Existantes et Hors Loi Jardé, Rouen, France) (2019/0252/OB). Informed consent was obtained from all individual participants included in the study or from legal representatives, including for photographs of patients 1, 2 and 4.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request. Deidentified clinical information tables are available upon request.

ORCID

Juliette Coursimault http://orcid.org/0000-0002-2668-5779
Kévin Cassinari https://orcid.org/0000-0003-2683-4073
François Lecoquierre https://orcid.org/0000-0002-9110-1856
Olivier Quenez https://orcid.org/0000-0002-8273-8505

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Céline Derambure https://orcid.org/0000-0002-6722-5955
Myriam Vezain https://orcid.org/0000-0002-8333-1360
Jamal Ghoumid http://orcid.org/0000-0002-7111-0050
Thomas Smol http://orcid.org/0000-0002-0119-5896
Didier Lacombe https://orcid.org/0000-0002-8956-2207
Pascale Saugier-Veber http://orcid.org/0000-0002-8045-6432
Gaël Nicolas http://orcid.org/0000-0001-9391-7800

REFERENCES

Alesi, V., Dentici, M. L., Loddo, S., Genovese, S., Orlando, V., Calacci, C., Pompili, D., Dallapiccola, B., Digilio, M. C., & Novelli, A. (2019). Confirmation of *BRD4* haploinsufficiency role in Cornelia de Lange–like phenotype and delineation of a 19p13.12p13.11 gene contiguous syndrome. *Annals of Human Genetics*, *83*(2), 100–109. https://doi.org/10.1111/ahg.12289

Aoi, H., Mizuguchi, T., Ceroni, J. R., Kim, V. E. H., Furquim, I., Honjo, R. S., Iwaki, T., Suzuki, T., Sekiguchi, F., Uchiyama, Y., Azuma, Y., Hamanaka, K., Koshimizu, E., Miyatake, S., Mitsuhashi, S., Takata, A., Miyake, N., Takeda, S., Itakura, A., & Matsumoto, N. (2019). Comprehensive genetic analysis of 57 families with clinically suspected Cornelia de Lange syndrome. *Journal of Human Genetics*, *64*(10), 967–978. https://doi.org/10.1038/s10038-019-0643-z

Borck, G., Zarhrate, M., Cluzeau, C., Bal, E., Bonnefont, J.-P., Munnich, A., Cormier-Daire, V., & Colleaux, L. (2006). Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5′ untranslated region of theNIPBL gene. *Human Mutation*, *27*(8), 731–735. https://doi.org/10.1002/humu.20380

Cassinari, K., Rovelet-Lecrux, A., Tury, S., Quenez, O., Richard, A., Charbonnier, C., Olaso, R., Boland, A., Deleuze, J., Besancenot, J., Delpont, B., Pouliquen, D., Lecoquierre, F., Chambon, P., Thauvin-Robinet, C., Campion, D., Frebourg, T., Battini, J., & Nicolas, G. (2020). Haploinsufficiency of the primary familial brain calcification gene *SLC20A2* mediated by disruption of a regulatory element. *Movement Disorders*, *35*(8), 1336–1345. https://doi.org/10.1002/mds.28090

Chan, K., & Gordenin, D. A. (2015). Clusters of multiple mutations: incidence and molecular mechanisms. *Annual Review of Genetics*, *49*(1), 243–267. https://doi.org/10.1146/annurev-genet-112414-054714

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J., Kruglyak, S., & Saunders, C. T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics (Oxford, England)*, *32*(8), 1220–1222. https://doi.org/10.1093/bioinformatics/btv710

Cheng, H., Capponi, S., Wakeling, E., Marchi, E., Li, Q., Zhao, M., Weng, C., Piatek, S. G., Ahlfors, H., Kleyner, R., Rope, A., Lumaka, A., Lukusa, P., Devriendt, K., Vermeesch, J., Posey, J. E., Palmer, E. E., Murray, L., Leon, E., & Lyon, G. J. (2020). Missense variants in *TAF1* and developmental phenotypes: challenges of determining pathogenicity. *Human Mutation*, *41*(2), 449–464. https://doi.org/10.1002/humu.23936

Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. V., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., & Talkowski, M. E. (2020). A structural variation reference for medical and population genetics. *Nature*, *581*(7809), 444–451. https://doi.org/10.1038/s41586-020-2287-8

Coursimault, J., Rovelet-Lecrux, A., Cassinari, K., Brischoux-Boucher, E., Saugier-Veber, P., Goldenberg, A., Lecoquierre, F., Drouot, N., Richard, A., Vera, G., Coutant, S., Quenez, O., Rolain, M., Bonnet, C., Bronner, M., Lecourtois, M., & Nicolas, G. (2022). uORF-introducing variants in the 5′UTR of the NIPBL gene as a cause of Cornelia de Lange syndrome. Human Mutation. Portico. https://doi.org/10.1002/humu.24384

Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E., Reddy, H. M., Zhao, F., Weisburd, B., Karczewski, K. J., O'Donnell-Luria, A. H., Birnbaum, D., Sarkozy, A., Hu, Y., & MacArthur, D. G. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine*, *9*(386), eaal5209. https://doi.org/10.1126/scitranslmed.aal5209

Deardorff, M. A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K. E., De Baere, E., Decroos, C., Di Donato, N., Ernst, S., Francey, L. J., Gyftodimou, Y., Hirashima, K., Hullings, M., & Shirahige, K. (2012). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature*, *489*(7415), 313–317. https://doi.org/10.1038/nature11316

Deardorff, M. A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodríguez, C., Arnedo, M., Loeys, B., Kline, A. D., Wilson, M., Lillquist, K., Siu, V., Ramos, F. J., Musio, A., Jackson, L. S., Dorsett, D., & Krantz, I. D. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *The American Journal of Human Genetics*, *80*(3), 485–494. https://doi.org/10.1086/511888

Deardorff, M. A., Porter, N. J., & Christianson, D. W. (2016). Structural aspects of HDAC8 mechanism and dysfunction in Cornelia de Lange syndrome spectrum disorders: structural aspects of HDAC8 mechanism. *Protein Science*, *25*(11), 1965–1976. https://doi.org/10.1002/pro.3030

Dossin, F., Pinheiro, I., Żylicz, J. J., Roensch, J., Collombet, S., Le Saux, A., Chelmicki, T., Attia, M., Kapoor, V., Zhan, Y., Dingli, F., Loew, D., Mercher, T., Dekker, J., & Heard, E. (2020). SPEN integrates transcriptional and epigenetic control of X-inactivation. *Nature*, *578*(7795), 455–460. https://doi.org/10.1038/s41586-020-1974-9

Dusl, M., Senderek, J., Muller, J. S., Vogel, J. G., Pertl, A., Stucka, R., Lochmuller, H., David, R., & Abicht, A. (2015). A 3'-UTR mutation creates a microRNA target site in the GFPT1 gene of patients with congenital myasthenic syndrome. *Human Molecular Genetics*, *24*(12), 3418–3426. https://doi.org/10.1093/hmg/ddv090

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Gardner, E. J., Lam, V. K., Harris, D. N., Chuang, N. T., Scott, E. C., Pittard, W. S., Mills, R. E., 1000 Genomes Project Consortium, & Devine, S. E. (2017). The mobile element locator tool (MELT): population-scale mobile element discovery and biology. *Genome Research*, *27*(11), 1916–1929. https://doi.org/10.1101/gr.218032.116

Geoffroy, V., Herenger, Y., Kress, A., Stoetzel, C., Piton, A., Dollfus, H., & Muller, J. (2018). AnnotSV: An integrated tool for structural variations annotation. *Bioinformatics (Oxford, England)*, *34*(20), 3572–3574. https://doi.org/10.1093/bioinformatics/bty304

