Received: 23 March 2023 | Revised: 2 May 2023 | Accepted: 3 May 2023

DOI: 10.1002/ajmg.a.63247

ORIGINAL ARTICLE

AMERICAN JOURNAL OF medical genetics PART A WILEY

Genomic analyses in Cornelia de Lange Syndrome and related diagnoses: Novel candidate genes, genotype–phenotype correlations and common mechanisms

Maninder Kaur[1] | Justin Blair[1] | Batsal Devkota[2] | Sierra Fortunato[1] | Dinah Clark[3] | Audrey Lawrence[1] | Jiwoo Kim[1] | Wonwook Do[1] | Benjamin Semeo[1] | Olivia Katz[1] | Devanshi Mehta[1] | Nobuko Yamamoto[4] | Emma Schindler[1] | Zayd Al Rawi[1] | Nina Wallace[1] | Jonathan J. Wilde[5] | Jennifer McCallum[6] | Jinglan Liu[7] | Dongbin Xu[8] | Marie Jackson[1] | Stefan Rentas[9] | Ahmad Abou Tayoun[10,11] | Zhang Zhe[12] | Omar Abdul-Rahman[13] | Bill Allen[14] | Moris A. Angula[15] | Kwame Anyane-Yeboa[16] | Jesús Argente[17,18] | Pamela H. Arn[19] | Linlea Armstrong[20,21] | Lina Basel-Salmon[22,23,24] | Gareth Baynam[25,26,27] | Lynne M. Bird[28,29] | Daniel Bruegger[30] | Gaik-Siew Ch'ng[31] | David Chitayat[32,33] | Robin Clark[34] | Gerald F. Cox[35] | Usha Dave[36] | Elfrede DeBaere[37,38] | Michael Field[39] | John M. Graham Jr[40] | Karen W. Gripp[41] | Robert Greenstein[42] | Neerja Gupta[43] | Randy Heidenreich[44] | Jodi Hoffman[45] | Robert J. Hopkin[46] | Kenneth L. Jones[47] | Marilyn C. Jones[28,29] | Ariana Kariminejad[48] | Jillene Kogan[49] | Baiba Lace[50] | Julian Leroy[37,38] | Sally Ann Lynch[51] | Marie McDonald[52] | Kirsten Meagher[20] | Nancy Mendelsohn[53] | Ieva Micule[50] | John Moeschler[54] | Sheela Nampoothiri[55] | Kaoru Ohashi[21] | Cynthia M. Powell[56] | Subhadra Ramanathan[34] | Salmo Raskin[57] | Elizabeth Roeder[58] | Marlene Rio[59] | Alan F. Rope[60] | Karan Sangha[20] | Angela E. Scheuerle[61] | Adele Schneider[62] | Stavit Shalev[63] | Victoria Siu[64,65] | Rosemarie Smith[66] | Cathy Stevens[67] | Tinatin Tkemaladze[68] | John Toimie[69] | Helga Toriello[70] | Anne Turner[71,72] | Patricia G. Wheeler[72] | Susan M. White[73,74] | Terri Young[75,76] | Kathleen M. Loomes[77,78] | Mary Pipan[78,79] | Ann Tokay Harrington[80] | Elaine Zackai[1,78] | Ramakrishnan Rajagopalan[81,82] | Laura Conlin[81,82] | Matthew A. Deardorff[83,84] | Deborah McEldrew[1] | Juan Pie[85] | Feliciano Ramos[86,87] | Antonio Musio[88] | Antonie D. Kline[89] | Kosuke Izumi[1,78] | Sarah E. Raible[1] | Ian D. Krantz[1,78]

For affiliations refer to page 15

Correspondence
Ian D. Krantz, Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA.
Email: krantz@chop.edu

Funding information
Cornelia de Lange Syndrome Foundation; National Institutes of Health

Abstract

Cornelia de Lange Syndrome (CdLS) is a rare, dominantly inherited multisystem developmental disorder characterized by highly variable manifestations of growth and developmental delays, upper limb involvement, hypertrichosis, cardiac, gastrointestinal, craniofacial, and other systemic features. Pathogenic variants in genes encoding cohesin complex structural subunits and regulatory proteins (NIPBL, SMC1A, SMC3, HDAC8, and RAD21) are the major pathogenic contributors to CdLS. Heterozygous or hemizygous variants in the genes encoding these five proteins have been found to be contributory to CdLS, with variants in *NIPBL* accounting for the majority (>60%) of cases, and the only gene identified to date that results in the severe or classic form of CdLS when mutated. Pathogenic variants in cohesin genes other than *NIPBL* tend to result in a less severe phenotype. Causative variants in additional genes, such as *ANKRD11*, *EP300*, *AFF4*, *TAF1*, and *BRD4*, can cause a CdLS-like phenotype. The common role that these genes, and others, play as critical regulators of developmental transcriptional control has led to the conditions they cause being referred to as disorders of transcriptional regulation (or "DTRs"). Here, we report the results of a comprehensive molecular analysis in a cohort of 716 probands with typical and atypical CdLS in order to delineate the genetic contribution of causative variants in cohesin complex genes as well as novel candidate genes, genotype–phenotype correlations, and the utility of genome sequencing in understanding the mutational landscape in this population.

KEYWORDS
CdLS, cohesin, Cornelia de Lange Syndrome, genome, HDAC8, NIPBL, RAD21, SMC1A, SMC3, transcription

1 | INTRODUCTION

Cornelia de Lange syndrome (CdLS, OMIM# 122470; 300,590; 300,882; 610,759; 614,701), also called Brachmann-de Lange syndrome, is a rare dominant multisystem developmental disorder with variable expression that affects approximately 1 in 10,000 to 1 in 30,000 live births (Krantz et al., 2004; Mannini et al., 2013). The first reports of CdLS were made by the Dutch anatomist and pathologist Willem Vrolik in 1849 and subsequently by Dr. Winfried Robert Clemens Brachmann in 1916, who both described single cases. However, the diagnosis was formally characterized by the Dutch physician Dr. Cornelia de Lange who described three unrelated cases in 1933 (Brachmann, 1916; De Lange, 1933; Oostra et al., 1994; Vrolik, 1849). The clinical hallmarks of CdLS include a distinct facial appearance and variable growth delay, intellectual disability, upper limb abnormalities, hypertrichosis, gastroesophageal dysfunction, cardiac, ocular, diaphragmatic, genitourinary, and other systemic involvement (Jackson et al., 1993; Kline et al., 2007). Craniofacial features can include microcephaly, synophrys, arched eyebrows, long and thick eyelashes, long philtrum, thin vermilion of the upper lip, depressed corners of the mouth, a high arched (and sometimes cleft) palate, and low-set/posteriorly rotated ears (Jackson et al., 1993; Kline et al., 2007; Kline et al., 2018; Figure 1a). Upper limb differences may range in severity from small hands, single palmar creases and fifth finger clinodactyly to various forms of oligodactyly and/or syndactyly with almost complete absence of the upper extremities being the most severe manifestation (Marino et al., 2002; Mehta et al., 2016; Figure 1b). Affected individuals may also present with intestinal malrotation, congenital diaphragmatic hernia (CDH), hearing loss, myopia, hypoplastic genitalia, autism, and self-injurious behavior (Ajmone et al., 2014; Grados et al., 2017; Jackson et al., 1993; Kline et al., 2007; Kline et al., 2018; Levin et al., 1990; Marino et al., 2002; Sataloff et al., 1990).

The classic CdLS phenotype—characterized by the craniofacial gestalt, growth and developmental delay, and limb differences—is striking and easily recognized. However, the broader CdLS phenotype is a spectrum that ranges from this classic presentation to milder or "non-classic" forms of CdLS (Kline et al., 2018; Figure 1a). Individuals with nonclassic CdLS may retain some of the cardinal features but may lack other clinical manifestations or manifest differing degrees of severity.

CdLS can be diagnosed clinically or by molecular confirmation of a pathogenic variant in one of five genes (*NIPBL*, *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*) associated with the cohesin pathway and rarely

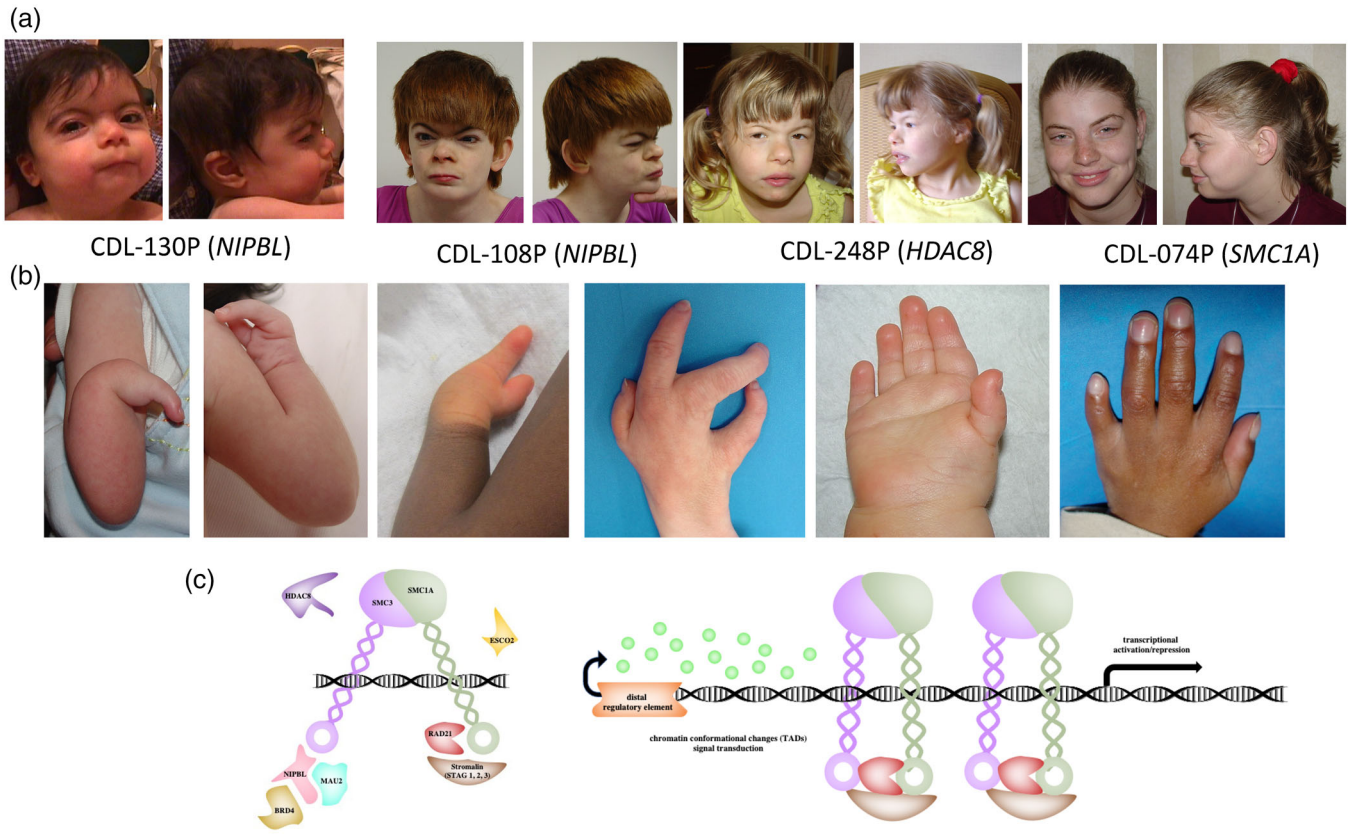


FIGURE 1 Overview of CdLS. (a) Typical facial features in CdLS with the classic features seen in the two individuals on the left with *NIPBL* pathogenic variants and more subtle/milder manifestations in the two individuals on the right with *HDAC8* and *SMC1A* pathogenic variants. (b) Variable upper limb differences seen in CdLS ranging from severe oligodactyly on the left to small hands with single palmar creases and hypoplasia of the fifth finger. (c) Simplified representation of the cohesin complex and core structural and regulatory proteins involved in CdLS that disrupt cohesin's noncanonical role in regulating developmental gene expression.

in two additional genes (*BRD4* and *ANKRD11*; Figure 1c). Cohesin plays a pivotal role in chromatid cohesion, gene expression, and DNA repair. The main cohesin genes that result in CdLS when mutated fall into two main categories: genes encoding cohesin regulatory proteins (e.g., *NIPBL*, *HDAC8*) and genes encoding cohesin structural proteins (e.g., *SMC1A*, *SMC3*, *RAD21*). *SMC1A*, *SMC3*, and *RAD21* encode core components of the cohesin complex, while *HDAC8* encodes a key regulator of cohesin that functions as a deacetylator of the SMC3 protein involved in regulating the dissociation of cohesin from chromatin (Deardorff, Bando, et al., 2012). The majority of affected individuals (>60%) have a pathogenic variant in *NIPBL*, a gene whose protein product is required for loading cohesin onto chromatin (Kline et al., 2018; Krantz et al., 2004). NIPBL's cohesin loading function is conserved across evolution, as demonstrated from experimental evidence obtained from model organisms (Ciosk et al., 2000; Gillespie & Hirano, 2004; Rollins et al., 2004; Takahashi et al., 2004). *NIPBL* is located on chromosome 5p13.2, spans more than 190 kb and consists of 47 exons that encodes two isoforms of delangin; A and B consisting of 2804 and 2697 amino acids, respectively (Krantz et al., 2004; Tonkin et al., 2004). SMC1A and SMC3 are structural maintenance of chromosomes (SMCs) proteins that are components of a large family of ring complexes that participate in DNA regulatory and repair functions. *SMC1A* encodes a subunit of the cohesin-core complex that tethers sister chromatids together to ensure correct chromosome segregation in both mitosis and meiosis. As a member of the cohesin ring, SMC1A takes part in gene transcription regulation and genome organization; and it participates in the DNA damage repair (DDR) pathway, being phosphorylated by Ataxia Telangiectasia Mutated (ATM) and Ataxia Telangiectasia and Rad3 Related (ATR) threonine/serine kinases. It is also a component of the recombination protein complex (RC-1) involved in DNA repair by recombination (Musio, 2020). *SMC1A* is located on the X chromosome, in a region that partially escapes X inactivation (Brown et al., 1995); both hemizygous male and heterozygous female individuals with CdLS have been identified with *SMC1A* pathogenic variants (Mannini et al., 2010). As one of the key components of the cohesion complex, SMC1A forms a tripartite ring structure with SMC3, RAD21, and stromal antigens (STAGs) that secure sister chromatids together by trapping them inside the ring (Haering et al., 2008). SMC3 forms a V-shaped SMC1A/SMC3 heterodimer in the tripartite ring structure via the interaction between the hinge domains (Deardorff et al., 2007). The structural and functional similarities between the gene products of *SMC1A* and *SMC3* imply that genetic variation in the two genes may result in similar phenotypes.