Gillis, L. A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A. D., Li, H., Devoto, M., Jackson, L. G., & Krantz, I. D. (2004). NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. *American Journal of Human Genetics*, *75*(4), 610–623. https://doi.org/10.1086/424698

Goldmann, J. M., Wong, W. S. W., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A. B., Glusman, G., Vissers, L. E. L. M., Hoischen, A., Roach, J. C., Vockley, J. G., Veltman, J. A., Solomon, B. D., Gilissen, C., & Niederhuber, J. E. (2016). Parent-of-origin-specific signatures of de novo mutations. *Nature Genetics*, *48*(8), 935–939. https://doi.org/10.1038/ng.3597

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D., Ohri, K., Viththiyapaskaran, S., Tarnopolsky, M. A., Mathews, K. D., Moore, S. A., Osorio, A. N., Villanova, D., Kemaladewi, D. U., Cohn, R. D., Brudno, M., & Dowling, J. J. (2019). Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *American Journal of Human Genetics*, *104*(3), 466–483. https://doi.org/10.1016/j.ajhg.2019.01.012

Hehir-Kwa, J. Y., Pfundt, R., & Veltman, J. A. (2015). Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert Review of Molecular Diagnostics*, *15*(8), 1023–1032. https://doi.org/10.1586/14737159.2015.1053467

Huisman, S. A., Redeker, E. J. W., Maas, S. M., Mannens, M. M., & Hennekam, R. C. M. (2013). High rate of mosaicism in individuals with Cornelia de Lange syndrome. *Journal of Medical Genetics*, *50*(5), 339–344. https://doi.org/10.1136/jmedgenet-2012-101477

Hurst, S. E., Liktor-Busa, E., Moutal, A., Parker, S., Rice, S., Szelinger, S., Senner, G., Hammer, M. F., Johnstone, L., Ramsey, K., Narayanan, V., Perez-Miller, S., Khanna, M., Dahlin, H., Lewis, K., Craig, D., Wang, E. H., Khanna, R., & Nelson, M. A. (2018). A novel variant in TAF1 affects gene expression and is associated with X-linked TAF1 intellectual disability syndrome. *Neuronal Signaling*, *2*(3), NS20180141. https://doi.org/10.1042/NS20180141

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki, J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A., Batzoglou, S., Sanders, S. J., & Farh, K. K.-H. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, *176*(3), 535–548.e24. https://doi.org/10.1016/j.cell.2018.12.015

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M. T., Hjorleifsson, K. E., Eggertsson, H. P., Gudjonsson, S. A., Ward, L. D., Arnadottir, G. A., Helgason, E. A., Helgason, H., Gylfason, A., Jonasdottir, A., Jonasdottir, A., Rafnar, T., Frigge, M., & Stefansson, K. (2017). Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature*, *549*(7673), 519–522. https://doi.org/10.1038/nature24018

Jouret, G., Heide, S., Sorlin, A., Faivre, L., Chantot-Bastaraud, S., Beneteau, C., Denis-Musquer, M., Turnpenny, P. D., Coutton, C., Vieville, G., Thevenon, J., Larson, A., Petit, F., Boudry, E., Smol, T., Delobel, B., Duban-Bedu, B., Fallerini, C., Mari, F., & Klink, B. (2022). Understanding the new BRD4-related syndrome: Clinical and genomic delineation with an international cohort study. *Clinical Genetics*, *102*(2), 117–122. https://doi.org/10.1111/cge.14141

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., & MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kim, A., Le Douce, J., Diab, F., Ferovova, M., Dubourg, C., Odent, S., Dupé, V., David, V., Diambra, L., Watrin, E., & de Tayrac, M. (2020). Synonymous variants in holoprosencephaly alter codon usage and impact the Sonic Hedgehog protein. *Brain*, *143*(7), 2027–2038. https://doi.org/10.1093/brain/awaa152