The vast majority of cohesin-related CdLS cases result from de novo causative variants with rare familial recurrences being due to germ line mosaicism or transmission from a mildly affected parent (Russell et al., 2001). Genotype–phenotype correlations have shown that *NIPBL* variants usually result in a classic and more severe CdLS phenotype than variants in other genes (Kaur et al., 2016; Mannini et al., 2013). A smaller number of affected individuals (totaling 5%–7%) have pathogenic variants in *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*. Individuals with pathogenic variants in these four genes tend to have milder or "non-classic" CdLS phenotypes (Deardorff et al., 2007; Deardorff, Bando, et al., 2012; Deardorff, Wilde, et al., 2012; Gil-Rodriguez et al., 2015; Kaur et al., 2016; Kline et al., 2018; Mannini et al., 2013; Musio et al., 2006).

BRD4 and *ANKRD11* have only recently been added to the list of known CdLS-causing genes. *BRD4* encodes a chromatin-associated protein that cooperates with *NIPBL* in transcriptional regulation and variants have been identified in a few individuals with CdLS (Olley et al., 2018; Rentas et al., 2020). *ANKRD11* is involved in regulating gene expression via chromatin remodeling (Cucco et al., 2020). Variants in *ANRKD11* have been reported in a few individuals with non-classic CdLs and overlapping features with KBG syndrome (Ansari et al., 2014; Parenti et al., 2016).

Alterations in cohesin and associated pathways caused by variants in genes encoding components of the transcriptional machinery as well as proteins involved in epigenetic modification, are causative of CdLS and related diagnoses when disrupted and have more broadly been termed "transcriptomopathies" or "disorders of transcriptional regulation" (Izumi, 2016; Yuan et al., 2015). Similarities between the clinical phenotypes of diverse syndromic diagnoses caused by disruption of developmental transcriptional regulation suggests that some commonalities exist in subsets of critical developmental genes that are misexpressed at key time points in organogenesis resulting in developmental diagnoses with overlapping phenotypes.

The high degree of clinical and genetic heterogeneity, especially among individuals with mild or "atypical" CdLS can often impede the diagnosis (Kline et al., 2018). Overlap between clinical features of CdLS and other diagnoses provides an additional challenge to confirming a diagnosis of CdLS (Ansari et al., 2014; Cucco et al., 2020; Gil-Rodriguez et al., 2015; Parenti et al., 2016). The presence of somatic mosaicism in some individuals with CdLS can also hinder establishing a molecular diagnosis with testing of tissue other than blood being needed in those with a negative result from blood (Ansari et al., 2014; Kline et al., 2018).

Although great progress has been made in identifying the genetic causes of CdLS, there remains a significant subset of affected individuals without an identifiable pathogenic variant, suggesting that there are additional mutational mechanisms likely not captured on standard targeted gene sequencing, panels or exome sequencing (e.g., noncoding variants in regulatory regions, deep intronic variants, complex structural rearrangements, undetected mosaicism) as well as additional CdLS-related genes yet to be discovered. In this study, we provide a comprehensive overview of all pathogenic genetic variants identified in our cohort of 716 molecularly screened CdLS probands and family members as well as assess the utility of genome sequencing in the subset of 178 probands who were not found to have an identifiable mutation through standard genetic screening. This paper represents a comprehensive review of genetic variation in CdLS and related diagnoses and offers insights into the diagnostic yield and contribution of the many genes involved, genotype–phenotype correlations, and potential novel candidate genes.

2 | MATERIALS AND METHODS

2.1 | Patients

All patients and family members were enrolled in the study under an institutional review board-approved protocol of informed consent at The Children's Hospital of Philadelphia (CHOP). All subjects were evaluated by clinical dysmorphologists with experience in the diagnosis of CdLS.

Patients were either seen at CHOP or were referred by experienced clinical geneticists or pediatricians. Clinical history and photographs were obtained for enrolled individuals.

All individuals, for whom sufficiently detailed phenotypic data were available, were assessed for diagnostic severity and classified into one of the five clinical diagnostic groups: definite, possible, atypical/overlapping (CdLS-like), unlikely or not CdLS and unknown (for probands with limited clinical information; based on prior study by Gillis et al., 2004). For the purposes of the phenotype classification, the following criteria were applied: (1) definite: characteristic facial features, typical limb involvement, small stature, microcephaly, cognitive impairment consistent with a clinical diagnosis of CdLS; (2) possible: meeting criteria for definite but some noncharacteristic features (e.g., normocephalic, no limb involvement); (3) atypical/overlapping (CdLS-like): many overlapping features with CdLS, however, the overall impression was not consistent with a definitive diagnosis of CdLS; and (4) unlikely or not CdLS: clinical features demonstrated overlap with CdLS however upon review features were not felt to be consistent with CdLS. Many of the probands who were referred and enrolled in this study who fell into group 3 and 4 classifications and subsequently found to have variations in genes associated with other diagnoses (most of which were unknown/undescribed at the time of initial enrollment) were enrolled due to their phenotypic overlap with CdLS and are now recognized as either having genetically distinct diagnoses that phenocopy CdLS or are more typical of CdLS even though they were subsequently found to have variants in genes related to a different diagnosis. Severity was assessed based on criteria outlined in Gillis et al., 2004.

2.2 | Sample cohort

We enrolled 2861 subjects including 2016 probands with suspected CdLS and 845 parents and siblings of the probands. A sample (DNA from blood, skin, or saliva) was received for 797 probands. Variant

screening (see details below) was performed on 716 probands. Of the probands tested: both parents were available for 309 (43%), one parent was available for 81 (11%) and 326 (46%) were tested as probands only. In 45 probands, clinical information was very limited and their diagnostic severity was listed as "unknown". The cohort was composed of 672 sporadic (94%) and 44 familial (6%) cases.

2.3 | Analytical methods

As this cohort has been enrolled and samples collected and tested over a 25-year period, various testing modalities have been employed to screen for pathogenic causative variants including: (1) targeted gene screening by conformation-sensitive gel electrophoresis (CSGE) with intronic exon flanking primers along the whole coding sequence, followed by direct Sanger sequencing; (2) multiplex ligation-dependent probe amplification (MLPA); (3) cDNA sequencing; (4) sequencing of conserved noncoding sequences (CNCs); (5) direct sanger sequencing using intronic exon flanking primers; (6) whole-genome SNP genotyping was performed with Illumina (San Diego, CA) Infinium HumanHap550 Beadchip or Affymetrix (Fremont, CA) Genome-Wide Human SNP 6.0 arrays according to the manufacturer's protocols. Copy-number calling was performed with custom algorithms and PennCNV. (Shaikh et al., 2009; Wang et al., 2007); (7) exome sequencing, and, most recently (8) genome sequencing; (9) deletion/duplication analysis of the *NIPBL* gene was performed using Illumina HapMap 550 K and multiplex ligation-dependent probe amplification (MLPA) analysis was performed using the SALSA P141/P142 MLPA kit (MRC-Holland, Amsterdam, the Netherlands). All findings were validated/confirmed by direct sequencing of a second independently amplified PCR product in both forward and reverse directions from the patient's DNA from the respective tissue source. Genome sequencing was performed at the Broad Institute using Illumina NovaSeq with an average read depth of 30×. Alignment and variant calling were carried out using GATK Best Practices workflows (Broad Institute), de novo mutation discovery using GEMINI, variant annotation using Annovar and SnpEff, copy number variation (CNV), and SV analysis conducted using CNVnator and Manta, visualizations done using Interactive Genomics Viewer. The nomenclature of the alterations was based on the mRNA sequence according to the recommendations of the Human Genome Variety Society. Variants were classified based on the American College of Medical Genetics (ACMG) recommendations. Detected pathogenic or potentially pathogenic variants were confirmed by independent PCR reactions followed by bidirectional Sanger sequencing.

2.4 | Mutation analysis methods

All variant nomenclature follows the HGVS nomenclature guidelines (http://www.hgvs.org/mutnomen). The GenBank reference sequences mentioned in this study use version GRCh38/hg38 of the Genome Reference Consortium Human Build. All results were compared with the reference sequences and variants were queried in the gnomAD (https://gnomad.broadinstitute.org), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), and Human Gene Mutation (http://www.hgmd.cf.ac.uk/ac/index.php) databases.

2.5 | Ontologic classification

Biological process, molecular function, and cellular component determinations were made using the gene ontology (GO) database and visualized with gProfiler. Protein families and domains were assigned using the Pfam database and visualized using the trackViewer library for R in R Studio.

3 | RESULTS

3.1 | Overall distribution of variants

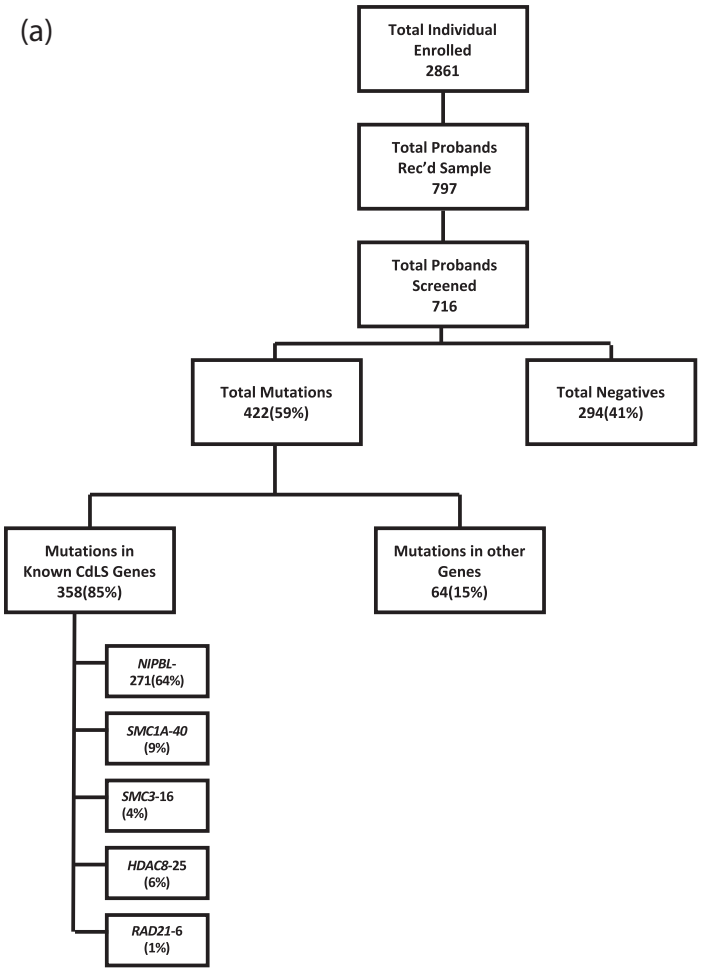
Of the 716 probands tested, pathogenic and likely pathogenic causative variants were identified in 422 (59%; Figure 2). The breakdown of genes in which suspected causative variants were identified is as follows: *NIPBL*: 271 (64%), (158 male: 113 female); *SMC1A*: 40 (9%), (11 male: 29 female); *HDAC8*: 25 (6%), (8 male: 17 female); *SMC3*: 16 (4%), (10 male: 6 female); *RAD21*: 6 (1%), (3 male: 3 females); other causative genes: 64 (15%; see Table S1). Of the identified variants, DNA was available from both parents in 210 families and of these, 207 (99%) variants were de novo.

3.2 | *NIPBL* variants

A total of 271 heterozygous pathogenic variants distributed across the *NIPBL* gene were identified in CdLS probands (Figure 3a): 209 (77%) falling in coding sequences, 50 (18%) in noncoding regions, and 12 (4%) involving gross genomic alterations (Table S1). GenBank NM_133433.4 was used as the *NIPBL* sequence reference. The majority of identified variants are nonsense, splice site, or frameshifts that result in a predicted truncated protein that presumably results in haploinsufficiency. A total of 222 different variants were identified. Causative variants include: 126 (46%) truncating [81 frameshifts (30%) and 45 nonsense (17%)], 76 (28%) missense; 47 (17%) splicing variants; and 7 (3%) in-frame deletions affecting the coding and consensus sequence; 1 mutation (0.4%) in the 5′UTR; 1 (0.4%) deep intronic mutation; 12 (4%) large intragenic deletions; and one balanced translocation.

Variants were identified in all exons with the exception of exons 5, 13, 14, and 25. This finding may suggest that variants within these regions are not tolerated indicating that these exons, and the protein domains they code for, could have critical functional roles that have yet to be determined or variants in these exons do not produce disease. Several exons had multiple variants, including exons 2, 3, 4, 7, 9, 10, 17, 22, 28, 29, 34, 37, 40, 42, 43, 44, 45, and 47. The largest

(a)



(b)

Gene Name	Number (%)	Number (females)	Number (males)
NIPBL	271 (64)	113	158
SMC1A	40 (9)	29	11
SMC3	16 (4)	6	10
HDAC8	25 (6)	17	8
RAD21	6 (1)	3	3
AFF4	3 (<1)	2	1
ANKRD11	6 (1)	4	2
ARCN1	1 (<1)	0	1
ARID1B	3 (<1)	3	0
ASXL2	1 (<1)	0	1
ASXL3	1 (<1)	0	1
BRD4	2 (<1)	0	2
CERT1	1 (<1)	1	0
CHD2	1 (<1)	1	0
EHMT1	1 (<1)	1	0
EP300	3 (<1)	3	0
IQSEC2	1 (<1)	1	0
ITGB8	1 (<1)	1	0
KCNH1	2 (<1)	0	2
KMT2A	2 (<1)	1	1
NAALADL2	2 (<1)	0	2
PACS1	2 (<1)	1	1
PHF6	1 (<1)	1	0
RASAL3	1 (<1)	1	0
SETD5	4 (<1)	1	3
SMARCA2	1 (<1)	0	1
SMARCA4	2 (<1)	0	2
SOX11	1 (<1)	0	1
STAG2	3 (<1)	2	1
TAF1	3 (<1)	0	3
USP7	1 (<1)	0	1
Micro del/dup	14 (3)	5	9
Total	**422 (59%)**		

FIGURE 2 (a) Summary of all probands screened and distribution of causative variants and (b) List of genes with causative variants and prevalence within this population.

single exon cluster of variants, 37 (14%) [11 nonsense, 25 frameshift, and 1 missense], were seen in exon 10 which encodes the coiled-coil region and for the undecapeptide repeat, however, this 1625-bp exon is ~8 times the size of the average exon (~200 bp) in the *NIPBL* gene (Gillis et al., 2004). The majority, 195 (72%), of *NIPBL* variants are unique; however 27 (10%) recurrent variants were identified (Figures 3 and 4) in unrelated probands.

3.3 | NIPBL point variants (missense)

The 76 identified missense pathogenic variants (46 previously reported by our group and 30 reported in this paper), were located along the entire coding sequence (Figure 3a). Additional NIPBL hot spots are suggested by recurrent variants at the same amino acid residue, which affect C1311, R1789, G2381, G2115, A18953, A2338, A2390, and R2298 respectively, and have been previously reported to be mutated in other CdLS probands (Figure 4a). Pathogenic missense substitutions at the highly conserved amino acid residue 2298 in exon 40 were mutated in 14 unrelated probands—six with R2298C, six with R2298H, one with R2298G, and one with R2298P. While most unrelated probands with identical variants in *NIPBL* have similar phenotypes there are exceptions suggesting that other modifying genetic or environmental factors likely impact the CdLS phenotype. While all six individuals identified to have the R2298C mutation were significantly affected and had the more "classic" CdLS phenotype, only two had structural limb-reduction defects and more severe phenotypes, while four had no limb defects and growth and developmental phenotypes ranging from mild to moderate (Figure 4b).

3.4 | In-frame deletions

In-frame deletion variants were identified in seven probands. Recurrent in-frame mutation 6653_6655delATA; N2218delN resulting in deletion of amino acid asparagine in exon 39 was identified in four unrelated probands (three males, one female) with varying degrees of definite mild to moderate phenotypes (Figure 4a). This suggests that, in the case of CdLS, the same genetic change does not always lead to the same degree of disease severity, a phenomenon which is probably influenced by additional, not yet specified, modifying factors (Gillis et al., 2004; Kuzniacka et al., 2013).

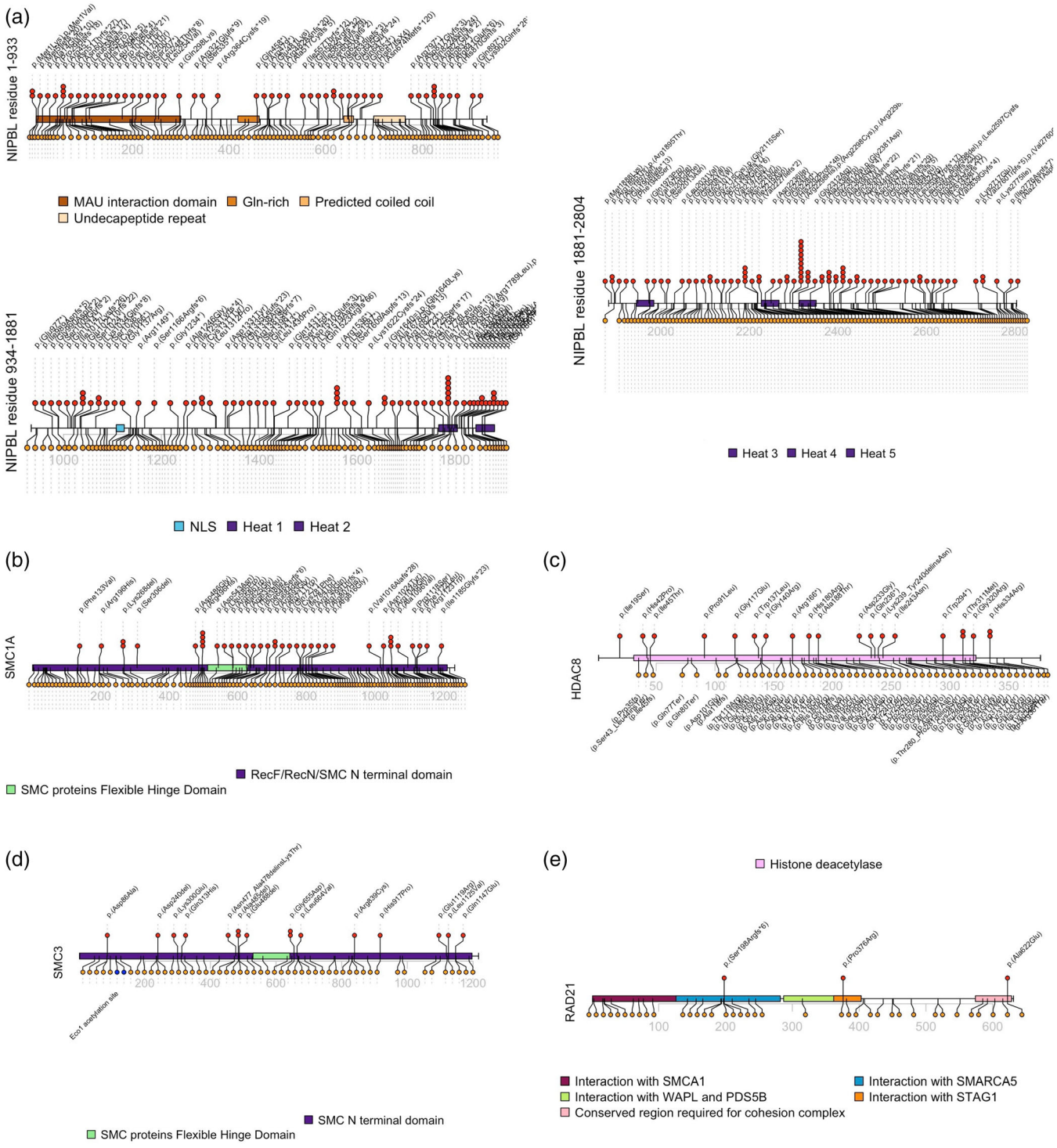


FIGURE 3 Schematic representation of pathogenic variants in (a) *NIPBL* variants, (b) *SMC1A*, (c) *HDAC8*, (d) *SMC3*, and (e) *RAD21* identified in this cohort.

3.5 | NIPBL splicing variants

Forty-seven splicing pathogenic variants were identified in noncoding regions of *NIPBL* affecting donor (36%) and acceptor (64%) splice sites. To our knowledge, the most common effect of splicing site changes is skipping of the downstream exon. Seven recurrent splice site changes that led to the formation of alternative transcripts by aberrant splicing were identified in our cohort: splice donor site variants c.64+1G>A (P6) and c.64+2_3insT (P2) lead to skipping of translation initiation codon carrying exon 2; c.7410+4A>G (P2), c.65-5A>G (P2), c.3855+1G>A (P2), and c.7686-1G>C (P2) were identified in unrelated probands with mild to moderate phenotype. We identified a

(a)

Gene	Nucleotide Change	Protein Change	Number of Probands
Recurrent Missense Mutations			
NIPBL	c.3932G>A	p.(Cys1311Tyr)†	1
NIPBL	c.3931T>C	p.(Cys1311Arg)†	1
NIPBL	c.5366G>A	p.(Arg1789Gln)†	2
NIPBL	c.5366G>A	p.(Arg1789Gln)†	1
NIPBL	c.5366G>T	p.(Arg1789Leu)†	1
NIPBL	c.5366G>C	p.(Arg1789Pro)†	1
NIPBL	c.6343G>T	p.(Gly2115Cys)†	1
NIPBL	c.6343G>A	p.(Gly2115Ser)†	1
NIPBL	c.6892C>T	p.(Arg2298Cys)†	6
NIPBL	c.6893G>A	p.(Arg2298His)†	5
NIPBL	c.6892C>G	p.(Arg2298Gly)†	1
NIPBL	c.6893G>C	p.(Arg2298Pro)†	1
NIPBL	c.7142G>C	p.(Gly2381Ala)†	2
NIPBL	c.7012G>C	p.(Ala2338Pro)	2
NIPBL	c.7168G>A	p.(Ala2390Thr)	4
NIPBL	c.5684G>C	p.(Arg1895Thr)	2
SMC1A	c.3146G>A	p.(Arg1049Gln)	3
SMC1A	c.1487G>A	p.(Arg496His)	4
SMC1A	c.2077C>G	p.(Arg693Gly)†	1
SMC1A	c.2078G>A	p.(Arg693Gln)†	1
SMC3	c.1964G>A	p.(Gly655Asp)	2
HDAC8	c.1001A>G	p.(His334Arg)	2
HDAC8	c.932C>T	p.(Thr311Met)	2
Recurrent In-frame deletions			
NIPBL	c.6653_6655delATA	p.(Asn2218del)	4
SMC1A	c.802_804del3	p.(Lys268del)	2
SMC3	c.1453_1455del3	p.(Ala485del)	2
Recurrent Splicesite Mutations			
NIPBL	c.64+1G>A		6
NIPBL	c.64+2_3insT		2
NIPBL	c.65-5A>G		2
NIPBL	c.3855+1G>A		2
NIPBL	c.7410+4A>G		2
NIPBL	c.7686-1G>C		2
NIPBL	c.5329-15A>G		3
Recurrent Protein truncating Mutations			
NIPBL	c.2479_2480delAG	p.(Arg827Glyfs*2)	3
NIPBL	c.3058_3061delAGAG	p.(Glu1021Thrfs*22)	3
NIPBL	c.3100_3106del7	p.(Lys1034Glnfs*8)	2
NIPBL	c.6679_6682delGTCA	p.(Val2227Ilefs*2)	2
NIPBL	c.7147_7151delAGACA	p.(Gln2384Argfs*4)	2
NIPBL	c.133C>T	p.(Arg45*)	3
NIPBL	c.1885C>T	p.(Arg629*)	2
NIPBL	c.4606C>T	p.(Arg1536*)	4
NIPBL	c.5440C>T	p.(Arg1814*)	2
NIPBL	c.5482C>T	p.(Arg1828*)	2
NIPBL	c.7219C>T	p.(Arg2407*)	2
NIPBL	c.6838C>T	p.(Gln2280*)	2

(b)

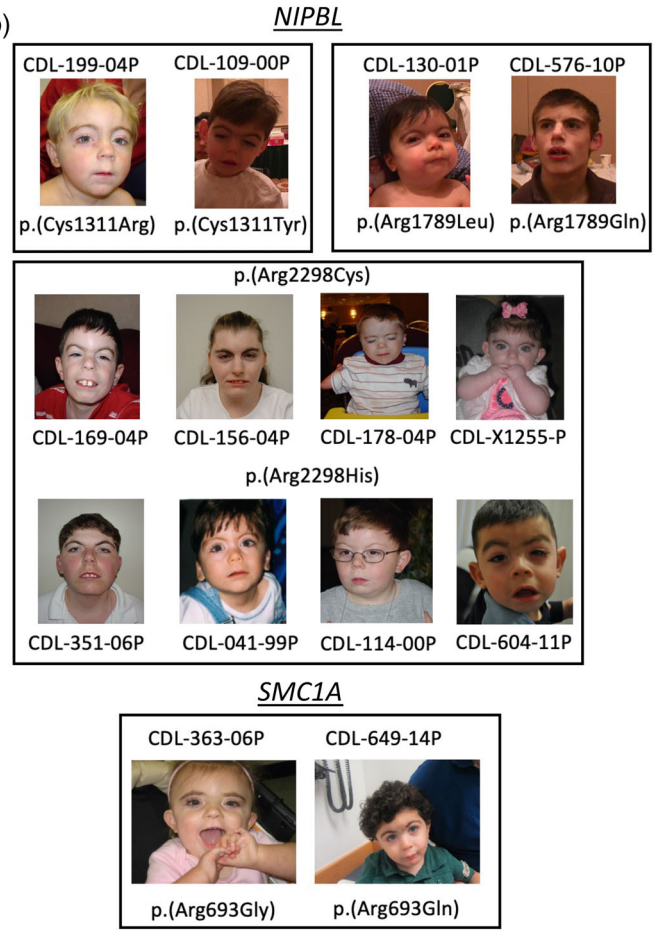


FIGURE 4 Recurrent pathogenic variants in CdLS genes and resultant phenotypes. (a) List of recurrent variants found in the known CdLS genes, † variants affecting same amino residues. (b) Phenotypic representations of a subset of probands listed in (a) with recurrent variants at the same amino acid residue. While most probands with recurrent variants had consistent phenotypic severities there were some exceptions (e.g., for the p.R2298C recurrent variants only 2/5 had severe limb reduction differences as seen in the proband on the right) indicating that while genotype is a strong driver of phenotype there are likely other genetic and environmental modifiers at play.

previously reported mutation in intron 27 of *NIPBL* (c.5329-15A>G) in three unrelated probands with consistent mild phenotypes (Figure 4). This mutation does not affect the conserved splice-donor or acceptor site but results in aberrant mRNA splicing. The resulting aberrantly spliced *NIPBL* transcript excludes a 99 bp fragment representing exon 28, but otherwise preserves the protein reading frame resulting in a slightly shortened, and presumably partially functional, protein (Teresa-Rodrigo et al., 2016). Three synonymous variants (de novo c.4920G>A, p.Gln1640=; de novo c.5427G>A, p.Arg1809=; c.7410G>A, p.Glu2470=) in the last nucleotide of exon 24, 28, and 43, respectively, that affect normal splicing and result in LOF and are predicted to be likely pathogenic were identified in three unrelated probands.

3.6 | NIPBL truncating/nonsense variants

Truncating variants were the most common type of variants 126 (46%). This subgroup included 54 (43%) deletions, 24 (19%) duplications, 3 (2%) insertions/deletions leading to frameshifts, and 45 (36%) nonsense variants, all resulting in premature protein truncation. Truncating variants in *NIPBL* result in a reduced level of functional NIPBL (haploinsufficiency) and typically results in a severe ("classic") CdLS phenotype. Exceptions to this were several probands with distal truncating variants involving the terminal 3′ exons of *NIPBL* resulting in a milder phenotype. Most of the variants are unique, but 12 (seven nonsense and five frameshift) recurrent variants were identified in unrelated probands (Figure 4).