Kline, A. D., Moss, J. F., Selicorni, A., Bisgaard, A.-M., Deardorff, M. A., Gillett, P. M., Ishman, S. L., Kerr, L. M., Levin, A. V., Mulder, P. A., Ramos, F. J., Wierzba, J., Ajmone, P. F., Axtell, D., Blagowidow, N., Cereda, A., Costantino, A., Cormier-Daire, V., FitzPatrick, D., & Hennekam, R. C. (2018). Diagnosis and management of Cornelia de Lange syndrome: First international consensus statement. *Nature Reviews Genetics*, *19*(10), 649–666. https://doi.org/10.1038/s41576-018-0031-0

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., Nowaczyk, M. J. M., Toriello, H., Bamshad, M. J., Carey, J. C., Rappaport, E., Kawauchi, S., Lander, A. D., Calof, A. L., Li, H.-H., & Jackson, L. G. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature Genetics*, *36*(6), 631–635. https://doi.org/10.1038/ng1364

Kremer, L. S., Wortmann, S. B., & Prokisch, H. (2018). « transcriptomics »: molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *Journal of Inherited Metabolic Disease*, *41*(3), 525–532. https://doi.org/10.1007/s10545-017-0133-4

Labrouche-Colomer, S., Soukarieh, O., Proust, C., Mouton, C., Huguenin, Y., Roux, M., Besse, C., Boland, A., Olaso, R., Constans, J., Deleuze, J.-F., Morange, P.-E., Jaspard-Vinassa, B., & Trégouët, D.-A., the GenMed consortium. (2020). A novel rare c. −39C$$T mutation in the PROS1 5'UTR causing PS deficiency by creating a new upstream translation initiation codon and inhibiting the production of the natural protein [preprint]. *Clinical Science*, *134*(10), 1181–1190. https://doi.org/10.1042/CS20200403

Lee, H., Huang, A. Y., Wang, L.-K., Yoon, A. J., Renteria, G., Eskin, A., Signer, R. H., Dorrani, N., Nieves-Rodriguez, S., Wan, J., Douine, E. D., Woods, J. D., Dell'Angelica, E. C., Fogel, B. L., Martin, M. G., Butte, M. J., Parker, N. H., Wang, R. T., Shieh, P. B., & Nelson, S. F. (2020). Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *22*(3), 490–499. https://doi.org/10.1038/s41436-019-0672-1

MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The database of genomic variants: A curated collection of structural variation in the human genome. *Nucleic Acids Research*, *42*(Database issue), D986–D992. https://doi.org/10.1093/nar/gkt958

Martin, H. C., Gardner, E. J., Samocha, K. E., Kaplanis, J., Akawi, N., Sifrim, A., Eberhardt, R. Y., Tavares, A. L. T., Neville, M. D. C., Niemi, M. E. K., Gallone, G., McRae, J., Deciphering Developmental Disorders Study, Wright, C. F., FitzPatrick, D. R., Firth, H. V., & Hurles, M. E. (2021). The contribution of X-linked coding variation to severe developmental disorders. *Nature Communications*, *12*(1), 627. https://doi.org/10.1038/s41467-020-20852-3

Murdock, D. R., Dai, H., Burrage, L. C., Rosenfeld, J. A., Ketkar, S., Müller, M. F., Yépez, V. A., Gagneur, J., Liu, P., Chen, S., Jain, M., Zapata, G., Bacino, C. A., Chao, H.-T., Moretti, P., Craigen, W. J., Hanchard, N. A., Undiagnosed Diseases Network, & Lee, B. (2021). Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *The Journal of Clinical Investigation*, *131*(1), 141500. https://doi.org/10.1172/JCI141500

Nizon, M., Henry, M., Michot, C., Baumann, C., Bazin, A., Bessières, B., Blesson, S., Cordier-Alex, M.-P., David, A., Delahaye-Duriez, A., Delezoïde, A.-L., Dieux-Coeslier, A., Doco-Fenzy, M., Faivre, L., Goldenberg, A., Layet, V., Loget, P., Marlin, S., Martinovic, J., & Cormier-Daire, V. (2016). A series of 38 novel germline and somatic mutations of NIPBL in Cornelia de Lange syndrome. *Clinical Genetics*, *89*(5), 584–589. https://doi.org/10.1111/cge.12720

Olley, G., Ansari, M., Bengani, H., Grimes, G. R., Rhodes, J., von Kriegsheim, A., Blatnik, A., Stewart, F. J., Wakeling, E., Carroll, N., Ross, A., Park, S.-M., Bickmore, W. A., Pradeepa, M. M., & FitzPatrick, D. R. (2018). BRD4 interacts with NIPBL and BRD4 is mutated in a Cornelia de Lange–like syndrome. *Nature Genetics*, *50*(3), 329–332. https://doi.org/10.1038/s41588-018-0042-y

O'Rawe, J. A., Wu, Y., Dörfel, M. J., Rope, A. F., Au, P. Y. B., Parboosingh, J. S., Moon, S., Kousi, M., Kosma, K., Smith, C. S., Tzetis, M., Schuette, J. L., Hufnagel, R. B., Prada, C. E., Martinez, F., Orellana, C., Crain, J., Caro-Llopis, A., Oltra, S., & Lyon, G. J. (2015). TAF1 Variants are associated with dysmorphic features, intellectual disability, and neurological manifestations. *American Journal of*

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Human Genetics, *97*(6), 922–932. https://doi.org/10.1016/j.ajhg.2015.11.005

Parenti, I., Diab, F., Gil, S. R., Mulugeta, E., Casa, V., Berutti, R., Brouwer, R. W. W., Dupé, V., Eckhold, J., Graf, E., Puisac, B., Ramos, F., Schwarzmayr, T., Gines, M. M., van Staveren, T., van IJcken, W. F. J., Strom, T. M., Pié, J., Watrin, E., & Wendt, K. S. (2020). MAU2 and NIPBL variants impair the heterodimerization of the Cohesin Loader Subunits and cause Cornelia de Lange Syndrome. *Cell Reports*, *31*(7), 107647. https://doi.org/10.1016/j.celrep.2020.107647

Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: A new computational method for splice site prediction. *Nucleic Acids Research*, *29*(5), 1185–1190. https://doi.org/10.1093/nar/29.5.1185

Piché, J., Van Vliet, P. P., Pucéat, M., & Andelfinger, G. (2019). The expanding phenotypes of cohesinopathies: One ring to rule them all. *Cell Cycle (Georgetown, Tex.)*, *18*(21), 2828–2848. https://doi.org/10.1080/15384101.2019.1658476

Plesser Duvdevani, M., Pettersson, M., Eisfeldt, J., Avraham, O., Dagan, J., Frumkin, A., Lupski, J. R., Lindstrand, A., & Harel, T. (2020). Whole-genome sequencing reveals complex chromosome rearrangement disrupting *NIPBL* in infant with Cornelia de Lange syndrome. *American Journal of Medical Genetics, Part A*, *182*(5), 1143–1151. https://doi.org/10.1002/ajmg.a.61539

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, *36*(10), 983–987. https://doi.org/10.1038/nbt.4235

Quenez, O., Cassinari, K., Coutant, S., Lecoquierre, F., Le Guennec, K., Rousseau, S., Richard, A.-C., Vasseur, S., Bouvignies, E., Bou, J., Lienard, G., Manase, S., Fourneaux, S., Drouot, N., Nguyen-Viet, V., Vezain, M., Chambon, P., Joly-Helas, G., Le Meur, N., & Nicolas, G. (2021). Detection of copy-number variations from NGS data using read depth information: A diagnostic performance evaluation. *European Journal of Human Genetics: EJHG*, *29*(1), 99–109. https://doi.org/10.1038/s41431-020-0672-2

Radio, F. C., Pang, K., Ciolfi, A., Levy, M. A., Hernández-García, A., Pedace, L., Pantaleoni, F., Liu, Z., de Boer, E., Jackson, A., Bruselles, A., McConkey, H., Stellacci, E., Lo Cicero, S., Motta, M., Carrozzo, R., Dentici, M. L., McWalter, K., Desai, M., & Tartaglia, M. (2021). SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an episignature of X chromosomes in females. *The American Journal of Human Genetics*, *108*(3), 502–516. https://doi.org/10.1016/j.ajhg.2021.01.015

Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in genie. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *4*(3), 311–323. https://doi.org/10.1089/cmb.1997.4.311

Rentas, S., Rathi, K. S., Kaur, M., Raman, P., Krantz, I. D., Sarmady, M., & Tayoun, A. A. (2020). Diagnosing Cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genetics in Medicine*, *22*(5), 927–936. https://doi.org/10.1038/s41436-019-0741-5

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (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. *Genetics in Medicine*, *17*(5), 405–423. https://doi.org/10.1038/gim.2015.30

Roller, E., Ivakhno, S., Lee, S., Royce, T., & Tanner, S. (2016). Canvas: Versatile and scalable detection of copy number variants. *Bioinformatics*, *32*(15), 2375–2377. https://doi.org/10.1093/bioinformatics/btw163

Saeidian, A. H., Youssefian, L., Vahidnezhad, H., & Uitto, J. (2020). Research techniques made simple: Whole-transcriptome sequencing by RNA-Seq for diagnosis of monogenic disorders. *The Journal of Investigative Dermatology*, *140*(6), 1117–1126.e1. https://doi.org/10.1016/j.jid.2020.02.032

Samocha, K. E., Robinson, E. B., Sanders, S. J., Stevens, C., Sabo, A., McGrath, L. M., Kosmicki, J. A., Rehnström, K., Mallick, S., Kirby, A., Wall, D. P., MacArthur, D. G., Gabriel, S. B., DePristo, M., Purcell, S. M., Palotie, A., Boerwinkle, E., Buxbaum, J. D., Cook, E. H., & Daly, M. J. (2014). A framework for the interpretation of de novo mutation in human disease. *Nature Genetics*, *46*(9), 944–950. https://doi.org/10.1038/ng.3050

Selicorni, A., Mariani, M., Lettieri, A., & Massa, V. (2021). Cornelia de lange syndrome: From a disease to a broader spectrum. *Genes*, *12*(7), 1075. https://doi.org/10.3390/genes12071075

Selicorni, A., Russo, S., Gervasini, C., Castronovo, P., Milani, D., Cavalleri, F., Bentivegna, A., Masciadri, M., Domi, A., Divizia, M., Sforzini, C., Tarantino, E., Memo, L., Scarano, G., & Larizza, L. (2007). Clinical score of 62 Italian patients with Cornelia de Lange syndrome and correlations with the presence and type of *NIPBL* mutation. *Clinical Genetics*, *72*(2), 98–108. https://doi.org/10.1111/j.1399-0004.2007.00832.x

Shen, Y., Shu, S., Ren, Y., Xia, W., Chen, J., Dong, L., Ge, H., Fan, S., Shi, L., Peng, B., & Zhang, X. (2021). Case report: Two novel frameshift mutations in SLC20A2 and one novel splice donor mutation in PDGFB associated with primary familial brain calcification. *Frontiers in Genetics*, *12*, 643452. https://doi.org/10.3389/fgene.2021.643452

Snijders Blok, L., Kleefstra, T., Venselaar, H., Maas, S., Kroes, H. Y., Lachmeijer, A. M. A., van Gassen, K. L. I., Firth, H. V., Tomkins, S., Bodek, S., Õunap, K., Wojcik, M. H., Cunniff, C., Bergstrom, K., Powis, Z., Tang, S., Shinde, D. N., Au, C., Iglesias, A. D., & Fisher, S. E. (2019). De novo variants disturbing the transactivation capacity of POU3F3 cause a characteristic neurodevelopmental disorder. *The American Journal of Human Genetics*, *105*(2), 403–412. https://doi.org/10.1016/j.ajhg.2019.06.007