3.7 | NIPBL regulatory variants

A de novo alteration in the 5′ untranslated region of the NIPBL gene c. -79-2A>G was found in a patient with a definite moderate phenotype. This nucleotide change close to the transcription start site presumably results in an alternative transcript or reduction of mRNA level producing haploinsufficiency. The small number of reported regulatory variants suggest that variants in the 5′UTR of the *NIPBL* gene

are rare events and likely not a significant contributor to the ~30% of mutation-negative CdLS probands.

3.8 | NIPBL intragenic copy number variations

Intragenic deletions in *NIPBL* are present in ~2%–5% of patients with CdLS (Bhuiyan et al., 2007; Pehlivan et al., 2012; Russo et al., 2012). Genomic alterations ranging from single to multiple exons, including portions of the nontranslated regulatory regions were detected in 12 (4%) probands. The CNVs ranged in size from 0.8 to 832 kb with the smallest encompassing one exon to those encompassing the entire *NIPBL* gene. Regardless of the size or location of their deletion/duplication, all 12 patients had typical phenotypes consistent with CdLS although some phenotypic variability was observed, with more severe phenotypes correlating with larger deletions. A relatively small deletion of exon 11 (4.2 kb) and exon 2 (4.5 kb) were identified in probands CDL266P, and CDL340P respectively both with definite mild phenotypes. A large deletion of ~832 kb involving almost the entire *NIPBL* gene was identified in CDL341P with a definite severe phenotype. The deletions observed in probands CDL283 (possible moderate phenotype) and CDL454P (definite severe phenotype) both encompassed exons 2–9; although the deletion in each patient is different in size: 32 and 85 kb, respectively. These cases are summarized in Figure 5.

3.9 | SMC1A and SMC3 variants

GenBank NM_006306.4 was used as the *SMC1A* gene reference sequence and GenBank NM_005445.4 was used as the *SMC3* gene reference sequence. Forty (9%) pathogenic variants in *SMC1A* and 16 (4%) pathogenic variants in *SMC3* were identified in this cohort. Parental samples were unavailable for 22 probands, 1 parental sample was unavailable for 3 probands and both parents were available for 15 probands. In all probands in whom both parental samples were available, all variants were confirmed to have arisen de novo. The amino acid residue (R496) was mutated in four unrelated probands, two of which are familial cases (Deardorff et al., 2007); residue R1049Q was mutated in three unrelated probands; and the in-frame deletion c.802_804del3; K268del was observed in two unrelated probands (Figure 4), all other variants were unique (4 in-frame deletions, 4 frameshifts, 1 nonsense, 1 splice site, and 24 missense variants; Table S1). Although variants were seen in more female probands

(a)

CDL#	Dx Certainty	Dx Severity	Inheritance	Deletion	Genomic Location (hg38)
CDL-034-99P	unlikely		de novo	deletion 832.7kb incl entire gene	chr5:36,255,992-37,088,772
CDL-141-03P	possible	Mild	de novo	deletion 98.5kb incl ex. 1-8	chr5:36,902,490-37,001,062
CDL-223-05P	definite	Severe	unknown*	deletion 66kb incl. ex. 2-17	chr5:36,976,473-37,042,330
CDL-266-05P	definite	Mild	de novo	deletion 4.2kb incl ex. 11	chr5:37,029,127-37,033,402
CDL-283-05P	possible	Moderate	unknown*	deletion 32kb incl ex. 2-9	chr5:36,987,804-37,019,560
CDL-340-06P	definite	Mild	de novo	deletion ex. 2	chr5:36,985,357-36,989,954
CDL-341-06P	definite	Severe	de novo	deletion ~832kb incl ex. 1-42	chr5:36,220,235-37,053,015
CDL-406-07P	possible	Mild	unknown*	deletion 18kb incl ex. 2-6	chr5:36,982,357-37,000,404
CDL-434-07P	definite	Severe	unknown*	deletion 3.4kb incl intron 1	chr5:36,879,320-36,882,710
CDL-X1296-P	definite	Severe	unknown*	deletion incl ex. 37-38	
CDL-454-07P	definite	Mild	de novo	deletion ~85kb incl ex. 2-9	
CDL-632-13P	atypical/overlap		de novo	5p13.2 duplication 1.63Mb incl *NIPBL* gene	chr5:35,805,105-37,445,000

(b)

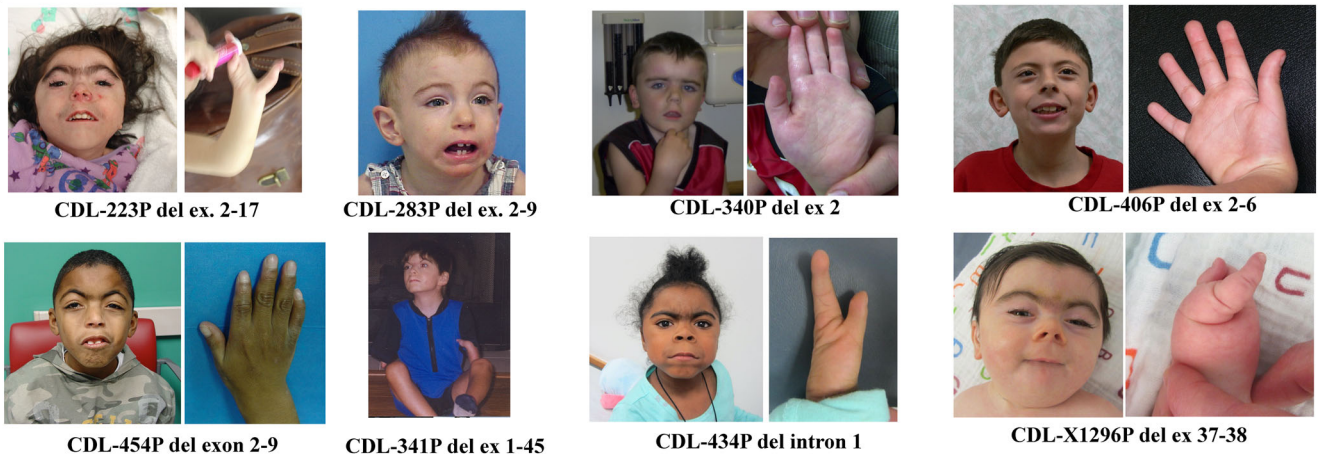


FIGURE 5 *NIPBL* intragenic copy number variations (CNVs) in CdLS probands. (a) List of *NIPBL* CNVs identified in this cohort with diagnostic certainty and severity scores. (b) Phenotypic representation of a subset of these probands with characteristic but variable involvement of facial features and upper limbs.

(29) compared to male probands (11) the level of severity between male and female probands is equal (Liu et al., 2009). All mutated residues affect evolutionarily conserved amino acids (Deardorff et al., 2007). Notably, the *SMC3*- and *SMC1A*-mutation-positive probands demonstrated a milder phenotype overall than probands with *NIPBL* pathogenic variants with an absence of major structural limb differences, although other organ anomalies were observed this including in *SMC1A*: CDH 6/40 (15%), cleft, GI, renal, brain malformation 1/40 (3%) and in *SMC3*: CDH 3/16 (19%), cleft and GI 1/16 (6%). A total of 14 different *SMC3* pathogenic variants in 16 unrelated probands with mild to moderate phenotypes were identified. Of the unique variants, 10 were missense, 3 were in-frame deletion, and 1 was del/ins frameshift. One recurrent in-frame deletion c.1453_1455del3 resulting in the deletion of a single amino acid (p.Ala485del) was found in two unrelated probands (Figure 4).

3.10 | *HDAC8* variants

GenBank NM 018486.3 was used as *HDAC8* sequence reference. *HDAC8*, located on chromosome Xq13.1, encodes for a histone deacetylase that deacetylates SMC3 during S-phase to establish the cohesiveness of chromatin-loaded cohesin. We have identified 25 individuals (6%) with causative variants in *HDAC8*, ranging from copy number abnormalities through single nucleotide missense substitutions.

Variants in *HDAC8* account for ~6% of variants in patients in our cohort (8 males, 17 females). De novo pathogenic variants include 16 missense, 3 nonsense, 1 splice site, 1 in-frame deletion leading to the deletion of an amino acid, and 4 microdeletions. Hemizygous males are more severely affected; in females the severity is strongly influenced by the level of X inactivation of the mutation. The c.1001A>G variant has been identified in a family with an affected boy, his mildly affected sister, and his unaffected mother, in which the mutant allele was inactivated in her blood. Functional studies showing complete skewing toward the normal allele in the blood of affected females suggests a strong selection against the *HDAC8* mutant allele. The missense variants c.539A>G;p.H180R and c.958G>A;p.G320R seen in patients with definite moderate phenotypes both led to a reduced level of HDCA8 protein in fibroblasts and lymphoblastoid cells suggesting these *HDAC8* variants can cause protein instability (Mannini et al., 2013).

3.11 | *RAD21* variants

GenBank NM 006265.3 was used as *RAD21* sequence reference. *RAD21* (also known as *KIAA0078*, *NXP1*, *HR21*, *Mcd1*, *Scc1*, and hereafter called *RAD21*) encodes a DNA double-strand break repair protein that is evolutionarily conserved in all eukaryotes from budding yeast to humans (Cheng et al., 2020). RAD21 (ENSG00000164754; OMIM *606462) is a key structural component of the cohesin complex, it forms a tri-partite ring together with SMC1A and SMC3.

RAD21 variants are found in a minority of CdLS individuals. To date, nine missense variants and five microdeletions have been reported in CdLS individuals (Kline et al., 2018). We identified two missense pathogenic variants (p.Pro376Arg and p.Ala622Glu), one frameshift pathogenic variant (p.Ser1286Leufs*84), and three de novo deletions including *RAD21* in our cohort. RAD21 interacts with the other cohesin subunits, SMC1A, SMC3, and STAG, to maintain the ring-like structure of the cohesin complex. It has been suggested that p.Pro376Arg variant might interfere with cohesin activity by increasing the binding of STAG to RAD21 (Deardorff, Bando, et al., 2012).

3.12 | CNVs not encompassing known CdLS loci

Rare CNVs not encompassing known CdLS Loci were identified in 15 probands (4%). These likely pathogenic variants calls based on absence of CNV in unaffected parents when available, size of the CNV (generally greater than 1 Mb, and absence of CNV in control databases [e.g., database of genomic variations DGV] http://dgv.tcag.ca/dgv/app/home) include single heterozygous microdeletions: de novo 1q25.3-q32.1del, 15.7 Mb; de novo 9q34.3 del, 1.49 Mb; de novo 11q24.3ter del, 115.8 Mb; unknown 2q23.1q23.3del, 5.14 Mb; de novo 4q21.1-q21.22 del, 5.32 Mb and microduplications: de novo 22q11.22dup, 2.6Kb; de novo 19p13.2p13.13dup, 0.51 Mb; unknown 4p11dup, 49.5 Mb (Figure 6). More than one CNV was identified in four probands. 2.5 Mb and 12.8 Mb deletions of 3q24-26.1 were identified in CDL091P; the mother was negative for both deletions and father's sample was unavailable. Deletion of chromosome 1q43q44del and Xq22.3del along with 3.92 Mb duplication at 4q35.1q35.2 were found in CDL516P, parental samples were unavailable. A 3.23 Mb duplication of 12p13.33p13.32 and 9.6 Mb deletion of 3p26.3-p26.1 that includes possible contributing gene *SETD5*, were identified in CDL142P and his affected sibling; both CNVs were absent in the mother, and father was not tested. Multiple de novo duplications at 1q23.3, 1q24.1-q24.2, 1q24.3, 1q32.2-q41, 1q41-43 ranging in size from 0.2 to 15.9 Mb along with 15.9 Mb deletion of chromosome 1q24.3-q32.1 were found in CDL219P. The 4.8 Mb 1q32.2-q41 duplication includes a likely contributory gene *KCNH1* (Tables S1 and S2).

3.13 | Genome sequencing in mutational negative CdLS probands

Genome sequencing was performed on 178 CdLS probands for whom targeted CdLS gene mutational analyses failed to identify a cause. In 60 probands (34%), causative variants were identified. In 23 probands (13%), genome sequencing identified variants in known cohesin genes, that were not screened or missed on earlier panels or were present in deep intronic regions not captured on gene panels or exomes. In 37 probands (21%), causative variants were identified in known disease-causing genes that were typically associated with other diagnoses that either overlap or resemble the CdLs phenotype (*ANKRD11*,

(a)

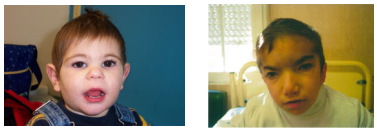
CDL#	Dx Certainty	Dx Severity	Inheritance	CNV-genomic location(hg38)	Type
CDL-042-99P	possible	Moderate	de novo	1q25.3-q32.1del, 15.7Mb (chr1:185,795,497-201,517,800)	Microdeletion
CDL-383-06P	possible	Moderate	unknown	2q23.1q23.3del, 5.14Mb (chr2:148705302-153843317)	Microdeletion
CDL-078-99P	possible	Mild	unknown	4p11dup, 49.5Mb (chr4:73,616-49,630,633)	Microduplication
CDL-467-08P	unlikely	Mild	de novo	4q21.1-q21.22 del, 5.32Mb (chr4:77,722,251-83,038,589)	Microdeletion
CDL-103-00P	atypical	moderate	de novo	9q34.3 del, 1.49Mb (chr9:136,640,251-138,133,487)	Microdeletion
CDL-117-01P	definite	Mild	de novo	11q24.3ter del, 115.8Mb (chr11:13,467,115-129,331,531)	Microdeletion
CDL-257-05P	possible	Mild	de novo	19q13.32dup, 0.8Mb(chr19:46,640,111-47,446,095)	Microduplication
CDL-324-06P	atypical/overlap		de novo	19p13.2p13.13dup, 0.51Mb (chr19:12,607,805-13,121,181)	Microduplication
CDL-348-06P	possible	Mild	de novo	22q11.22dup, 2.6Kb (chr22:21,960,940-22,225,921)	Microduplication
CDL-091-99P	unlikely		unknown	3q24del, 2.5Mb (chr3:145,075,235-147,588,004)	Microdeletion
				3q24-26.1del, 12.8Mb (chr3:148,529,532-161,405,738)	Microdeletion
CDL-516-08P	unlikely		unknown	ish der(1)t(1;4) (q43;q35.1) (D1Z1+, 1QTEL10-, D4S2930+) 1q43q44del (chr1:242,060,041-248,916,508) 4q35.1q35.2 dup (chr4:186119275-190044201)	FISH Microdeletion Microduplication
				Xq22.3del (chrX:105,935,140-106,429,487)	Microdeletion
CDL-142-03P	possible		de novo	12p13.33p13.32 dup, 3.23Mb (chr12:84,652-3,318,614)	Microduplication
				3p26.3-p26.1del, 9.6Mb (chr3:21,733-9,664,493)	Microdeletion
CDL-219-05P	possible	Moderate	de novo	1q23.3dup, 0.5Mb (chr1:164,292,317-164,811,778) 1q24.1-q24.2dup, 0.38Mb (chr1:166,301,458-170,147,101) 1q24.3dup, 0.2Mb (chr1:171,314,704-171,517,773) 1q24.3-q32.1del, 26.3Mb (chr1:172,883,263-199,192,634) 1q32.2-q41dup, 4.8Mb (chr1:210,240,394-215,019,935)	Microduplication Microduplication Microduplication Microdeletion Microduplication
				1q41-43dup, 15.9Mb (chr1:222,311,403-238,288,928)	Microduplication
CDL-154-00P	unknown	unknown/limited clinical info	de novo	4q32.2ter del, 13q31.1ter dup	unbalanced translocation
CDL-512-09P	unlikely	Mild	unknown	10q26.3del, 4.6Mb (128,934,414-133,622,588) & 18p11.22-p11.32dup, 7.5Mb (chr18:85,432-7,872,307)	unbalanced translocation

(b)

Microdeletions

CDL-117-01P CDL-467-08P CDL-348-06P CDL-383-06P

Microduplications



CDL-257-05P CDL-219-05P

FIGURE 6 Chromosomal position and boundaries of rare CNVs not encompassing known CdLS Loci. (a) Chromosomal coordinate and phenotypes of 15 probands with CNVs. (b) Representative facial features of six of these probands.

(a)

Gene	Location	Phenotype	OMIM #	Gene Function	MOI	# probands
AFF4	5q31.1	CHOPS Syndrome	616368	Component of super elongation complex (SEC)	AD	3
ANKRD11	16q24.3	KBG syndrome	148050	Chromatin regulator inhibits ligand-dependent activation of transcription	AD	6
ARCN1	11q23.3	Short stature, rhizomelic, with microcephaly, micrognathia, and developmental delay	617164	Component of the coatomer protein complex responsible for intracellular protein transport between endoplasmic reticulum and golgi	AD	1
ARID1B	6q25.3	Coffin-Siris syndrome 1	135900	Member of the human SWI/SNF chromatin remodeling complex	AD	3
ASXL2	2q24.3	Shashi-Pena syndrome	617190	Putative Polycomb group (PcG) protein required to maintain transcriptionally repressive state of homeotic genes throughout development.	AD	1
ASXL3	18q12.1	Bainbridge-Ropers syndrome	615485	Chromatin remodeling and transcriptional regulation.	AD	1
BRD4	19q13.33		608749	Transcriptional and epigenetic regulator that plays a pivotal role during embryogenesis and cancer development.		2
CERT1	5q13.3	Mental retardation, autosomal dominant 34	616351	Mediates intracellular trafficking of ceramides	AD	1
CHD2	15q26.1	Epileptic encephalopathy, childhood-onset	615369	Chromatin remodeling	AD	1
EP300	22q13.2	Rubinstein-Taybi Syndrome(RSTS	613684	Histone acetyltransferase that regulates transcription via chromatin remodeling	AD	3
IQSEC2	Xp11.22	Intellectual developmental disorder, X-linked	309530	Guanine nucleotide exchange factor for the ARF GTP-binding proteins, expressed in neurons, is involved in cytoskeletal organization, dendritic spine morphology, and excitatory synaptic organization	XLD	1
ITGB8	7p21.3		604160	Receptor for fibronectin		1
KCNH1	1q32.2	Temple-Baraitser syndrome	611816	Pore-forming (alpha) subunit of a voltage-gated delayed rectifier potassium channel	AD	2
KMT2A	11q23.3	Wiedemann-Steiner syndrome	605130	Histone H3K4 methyltransferase regulation of gene expression	AD	2
NAALADL2	3q26.31		608806	Unknown function: disrupted at the balanced *de novo* translocation breakpoint in proband (Tonkin et al 2004)		2
PACS1	11q13.1-q13.2	Schuurs-Hoeijmakers syndrome	615009	Plays a putative role in the localisation of trans-Golgi network (TGN) membrane proteins.	AD	2
PHF6	Xq26.2	Borjeson-Forssman-Lehmann syndrome	301900	Transcriptional regulator that associates with ribosomal RNA promoters and suppresses ribosomal RNA (rRNA) transcription.	XL	1
RASAL3	19q13.32		616561	Functions as a Ras GTPase-activating protein.		1
SETD5	3p25.3	Mental retardation, autosomal dominant 23	615761	Chromatin regulator	AD	4
SMARCA2	9p24.3	Nicolaides-Baraitser syndrome	601358	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin	AD	2
SMARCA4	19q13.2	Coffin-Siris syndrome 4	614609	Component of SWI/SNF chromatin remodeling complexes, transcriptional activation and repression of select genes by chromatin remodeling	AD	2
SOX11	2p25.2	Coffin-Siris syndrome 9	615866	Transcription factor that acts as a transcriptional activator	AD	1
STAG2	Xp25	Mullegama-Klein-Martinez syndrome	301022	Subunit of the cohesin complex, which regulates the separation of sister chromatids during cell division.	XL	3
TAF1	X q13.1	Intellectual disability syndrome	300966	Transcription initiation factor	XL	3
USP7	16p13.2	Hao-Fountain syndrome	616863	Hydrolase that deubiquitinates target proteins such as p53 and WASH	AD	1

(b)

FIGURE 7 (a) Novel and atypical genes identified to have causative variants in this CdLS cohort with (b) representative photos of affected individuals.

ARCN1, *ARID1B*, *ASXL2*, *ASXL3*, *BRD4*, *CERT1*, *CHD2*, *EP300*, *IQSEC2*, *KCNH1*, *KMT2A*, *PACS1*, *PHF6*, *SETD5*, *SMARCA2*. *SMARCA4*, *SOX11*, *STAG2*, *TAF1*, *USP7*). In four probands (2%), a strong novel CdLS candidate gene was identified (*NAALADL2*, *ITGB8*, and *RASAL3* on genome sequencing; summarized in Figure 7). The clinical overlap between these syndromes suggests dysregulation of common genes and pathways (Izumi, 2016; Sarogni et al., 2020). Variants in these genes were identified in a single proband except for *ARID1B* in three probands, *SETD5* in four probands, *ANKRD11* in six probands, *EP300*, *KCNH1*, *KMT2A*, *SMARCA4*, *NAALADL2* in two probands.

4 | DISCUSSION AND CONCLUSION

CdLS is a genetically heterogeneous diagnosis that presents with extensive phenotypic variability, including facial dysmorphism, developmental delay/intellectual disability, behavioral differences, hypertrichosis, and variable structural abnormalities of the limbs, heart, palate, intestines, diaphragm, genitourinary system, and others. Features vary widely among affected patients and range from relatively mild involvement to severe manifestations. This study examined the genetic contributors to CdLS in a cohort of 716 probands with a

diagnosis of CdLS referred to the CdLS Center at CHOP over a 25-year period. Through variable molecular diagnostic approaches (driven by technology changes over the years of enrollment), a molecular etiology was able to be identified in 423/716 (59%). Of the identified causative variants 85% were in the previously known CdLS genes (*NIPBL*, *SMC1A*, *SMC3*, *HDAC8*, and *RAD21*) with *NIPBL* variants representing the vast majority (64% of variants overall). Causative variants in other genes (*AFF4*, *ANKRD11*, *ARCN1*, *ARID1B*, *ASXL2*, *ASXL3*, *BRD4*, *CERT1*, *CHD2*, *EP300*, *IQSEC2*, *ITGB8*, *KCNH1*, *KMT2A*, *NAALADL2*, *PACS1*, *PHF6*, *RASAL3*, *SETD5*, *SMARCA2*, *SMARCA4*, *SOX11*, *TAF1*, *USP7*) accounted for a small but significant number (15%) of etiologies. These additional genes fall into the following categories: genes that cause other well-recognized syndromes with phenotypes that overlap with CdLS (e.g., *EP300*, *ANKRD11*, *ASXL2*, and *3*, *SMARCA2*), CdLS phenocopy diagnoses caused by genes that have a role in cohesin function (e.g., *BRD4*, *AFF4*), CdLS phenocopy diagnoses caused by genes not known to have a direct role in cohesin function (e.g., *IQSEC2*, *SETD5*, *PACS1*), and potentially novel CdLS genes (e.g., *NAALADL2*, *ITGB8*, *RASAL3*). The probands with *NAALADL2* variants (one de novo missense c.511A>C, p.Thr171Pro and a nonsense mutation of unknown inheritance, c.2098A>T, p.Arg700*) are of particular interest as this gene lies at the chromosome 3q26.3 breakpoint in a CdLS patient with an apparently balanced translocation [t(3;17)(q26.3;q23.1)] (Ireland et al., 1991; Tonkin et al., 2004), although screening for variants in this gene in other CdLS probands was negative. These two probands have significant clinical overlap with the CdLS phenotype and were classified as "possible mild" and "possible moderate," respectively.

All genes identified to date are either autosomal or X-linked dominant, with 99% of variants for which a sample was available from both parents being de novo. In this cohort, 27 familial recurrent variants were identified. There is a bias in the number of recurrent familial cases in our cohort as these families were specifically recruited to the study due to their importance in gene mapping/identification studies and are therefore overrepresented. Familial recurrences are due to three reported mechanisms: dominant transmission directly from a mildly affected parent (Russell et al., 2001), X-linked transmission from an unaffected carrier mother (this is extremely rare, with most X-linked gene (*SMC1A*, *HDAC8*) variants occurring as de novo events; Musio et al., 2006), and transmission through germline mosaicism (Russell et al., 2001; Slavin et al., 2012). Germline mosaicism was proposed as the pathogenic mechanism for families in which parents were mutation negative in their blood but who had multiple affected children carrying the same *NIPBL* mutation (Slavin et al., 2012), and proven in a father with multiple affected offspring who carried a pathogenic *NIPBL* missense mutation c.7298A>G in his semen but not in his peripheral blood (Niu et al., 2006). In 20 familial cases in which a *NIPBL* mutation has been identified, all of the variants were identified in the probands and affected family members only, and no unaffected family members carried the mutation, with paternity confirmed in all cases (Krantz et al., 2004). In our cohort, all familial recurrences were due to presumed germline mosaicism.

The identification of causative variants in genes encoding components of chromosome remodeling, chromatin regulators, and transcription regulation machineries, suggests that common pathways involving developmental transcriptional control (Figure 8a,b) result in developmental diagnoses that have overlapping phenotypes, likely due to disruption of common downstream pathways. Critical downstream developmental genes could either be disrupted through different chromatin/transcriptional pathways or roles yet to be identified in the regulation of the complex. This concept of a common molecular etiopathogenesis has been suggested by several authors who argue that this group of diagnoses be termed "transcriptomopathies" or "disorders of transcriptional regulation (DTRs)" (Cucco et al., 2020; Izumi, 2016; Parenti et al., 2016, 2017; Sarogni et al., 2020; Yuan et al., 2015). Figure 8c hierarchically represents the statistically significant HPO terms seen in common among probands with variants in the causative genes identified in this study. Some of these features include microcephaly, facial dysmorphism (long eyelashes, abnormal eyebrow morphology/synophrys, anteverted nares, differences of the philtrum and thin vermilion of the upper lip), limb differences, hypertrichosis, and structural differences that are commonly associated with CdLS such as volvulus and CDH.

It is important to note that in 41% of our cohort, we were unable to find a convincing molecular etiology. Given the striking clinical involvement of all enrolled probands with a CdLS or CdLS-like phenotype, we feel certain that this cohort is extremely strongly enriched for underlying genetic causes for their clinical presentations, and for likely dominantly acting de novo pathogenic variants. There are multiple possible explanations for this relatively high mutation negative rate including: (1) Yet to be identified novel CdLS genes; (2) cryptic noncoding mutational events such as regulatory or deep intronic variants in known or novel CdLS genes that were not identified on DNA-based screening. These types of mutational mechanisms have been documented in CdLS with cryptic deep intronic variants affecting splicing being identified by RNA sequencing in the *NIPBL* gene (Rentas et al., 2020); or (3) the known presence of tissue-specific mosaicism documented in CdLS, that has been predicted to be as high as 15%–20% (Ansari et al., 2014; Gonzalez Garcia et al., 2020; Krawczynska et al., 2019; Latorre-Pellicer et al., 2021). In our cohort, the vast majority of collected samples on which mutational analyses were performed were from blood samples. We are in the process of collecting buccal swabs from the mutation-negative probands to assess for the potential of tissue-specific mosaicism. To date, we have identified only 0.9% of probands with a mosaic contribution for any of the known CdLS genes.