Carlson, J., Locke, A. E., Flickinger, M., Zawistowski, M., Levy, S., Myers, R. M., Boehnke, M., Kang, H. M., Scott, L. J., Li, J. Z., & Zöllner, S., The BRIDGES Consortium. (2018). Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *Nature Communications*, *9*(1), 3753. https://doi.org/10.1038/s41467-018-05936-5

Tonkin, E. T., Smith, M., Eichhorn, P., Jones, S., Imamwerdi, B., Lindsay, S., Jackson, M., Wang, T.-J., Ireland, M., Burn, J., Krantz, I. D., Carr, P., & Strachan, T. (2004). A giant novel gene undergoing extensive alternative splicing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3. *Human Genetics*, *115*(2), 139–148. https://doi.org/10.1007/s00439-004-1134-6

Wang, T., Hoekzema, K., Vecchio, D., Wu, H., Sulovari, A., Coe, B. P., Gillentine, M. A., Wilfert, A. B., Perez-Jurado, L. A., Kvarnung, M., Sleyp, Y., Earl, R. K., Rosenfeld, J. A., Geisheker, M. R., Han, L., Du, B., Barnett, C., Thompson, E., Shaw, M., & Eichler, E. E. (2020). Large-scale targeted sequencing identifies risk genes for neurodevelopmental disorders. *Nature Communications*, *11*(1), 4932. https://doi.org/10.1038/s41467-020-18723-y

Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., McRae, J. F., van Kogelenberg, M., King, D. A., Ambridge, K., Barrett, D. M., Bayzetinova, T., Bevan, A. P., Bragin, E., Chatzimichali, E. A., Gribble, S., Jones, P., Krishnappa, N., Mason, L. E., Miller, R., Morley, K. I., & Firth, H. V. (2015). Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of genome-wide research data. *The Lancet*, *385*(9975), 1305–1314. https://doi.org/10.1016/S0140-6736(14)61705-0

Wright, C. F., Quaife, N. M., Ramos-Hernández, L., Danecek, P., Ferla, M. P., Samocha, K. E., Kaplanis, J., Gardner, E. J., Eberhardt, R. Y., Chao, K. R., Karczewski, K. J., Morales, J.,

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gallone, G., Balasubramanian, M., Banka, S., Gompertz, L., Kerr, B., Kirby, A., Lynch, S. A., & Whiffin, N. (2021). Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. *The American Journal of Human Genetics*, *108*(6), 1083–1094. https://doi.org/10.1016/j.ajhg.2021.04.025

Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *11*(2-3), 377–394. https://doi.org/10.1089/1066527041410418

Zhang, M. Q. (1998). Statistical features of human exons and their flanking regions. *Human Molecular Genetics*, *7*(5), 919–932. https://doi.org/10.1093/hmg/7.5.919

Zuin, J., Casa, V., Pozojevic, J., Kolovos, P., van den Hout, M. C. G. N., van Ijcken, W. F. J., Parenti, I., Braunholz, D., Baron, Y., Watrin, E., Kaiser, F. J., & Wendt, K. S. (2017). Regulation of the cohesin-loading factor NIPBL: Role of the lncRNA NIPBL-AS1 and identification of a distal enhancer element. *PLoS Genetics*, *13*(12), e1007137. https://doi.org/10.1371/journal.pgen.1007137

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Coursimault, J., Cassinari, K., Lecoquierre, F., Quenez, O., Coutant, S., Derambure, C., Vezain, M., Drouot, N., Vera, G., Schaefer, E., Philippe, A., Doray, B., Lambert, L., Ghoumid, J., Smol, T., Rama, M., Legendre, M., Lacombe, D., Fergelot, P., … Nicolas, G. (2022). Deep intronic *NIPBL de novo* mutations and differential diagnoses revealed by whole genome and RNA sequencing in Cornelia de Lange syndrome patients. *Human Mutation*, 43, 1882–1897. https://doi.org/10.1002/humu.24438

10981004, 2022, 12, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/humu.24438 by CHU Bordeaux, Wiley Online Library on [24/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License