5 | GENOTYPE–PHENOTYPE CORRELATION

There is a wide range of severity of clinical characteristics observed in patients with CdLS, including typical facial features, growth retardation, intellectual disability, limb defects, and involvement of other systems. These features vary widely among affected patients and range from relatively mild to severe. In the majority of unrelated probands sharing the same mutation (and in the vast majority of familial recurrences), the phenotypes are generally consistent. However, some

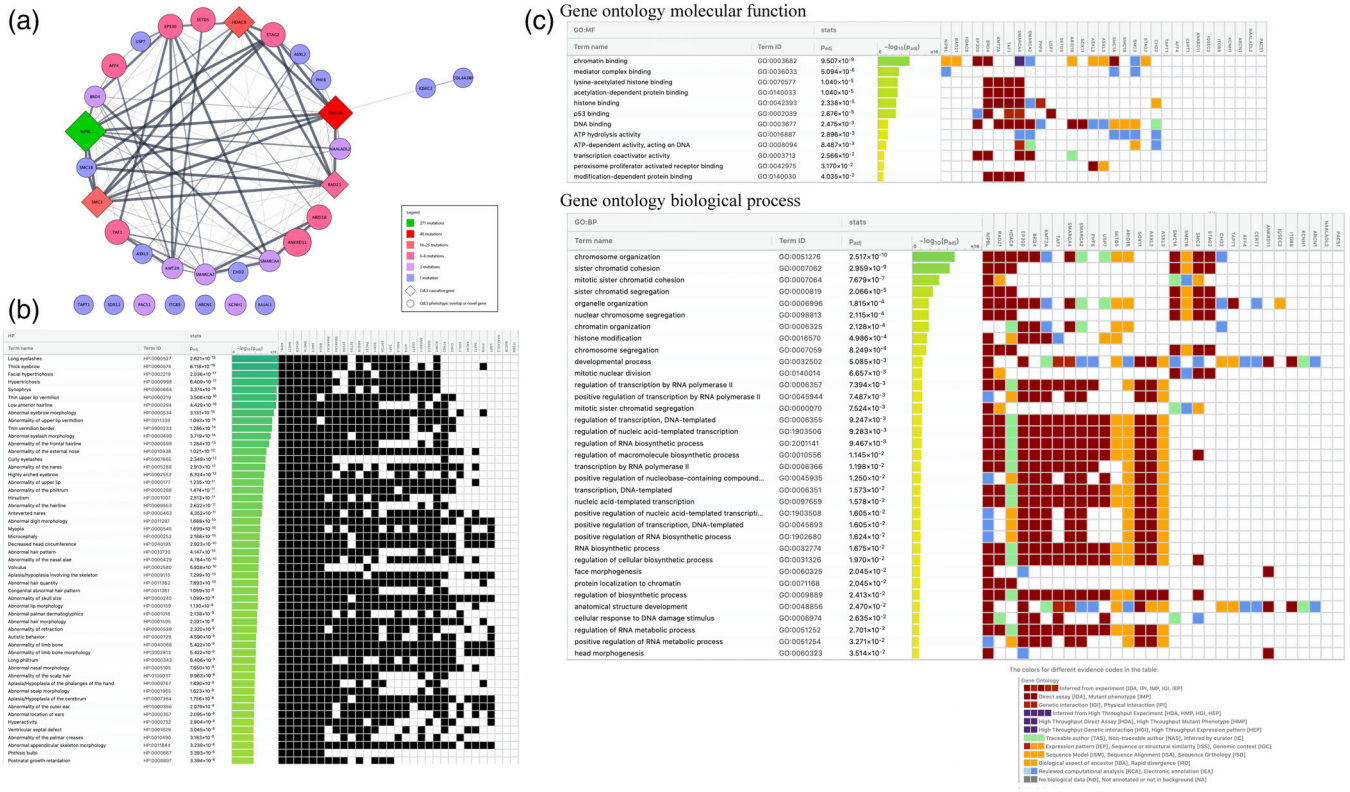


FIGURE 8 (a) Protein–protein interactions among genes with identified variants. Core CdLS genes are indicated by diamond shapes, the prevalence of variants indicated by the size of shapes, and the strength of interactions between proteins indicated by the width of lines. (b) HPO terms associated with mutated genes identified in this study. (c) Gene ontologies by molecular function and biological processes.

differences in phenotype are observed, suggesting that other genetic, or environmental, factors may modify expression of the CdLS phenotype. Analysis of the mutational spectrum reveals a genotype–phenotype correlation. Patients with *NIPBL* variants are likely to present with more severe "classic" clinical features and to have more impaired cognitive function than those with other causal variants. A presumably nonfunctional NIPBL protein caused by truncating variants (nonsense, splice site, and frameshift variants) usually produces a more severe cognitive and structural phenotype. Frame-preserving variants (splice variants and small deletions) of *NIPBL* are generally associated with a milder phenotype. Missense variants produce a variable effect depending on the involved protein region, for example, missense variants in the HEAT domain of *NIPBL* result in severe clinical phenotypes (Mannini et al., 2013) while most other missense variants produce milder phenotypes. Limb reduction defects are almost exclusively seen in *NIPBL*-mutated individuals, in agreement with a previous study (Gillis et al., 2004). The association of splice-site variants with a moderate phenotype is interesting and deserves further study (Figure 9).

The clinical picture of patients with CdLS carrying SMC*1A*, *SMC3*, and *RAD21* pathogenic variants is more uniform, characterized by a mild to moderate phenotype. Individuals with *SMC1A* and *SMC3* pathogenic variants are very similar in their phenotypic presentations and have minimal structural involvement beyond the craniofacial dysmorphism, impaired growth, and malrotations, however, cardiac defects were found in a small percent of individuals with variants in *SMC1A* (15%) and *SMC3* (19%), and a small percent of individuals (3%–8%) had cleft palates. Individuals with *RAD21* causative variants tend to have characteristic facial dysmorphism and growth delays but milder cognitive involvement. Finally, individuals harboring pathogenic variants in the *HDAC8* gene show typical facial dysmorphism and severe cognitive delay with additional features (e.g., delayed closure of the anterior fontanelle, hooded eyelids, widely spaced eyes, and a wide nose) that are specific to individuals with variants in this gene (Boyle et al., 2015; Cucco et al., 2020; Deardorff et al., 2007; Deardorff, Bando, et al., 2012; Deardorff, Wilde, et al., 2012; Gillis et al., 2004; Gil-Rodriguez et al., 2015; Mannini et al., 2013; Sarogni et al., 2020).

SMC1A pathogenic variants can result in two different phenotypes depending on the underlying mutational mechanism. Individuals with missense and frame-preserving del/dup variants present with a typical CdLS phenotype, with males and females similarly affected. Truncating variants in *SMC1A* result in a severe neurocognitive seizure phenotype termed "SMC1A-related neurocognitive disorder" or "Developmental and epileptic encephalopathy 85 with or without midline brain defects (DEE85)," a severe disorder with epilepsy and the often profound neurocognitive impact that can mimic Rett syndrome. While affected individuals can be small and microcephalic they do not typically have the other characteristic phenotypic findings seen in CdLS. Truncating variants in *SMC1A* are almost exclusively seen in

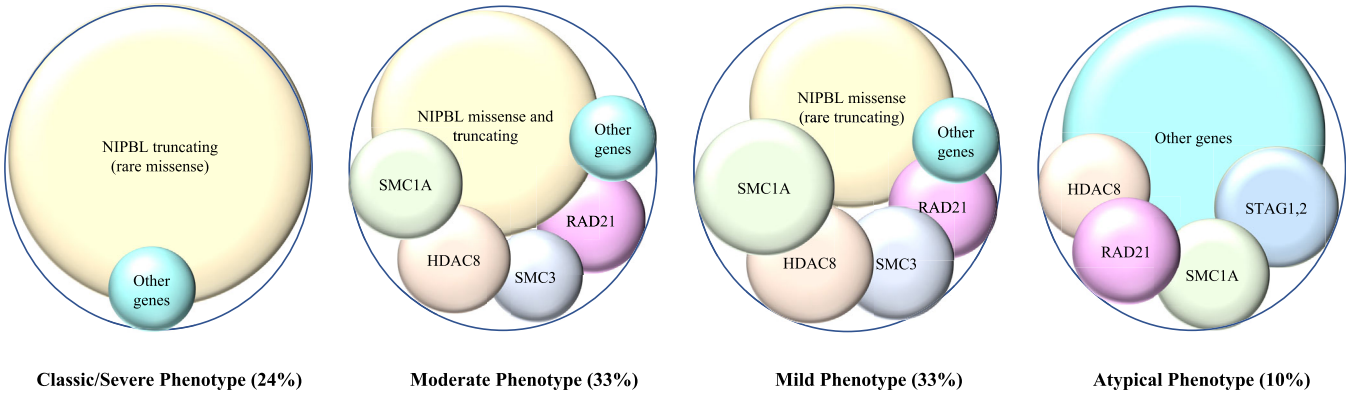


FIGURE 9 Genotype–phenotype correlations in CdLS and related diagnoses. The genetic contributors to the various phenotypic subclassifications of CdLS include a predominance of *NIPBL* truncating variants contributing to the "classic/severe" CdLS phenotype with rare *NIPBL* missense variants in critical domains as well as possible other mutational mechanisms/novel genes contributing to the small percent classic/severe CdLS probands in which a mutation has not been identified. The moderate phenotype is caused predominantly by missense and more terminal truncating variants in NIPBL as well as by variants in most of the other cohesin-related CdLS genes (*SMC1A*, *SMC3*, *HDAC8*, *RAD21*) with some variants in non-cohesin related genes and additional mechanisms/genes still to be identified. The mild CdLS phenotype demonstrates a similar distribution with a greater representation of non-NIPBL-related variants. The "atypical" phenotypes are primarily caused by variants in non-cohesin related genes, however, there is a smaller contribution of cohesin gene mutation as well (e.g., truncating variants in *SMC1A*, *HDAC8*, *RAD21*, and the *STAG* genes).

females and presumably are nonviable in males (Huisman et al., 2017). Rarely, we have observed some affected females with truncating variants that appear to have an intermediate phenotype between CdLS and DEE85.

These studies underscore the broad phenotypic variability of CdLS and the significant overlap of this diagnosis with many other diagnoses caused by genes that encode proteins involved in chromatin and transcriptional regulation. The significant number of individuals diagnosed with CdLS for which an underlying genetic etiology has not been identified to date suggests that additional mutational mechanisms (e.g., mosaicism, noncoding variations) need to be more thoroughly investigated in this cohort and that additional CdLS-related genes and mechanisms remain to be identified. The growing recognition of phenotypic and molecular overlap among DTRs suggests common developmental pathways/downstream genes, which could be targeted for therapies that could benefit many disorders in addition to CdLS.

AUTHOR CONTRIBUTIONS

Maninder Kaur: Designed and performed the experiments, processed the experimental data, performed the analysis collected and analyzed data, drafted and wrote the paper with input from all authors, designed the figures. **Justin Blair:** Contributed data acquisition and data analysis tools and performed the manuscript's analysis and writing. **Batsal Devkota:** Contributed data acquisition, and data analysis tools and performed the analysis. **Sierra Fortunato:** Sample and data collection. **Dinah Clark:** Sample and data collection. **Audrey Lawrence:** Data collection. **Jiwoo Kim:** Whole genome sequencing data analysis. **Wonwook Do:** Performed the experiments and analyzed the data. **Benjamin Semeo:** Performed the experiments and analyzed the data. **Olivia Katz:** Performed the experiments and analyzed the data. **Devanshi Mehta:** Performed the experiments and analyzed the data. **Nobuko Yamamoto:** Performed whole genome sequencing data analysis. **Emma Schindler:** Data analysis and helped write the manuscript. **Zayd Al Rawi:** data analysis. **Nina Wallace:** Performed the experiments and analyzed the data. **Jonathan J. Wilde:** Performed the experiments and analyzed the data. **Jennifer McCallum:** Performed the experiments and analyzed the data. **Jinglan Liu:** Performed the experiments and analyzed the data. **Dongbin Xu:** Performed the experiments and analyzed the data. **Marie Jackson:** Sample and data collection. **Stefan Rentas:** Performed RNAseq and data analysis. **Ahmad Abou Tayoun:** Performed RNAseq and data analysis. **Zhang Zhe:** Processed the experimental data, and performed the analysis. **Omar Abdul-Rahman:** Contributed patient information and samples. **Bill Allen:** Contributed patient information and samples. **Moris A Angula:** Contributed patient information and samples. **Kwame Anyane-Yeboa:** Contributed patient information and samples. **Jesús Argente:** Contributed patient information and samples. **Pamela H Arn:** Contributed patient information and samples. **Linlea Armstrong:** Contributed patient information and samples. **Lina Basel-Salmon:** Contributed patient information and samples. **Gareth Baynam:** contributed patient information and samples. **Lynne M. Bird:** Contributed patient information and samples, reviewing and editing of manuscript: **Daniel Bruegger:** Contributed patient information and samples. **Gaik-Siew Ch'ng:** Contributed patient information and samples. **David Chitayat:** Contributed patient information and samples. **Robin Clark:** Contributed patient information and samples. **Gerald F. Cox:** Contributed patient information and samples. **Usha Dave:** Contributed patient information and samples. **Elfrede DeBaere:** Contributed patient information and samples. **Michael Field:** Contributed patient information and samples. **John M. Graham:** Contributed patient information and samples. **Karen W. Gripp Jr.:** Contributed patient information and

samples. **Robert Greenstein:** Contributed patient information and samples. **Neerja Gupta:** Contributed patient information and samples. **Randy Heidenreich:** Contributed patient information and samples. **Jodi Hoffman:** Contributed patient information and samples. **Robert J. Hopkin:** Contributed patient information and samples. **Kenneth L. Jones:** Contributed patient information and samples. **Marilyn C. Jones:** Contributed patient information and samples. **Ariana Kariminejad:** Contributed patient information and samples. **Jillene Kogan:** Contributed patient information and samples. **Baiba Lace:** Contributed patient information and samples. **Julian Leroy:** Contributed patient information and samples. **Sally Ann Lynch:** Contributed patient information and samples. **Marie McDonald:** Contributed patient information and samples. **Kirsten Meagher:** Contributed patient information and samples. **Nancy Mendelsohn:** Contributed patient information and samples. **Ieva Micule:** Contributed patient information and samples. **John Moeschler:** Contributed patient information and samples. **Sheela Nampoothiri:** Contributed patient information and samples. **Kaoru Ohashi:** Contributed patient information and samples. **Cynthia M. Powell:** Contributed patient information and samples. **Subhadra Ramanathan:** Contributed patient information and samples. **Salmo Raskin:** Contributed patient information and samples. **Elizabeth Roeder:** Contributed patient information and samples. **Marlene Rio:** Contributed patient information and samples. **Alan F. Rope:** Contributed patient information and samples. **Karan Sangha:** Contributed patient information and samples. **Angela E. Scheuerle:** Contributed patient information and samples. **Adele Schneider:** Contributed patient information and samples. **Stavit Shalev:** Contributed patient information and samples. **Victoria Siu:** Contributed patient information and samples. **Rosemarie Smith:** Contributed patient information and samples. **Cathy Stevens:** Contributed patient information and samples. **Tinatin Tkemaladze:** Contributed patient information and samples. **John Toimie:** Contributed patient information and samples. **Helga Toriello:** Contributed patient information and samples. **Anne Turner:** Contributed patient information and samples. **Patricia G. Wheeler:** Contributed patient information and samples. **Susan M. White:** Contributed patient information and samples. **Terri Young:** Contributed patient information and samples. **Kathleen M. Loomes:** Contributed patient information and samples, review and editing of the manuscript. **Mary Pipan:** Contributed patient information and samples, review, and editing of the manuscript. **Ann Tokay Harrington:** Contributed patient information and samples. **Elaine Zackai:** Contributed patient information and samples. **Ramakrishnan Rajagopalan:** Whole genome sequencing data analysis. **Laura Conlin:** Contributed data acquisition, and data analysis tools and performed the analysis. **Matthew A. Deardorff:** Data collection, analysis and interpretation of results, contributed patient information and samples, performed experiments, reviewed and editing of manuscript. **Deborah McEldrew:** Performed experiments and data analysis. **Juan Pie:** Contributed patient information and samples. **Feliciano Ramos:** Data analysis and data sharing. **Antonio Musio:** Data sharing. **Antonie D. Kline:** Contributed patient information and samples, review and editing of manuscript. **Kosuke Izumi:** Contributed patient information and samples, performed the experiments and analyzed the data. **Sarah E. Raible:** Sample and data collection, manuscript review and editing, administration. **Ian D. Krantz:** Study conception and design, drafting and writing the manuscript, data curation, methodology, project administration, acquisition, analysis, and interpretation of data.

AFFILIATIONS

[1]Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[2]Illumina Inc, San Diego, California, USA
[3]Natera, Inc., Austin, Texas, USA
[4]Division of Otolaryngology, National Center for Child Health and Development, Tokyo, Japan
[5]Emugen Therapeutics, Woburn, Massachusetts, USA
[6]Department of Cancer Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[7]Department of Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, Philadelphia, Pennsylvania, USA
[8]Hematologics Inc, Seattle, Washington, USA
[9]Department of Pathology, Duke University School of Medicine, Durham, North Carolina, USA
[10]Al Jalila Genomics Center, Al Jalila Children's Hospital, Dubai, United Arab Emirates
[11]Center for Genomic Discovery, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates
[12]Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[13]Department of Genetic Medicine, Munroe-Meyer Institute, University of Nebraska Medical Center, Omaha, Nebraska, USA
[14]Fullerton Genetics Center, Mission Health, Asheville, North Carolina, USA
[15]Department of Pediatrics, NYU Langone Hospital-Long Island, Mineola, New York, USA
[16]Pediatrics, Columbia University Irving Medical Center, New York, New York, USA
[17]Hospital Infantil Universitario Niño Jesús & Universidad Autónoma de Madrid, Madrid, Spain
[18]CIBER Fisiopatología de la obesidad y nutrición (CIBEROBN) and IMDEA Food Institute, Madrid, Spain
[19]Department of Pediatrics, Nemours Children's Specialty Care, Jacksonville, Florida, USA
[20]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada
[21]Department of Medical Genetics, BC Women's Hospital, Vancouver, British Columbia, Canada
[22]Rabin Medical Center-Beilinson Hospital, Raphael Recanati Genetics Institute, Petach Tikva, Israel
[23]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
[24]Felsenstein Medical Research Center, Petach Tikva, Israel
[25]Western Australian Register of Developmental Anomalies and Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Western Australia, Australia

[26]Faculty of Health and Medical Sciences, Division of Pediatrics and Telethon Kids Institute, University of Western Australia, Perth, Western Australia, Australia
[27]Rare Care Centre, Perth Children's Hospital, Perth, Western Australia, Australia
[28]Department of Pediatrics, University of California San Diego, San Diego, California, USA
[29]Division of Genetics & Dysmophology, Rady Children's Hospital San Diego, San Diego, California, USA
[30]Department of Otolaryngology-Head and Neck Surgery, University of Kansas School of Medicine, Kansas City, Kansas, USA
[31]Department of Genetics, Kuala Lumpur Hospital, Kuala Lumpur, Malaysia
[32]The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada
[33]Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for SickKids, University of Toronto, Toronto, Ontario, Canada
[34]Department of Pediatrics, Division of Medical Genetics, Loma Linda University School of Medicine, Loma Linda, California, USA
[35]Division of Genetics and Genomics, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA
[36]R & D MILS International India, Mumbai, India
[37]Department of Biomolecular Medicine, Ghent University, Ghent, Belgium
[38]Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium
[39]Genetics of Learning Disability Service, Hunter Genetics, Waratah, New South Wales, Australia
[40]Division of Medical Genetics, Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, California, USA
[41]Nemours Children's Health, Wilmington, Delaware, USA
[42]University of Connecticut Health Center, Farmington, Connecticut, USA
[43]Division of Genetics, Department of Paediatrics, All India Institute of Medical Sciences, New Delhi, India
[44]Department of Pediatrics, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA
[45]Department of Pediatrics, Boston University School of Medicine, Boston, Massachusetts, USA
[46]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, and Department of Pediatrics University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
[47]Division of Dysmorphology & Teratology, Department of Pediatrics, University of California San Diego School of Medicine, San Diego, California, USA
[48]Kariminejad-Najmabadi Pathology and Genetics Center, Tehran, Iran
[49]Division of Genetics, Advocate Children's Hospital, Park Ridge, Illinois, USA
[50]Children's Clinical University Hospital, Riga, Latvia
[51]Department of Clinical Genetics, Children's Health Ireland, Dublin, Ireland
[52]Duke University Medical Center, Durham, North Carolina, USA
[53]Complex Health Solutions, United Healthcare, Minneapolis, Minnesota, USA
[54]Department of Pediatrics, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire, USA
[55]Department of Pediatric Genetics, Amrita Institute of Medical Sciences & Research Centre, Cochin, India
[56]Division of Genetics and Metabolism, Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA
[57]Genetika—Centro de aconselhamento e laboratório de genética, Curitiba, Brazil
[58]Department of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, San Antonio, Texas, USA
[59]Department of Genetics, Hôpital Necker-Enfants Malades, Paris, France
[60]Genome Medical, South San Francisco, California, USA
[61]Division of Genetics and Metabolism, Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas, USA
[62]Department of Pediatrics and Oculogenetics, Wills Eye Hospital, Philadelphia, Pennsylvania, USA
[63]Rappaport Faculty of Medicine, Technion, The Genetics Institute, Emek Medical Center, Afula, Haifa, Israel
[64]London Health Sciences Centre, London, Ontario, Canada
[65]Division of Medical Genetics, Department of Pediatrics, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[66]Division of Genetics, Department of Pediatrics, Maine Medical Center, Portland, Maine, USA
[67]Department of Pediatrics, University of Tennessee College of Medicine, T.C. Thompson Children's Hospital, Chattanooga, Tennessee, USA
[68]Department of Molecular and Medical Genetics, Tbilisi State Medical University, Tbilisi, Georgia
[69]Clinical Genetics Service, Laboratory Medicine Building, Southern General Hospital, Glasgow, UK
[70]Department of Pediatrics and Human Development, Michigan State University, East Lansing, Michigan, USA
[71]Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, New South Wales, Australia
[72]Division of Genetics, Arnold Palmer Hospital, Orlando, Florida, USA
[73]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, Australia
[74]Department of Paediatrics, University of Melbourne, Parkville, Australia
[75]Department of Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA
[76]Research to Prevent Blindness Inc, New York, New York, USA
[77]Division of Pediatric Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

[78]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[79]Behavioral Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[80]Center for Rehabilitation, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[81]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[82]Department of Pathology and Laboratory Medicine, Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA
[83]Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA
[84]Department of Pediatrics, Children's Hospital Los Angeles, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA
[85]Laboratorio de Genética Clínica y Genómica Funcional, Facultad de Medicina, Universidad de Zaragoza, Zaragoza, Spain
[86]Unidad de Genética Clínica, Servicio de Pediatría, Hospital Clínico Universitario "Lozano Blesa", Zaragoza, Spain
[87]Departamento de Pediatría, Facultad de Medicina, Universidad de Zaragoza, Zaragoza, Spain
[88]Consiglio Nazionale delle Ricerche, Pisa, Italy
[89]Greater Baltimore Medical Centre, Harvey Institute of Human Genetics, Baltimore, Maryland, USA

ACKNOWLEDGMENTS

The authors are exceptionally grateful to the individuals and families with CdLS who participated in this study, as well as to the referring physicians and colleagues including Salim Aftimos, Eric Haan, Maria Giovannucci Uzielli, Fred Gilbert, Elizabeth Loy, and others who have contributed samples and clinical information. The authors are indebted to the continued support of the National (USA) and the International Cornelia de Lange Syndrome Foundations. This work was supported by the National Institutes of Health Grants (RO1: RO1HD039323, PPG: PO1HD052860, XO1: X01 HL145697/Gabriella Miller Kids First, and grants from the National CdLS Foundation). The authors are also deeply indebted to the contributions over many years of the late Dr. Laird Jackson as well as the continued support of the endowed CdLS and Related Diagnoses Multispecialty Center and the Rare Diagnoses Program in the Roberts Individualized Medical Genetics Center (RIMGC) at CHOP.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Gabriella Miller Kids First Data Resource at https://kidsfirstdrc.org/.

ORCID

Maninder Kaur https://orcid.org/0009-0000-3826-9337
Emma Schindler https://orcid.org/0000-0002-3610-7868
Omar Abdul-Rahman https://orcid.org/0000-0002-1020-9989
Kwame Anyane-Yeboa https://orcid.org/0000-0002-4977-9719
Jesús Argente https://orcid.org/0000-0001-5826-0276
Lynne M. Bird https://orcid.org/0000-0003-4833-3747
Robin Clark https://orcid.org/0000-0002-7277-4975
John M. Graham Jr https://orcid.org/0000-0003-4297-1078
Neerja Gupta https://orcid.org/0000-0002-7454-5543
Marilyn C. Jones https://orcid.org/0000-0002-4512-2015
Ariana Kariminejad https://orcid.org/0000-0002-8467-4728
Sheela Nampoothiri https://orcid.org/0000-0002-9575-0722
Antonio Musio https://orcid.org/0000-0001-7701-6543
Antonie D. Kline https://orcid.org/0000-0002-7863-2994

REFERENCES

Ajmone, P. F., Rigamonti, C., Dall'Ara, F., Monti, F., Vizziello, P., Milani, D., Cereda, A., Selicorni, A., & Costantino, A. (2014). Communication, cognitive development and behavior in children with Cornelia de Lange Syndrome (CdLS): Preliminary results. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *165B*(3), 223–229. https://doi.org/10.1002/ajmg.b.32224

Ansari, M., Poke, G., Ferry, Q., Williamson, K., Aldridge, R., Meynert, A. M., Bengani, H., Chan, C. Y., Kayserili, H., Avci, S., Hennekam, R. C., Lampe, A. K., Redeker, E., Homfray, T., Ross, A., Falkenberg Smeland, M., Mansour, S., Parker, M. J., Cook, J. A., … FitzPatrick, D. R. (2014). Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. *Journal of Medical Genetics*, *51*(10), 659–668. https://doi.org/10.1136/jmedgenet-2014-102573

Bhuiyan, Z. A., Stewart, H., Redeker, E. J., Mannens, M. M., & Hennekam, R. C. (2007). Large genomic rearrangements in NIPBL are infrequent in Cornelia de Lange syndrome. *European Journal of Human Genetics*, *15*(4), 505–508. https://doi.org/10.1038/sj.ejhg.5201776

Boyle, M. I., Jespersgaard, C., Brondum-Nielsen, K., Bisgaard, A. M., & Tumer, Z. (2015). Cornelia de Lange syndrome. *Clinical Genetics*, *88*(1), 1–12. https://doi.org/10.1111/cge.12499

Brachmann, W. (1916). Ein fall von symmetrischer monodaktylie durch Ulnadefekt. *Jarb Kinder Phys Erzie*, *84*, 225–235.

Brown, C. J., Miller, A. P., Carrel, L., Rupert, J. L., Davies, K. E., & Willard, H. F. (1995). The DXS423E gene in Xp11.21 escapes X chromosome inactivation. *Human Molecular Genetics*, *4*(2), 251–255. https://doi.org/10.1093/hmg/4.2.251

Cheng, H., Zhang, N., & Pati, D. (2020). Cohesin subunit RAD21: From biology to disease. *Gene*, *758*, 144966. https://doi.org/10.1016/j.gene.2020.144966

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., & Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Molecular Cell*, *5*(2), 243–254. https://doi.org/10.1016/s1097-2765(00)80420-7

Cucco, F., Sarogni, P., Rossato, S., Alpa, M., Patimo, A., Latorre, A., Magnani, C., Puisac, B., Ramos, F. J., Pie, J., & Musio, A. (2020). Pathogenic variants in EP300 and ANKRD11 in patients with phenotypes overlapping Cornelia de Lange syndrome. *American Journal of Medical Genetics. Part A*, *182*(7), 1690–1696. https://doi.org/10.1002/ajmg.a.61611

De Lange, C. (1933). Surun type nouveau degeneration (typus Amestelodamensis). *Arch Med Enfants*, *36*, 713–719.

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–Rodriguez, 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. *American Journal of Human Genetics*, *80*(3), 485–494. https://doi.org/10.1086/511888

Deardorff, M. A., Wilde, J. J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Monnich, M., Yan, Y., Xu, W., Gil–Rodriguez, M. C., Clark, D., Hakonarson, H., Halbach, S., Michelis, L. D., Rampuria, A., Rossier, E., Spranger, S., Van Maldergem, L., Lynch, S. A., … Kaiser, F. J. (2012). RAD21 mutations cause a human cohesinopathy. *American Journal of Human Genetics*, *90*(6), 1014–1027. https://doi.org/10.1016/j.ajhg.2012.04.019

Gillespie, P. J., & Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in xenopus egg extracts. *Current Biology*, *14*(17), 1598–1603. https://doi.org/10.1016/j.cub.2004.07.053

Gillis, L. A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A. D., Li, H. 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

Gil-Rodriguez, M. C., Deardorff, M. A., Ansari, M., Tan, C. A., Parenti, I., Baquero-Montoya, C., Ousager, L. B., Puisac, B., Hernandez–Marcos, M., Teresa–Rodrigo, M. E., Marcos–Alcalde, I., Wesselink, J. J., Lusa–Bernal, S., Bijlsma, E. K., Braunholz, D., Bueno–Martinez, I., Clark, D., Cooper, N. S., Curry, C. J., … Pie, J. (2015). De novo heterozygous mutations in SMC3 cause a range of Cornelia de Lange syndrome-overlapping phenotypes. *Human Mutation*, *36*(4), 454–462. https://doi.org/10.1002/humu.22761

Gonzalez Garcia, A., Malone, J., & Li, H. (2020). A novel mosaic variant on SMC1A reported in buccal mucosa cells, albeit not in blood, of a patient with Cornelia de Lange-like presentation. *Cold Spring Harbor Molecular Case Studies*, *6*(3), a005322. https://doi.org/10.1101/mcs.a005322

Grados, M. A., Alvi, M. H., & Srivastava, S. (2017). Behavioral and psychiatric manifestations in Cornelia de Lange syndrome. *Current Opinion in Psychiatry*, *30*(2), 92–96. https://doi.org/10.1097/YCO.0000000000000311

Haering, C. H., Farcas, A. M., Arumugam, P., Metson, J., & Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. *Nature*, *454*(7202), 297–301. https://doi.org/10.1038/nature07098

Huisman, S., Mulder, P. A., Redeker, E., Bader, I., Bisgaard, A. M., Brooks, A., Cereda, A., Cinca, C., Clark, D., Cormier–Daire, V., Deardorff, M. A., Diderich, K., Elting, M., van Essen, A., FitzPatrick, D., Gervasini, C., Gillessen–Kaesbach, G., Girisha, K. M., Hilhorst–Hofstee, Y., … Hennekam, R. C. (2017). Phenotypes and genotypes in individuals with SMC1A variants. *American Journal of Medical Genetics. Part A*, *173*(8), 2108–2125. https://doi.org/10.1002/ajmg.a.38279

Ireland, M., English, C., Cross, I., Houlsby, W. T., & Burn, J. (1991). A de novo translocation t(3;17)(q26.3;q23.1) in a child with Cornelia de Lange syndrome. *Journal of Medical Genetics*, *28*(9), 639–640. https://doi.org/10.1136/jmg.28.9.639

Izumi, K. (2016). Disorders of transcriptional regulation: An emerging category of multiple malformation syndromes. *Molecular Syndromology*, *7*(5), 262–273. https://doi.org/10.1159/000448747

Jackson, L., Kline, A. D., Barr, M. A., & Koch, S. (1993). de Lange syndrome: A clinical review of 310 individuals. *American Journal of Medical Genetics*, *47*(7), 940–946. https://doi.org/10.1002/ajmg.1320470703

Kaur, M., Mehta, D., Noon, S. E., Deardorff, M. A., Zhang, Z., & Krantz, I. D. (2016). NIPBL expression levels in CdLS probands as a predictor of mutation type and phenotypic severity. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, *172*(2), 163–170. https://doi.org/10.1002/ajmg.c.31495

Kline, A. D., Krantz, I. D., Sommer, A., Kliewer, M., Jackson, L. G., FitzPatrick, D. R., Levin, A. V., & Selicorni, A. (2007). Cornelia de Lange syndrome: Clinical review, diagnostic and scoring systems, and anticipatory guidance. *American Journal of Medical Genetics. Part A*, *143A*(12), 1287–1296. https://doi.org/10.1002/ajmg.a.31757

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., 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

Krawczynska, N., Wierzba, J., & Wasag, B. (2019). Genetic mosaicism in a group of patients with Cornelia de Lange syndrome. *Frontiers in Pediatrics*, *7*, 203. https://doi.org/10.3389/fped.2019.00203

Kuzniacka, A., Wierzba, J., Ratajska, M., Lipska, B. S., Koczkowska, M., Malinowska, M., & Limon, J. (2013). Spectrum of NIPBL gene mutations in polish patients with Cornelia de Lange syndrome. *Journal of Applied Genetics*, *54*(1), 27–33. https://doi.org/10.1007/s13353-012-0126-9

Latorre-Pellicer, A., Gil-Salvador, M., Parenti, I., Lucia-Campos, C., Trujillano, L., Marcos-Alcalde, I., Arnedo, M., Ascaso, A., Ayerza–Casas, A., Antonanzas–Perez, R., Gervasini, C., Piccione, M., Mariani, M., Weber, A., Kanber, D., Kuechler, A., Munteanu, M., Khuller, K., Bueno–Lozano, G., … Pie, J. (2021). Clinical relevance of postzygotic mosaicism in Cornelia de Lange syndrome and purifying selection of NIPBL variants in blood. *Scientific Reports*, *11*(1), 15459. https://doi.org/10.1038/s41598-021-94958-z

Levin, A. V., Seidman, D. J., Nelson, L. B., & Jackson, L. G. (1990). Ophthalmologic findings in the Cornelia de Lange syndrome. *Journal of Pediatric Ophthalmology and Strabismus*, *27*(2), 94–102. https://doi.org/10.3928/0191-3913-19900301-11

Liu, J., Feldman, R., Zhang, Z., Deardorff, M. A., Haverfield, E. V., Kaur, M., Li, J. R., Clark, D., Kline, A. D., Waggoner, D. J., Das, S., Jackson, L. G., & Krantz, I. D. (2009). SMC1A expression and mechanism of pathogenicity in probands with X-linked Cornelia de Lange syndrome. *Human Mutation*, *30*(11), 1535–1542. https://doi.org/10.1002/humu.21095

Mannini, L., Cucco, F., Quarantotti, V., Krantz, I. D., & Musio, A. (2013). Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. *Human Mutation*, *34*(12), 1589–1596. https://doi.org/10.1002/humu.22430

Mannini, L., Liu, J., Krantz, I. D., & Musio, A. (2010). Spectrum and consequences of SMC1A mutations: The unexpected involvement of a core component of cohesin in human disease. *Human Mutation*, *31*(1), 5–10. https://doi.org/10.1002/humu.21129

Marino, T., Wheeler, P. G., Simpson, L. L., Craigo, S. D., & Bianchi, D. W. (2002). Fetal diaphragmatic hernia and upper limb anomalies suggest Brachmann-de Lange syndrome. *Prenatal Diagnosis*, *22*(2), 144–147. https://doi.org/10.1002/pd.281

Mehta, D., Vergano, S. A., Deardorff, M., Aggarwal, S., Barot, A., Johnson, D. M., Miller, N. F., Noon, S. E., Kaur, M., Jackson, L., & Krantz, I. D. (2016). Characterization of limb differences in children with Cornelia de Lange syndrome. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, *172*(2), 155–162. https://doi.org/10.1002/ajmg.c.31498

Musio, A. (2020). The multiple facets of the SMC1A gene. *Gene*, *743*, 144612. https://doi.org/10.1016/j.gene.2020.144612

Musio, A., Selicorni, A., Focarelli, M. L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., & Larizza, L. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nature Genetics*, *38*(5), 528–530. https://doi.org/10.1038/ng1779

Niu, D. M., Huang, J. Y., Li, H. Y., Liu, K. M., Wang, S. T., Chen, Y. J., Udaka, T., Izumi, K., & Kosaki, K. (2006). Paternal gonadal mosaicism of NIPBL mutation in a father of siblings with Cornelia de Lange syndrome. *Prenatal Diagnosis*, *26*(11), 1054–1057. https://doi.org/10.1002/pd.1554

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., Deciphering Developmental Disorders, S., 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

Oostra, R. J., Baljet, B., & Hennekam, R. C. (1994). Brachmann-de Lange syndrome "avant la lettre". *American Journal of Medical Genetics*, *52*(3), 267–268. https://doi.org/10.1002/ajmg.1320520303

Parenti, I., Gervasini, C., Pozojevic, J., Graul-Neumann, L., Azzollini, J., Braunholz, D., Watrin, E., Wendt, K. S., Cereda, A., Cittaro, D., Gillessen–Kaesbach, G., Lazarevic, D., Mariani, M., Russo, S., Werner, R., Krawitz, P., Larizza, L., Selicorni, A., & Kaiser, F. J. (2016). Broadening of cohesinopathies: Exome sequencing identifies mutations in ANKRD11 in two patients with Cornelia de Lange-overlapping phenotype. *Clinical Genetics*, *89*(1), 74–81. https://doi.org/10.1111/cge.12564

Parenti, I., Teresa-Rodrigo, M. E., Pozojevic, J., Ruiz Gil, S., Bader, I., Braunholz, D., Bramswig, N. C., Gervasini, C., Larizza, L., Pfeiffer, L., Ozkinay, F., Ramos, F., Reiz, B., Rittinger, O., Strom, T. M., Watrin, E., Wendt, K., Wieczorek, D., Wollnik, B., … Kaiser, F. J. (2017). Mutations in chromatin regulators functionally link Cornelia de Lange syndrome and clinically overlapping phenotypes. *Human Genetics*, *136*(3), 307–320. https://doi.org/10.1007/s00439-017-1758-y

Pehlivan, D., Hullings, M., Carvalho, C. M., Gonzaga-Jauregui, C. G., Loy, E., Jackson, L. G., Krantz, I. D., Deardorff, M. A., & Lupski, J. R. (2012). NIPBL rearrangements in Cornelia de Lange syndrome: Evidence for replicative mechanism and genotype-phenotype correlation. *Genetics in Medicine*, *14*(3), 313–322. https://doi.org/10.1038/gim.2011.13

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

Rollins, R. A., Korom, M., Aulner, N., Martens, A., & Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. *Molecular and Cellular Biology*, *24*(8), 3100–3111. https://doi.org/10.1128/MCB.24.8.3100-3111.2004

Russell, K. L., Ming, J. E., Patel, K., Jukofsky, L., Magnusson, M., & Krantz, I. D. (2001). Dominant paternal transmission of Cornelia de Lange syndrome: A new case and review of 25 previously reported familial recurrences. *American Journal of Medical Genetics*, *104*(4), 267–276. https://doi.org/10.1002/ajmg.10066

Russo, S., Masciadri, M., Gervasini, C., Azzollini, J., Cereda, A., Zampino, G., Haas, O., Scarano, G., Di Rocco, M., Finelli, P., Tenconi, R., Selicorni, A., & Larizza, L. (2012). Intragenic and large NIPBL rearrangements revealed by MLPA in Cornelia de Lange patients. *European Journal of Human Genetics*, *20*(7), 734–741. https://doi.org/10.1038/ejhg.2012.7

Sarogni, P., Pallotta, M. M., & Musio, A. (2020). Cornelia de Lange syndrome: From molecular diagnosis to therapeutic approach. *Journal of Medical Genetics*, *57*(5), 289–295. https://doi.org/10.1136/jmedgenet-2019-106277

Sataloff, R. T., Spiegel, J. R., Hawkshaw, M., Epstein, J. M., & Jackson, L. (1990). Cornelia de Lange syndrome. Otolaryngologic manifestations. *Archives of Otolaryngology – Head & Neck Surgery*, *116*(9), 1044–1046. https://doi.org/10.1001/archotol.1990.01870090060008

Shaikh, T. H., Gai, X., Perin, J. C., Glessner, J. T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L. K., D'Arcy, M., Frackelton, E. C., Geiger, E. A., Haldeman–Englert, C., Imielinski, M., Kim, C. E., Medne, L., Annaiah, K., Bradfield, J. P., Dabaghyan, E., … Hakonarson, H. (2009). High-resolution mapping and analysis of copy number variations in the human genome: A data resource for clinical and research applications. *Genome Research*, *19*(9), 1682–1690. https://doi.org/10.1101/gr.083501.108

Slavin, T. P., Lazebnik, N., Clark, D. M., Vengoechea, J., Cohen, L., Kaur, M., Konczal, L., Crowe, C. A., Corteville, J. E., Nowaczyk, M. J., Byrne, J. L., Jackson, L. G., & Krantz, I. D. (2012). Germline mosaicism in Cornelia de Lange syndrome. *American Journal of Medical Genetics. Part A*, *158A*(6), 1481–1485. https://doi.org/10.1002/ajmg.a.35381

Takahashi, T. S., Yiu, P., Chou, M. F., Gygi, S., & Walter, J. C. (2004). Recruitment of xenopus Scc2 and cohesin to chromatin requires the pre-replication complex. *Nature Cell Biology*, *6*(10), 991–996. https://doi.org/10.1038/ncb1177

Teresa-Rodrigo, M. E., Eckhold, J., Puisac, B., Pozojevic, J., Parenti, I., Baquero-Montoya, C., Gil–Rodriguez, M. C., Braunholz, D., Dalski, A., Hernandez–Marcos, M., Ayerza, A., Bernal, M. L., Ramos, F. J., Wieczorek, D., Gillessen–Kaesbach, G., Pie, J., & Kaiser, F. J. (2016). Identification and functional characterization of two intronic NIPBL mutations in two patients with Cornelia de Lange syndrome. *BioMed Research International*, *2016*, 8742939. https://doi.org/10.1155/2016/8742939

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

Vrolik, W. (1849). *Tabulae ad illustrandam embryogenesin hominis et mammalium, tam naturalem quam abnormem*. G. M. P. Londonck.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., Hakonarson, H., & Bucan, M. (2007). PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research*, *17*(11), 1665–1674. https://doi.org/10.1101/gr.6861907

Yuan, B., Pehlivan, D., Karaca, E., Patel, N., Charng, W. L., Gambin, T., Gonzaga–Jauregui, C., Sutton, V. R., Yesil, G., Bozdogan, S. T., Tos, T., Koparir, A., Koparir, E., Beck, C. R., Gu, S., Aslan, H., Yuregir, O. O., Al Rubeaan, K., Alnaqeb, D., … Lupski, J. R. (2015). Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. *The Journal of Clinical Investigation*, *125*(2), 636–651. https://doi.org/10.1172/JCI77435

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: Kaur, M., Blair, J., Devkota, B., Fortunato, S., Clark, D., Lawrence, A., Kim, J., Do, W., Semeo, B., Katz, O., Mehta, D., Yamamoto, N., Schindler, E., Al Rawi, Z., Wallace, N., Wilde, J. J., McCallum, J., Liu, J., Xu, D., … Krantz, I. D. (2023). Genomic analyses in Cornelia de Lange Syndrome and related diagnoses: Novel candidate genes, genotype–phenotype correlations and common mechanisms. *American Journal of Medical Genetics Part A*, 1–19. https://doi.org/10.1002/ajmg.a.63247