Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies

Bo Yuan, PhD1,2, Juanita Neira, MD1,3, Davut Pehlivan, MD1,4, Teresa Santiago-Sim, PhD1,2, Xiaofei Song, BS1, Jill Rosenfeld, MS1,2, Jennifer E. Posey, MD, PhD1, Vipulkumar Patel, MS2, Weihong Jin, MS2, Margaret P. Adam, MD5,6, Emma L. Baple, MBBS, PhD7,8, John Dean, FRCP9, Chin-To Fong, MD10, Scott E. Hickey, MD11, Louanne Hudgins, MD12, Eyby Leon, MD13, Suneeta Madan-Khetarpal, MD14, Lettie Rawlins, MBChB7,8, Cecilie F. Rustad, MD15, Asbjørg Stray-Pedersen, MD, PhD16, Kristian Tveten, PhD17, Olivia Wenger, MD18, Jullianne Diaz, MS13, Laura Jenkins, MS14, Laura Martin, MS10, Marianne McGuire, MS1,2, Marguerite Pietryga, MS1,1, Linda Ramsdell, MS5,6, Leah Slattery, MS1,1, DDD Study, Farida Abid, MD1,3,4, Alison A. Bertuch, MD, PhD1,3, Dorothy Grange, MD19, LaDonna Immken, MD20, Christian P. Schaaf, MD, PhD1,2,2,21,22,23, Hilde Van Esch, MD, PhD24, Weimin Bi, PhD1,2, Sau Wai Cheung, PhD, MBA1,2, Amy M. Breman, PhD1,2, Janice L. Smith, PhD1,2, Chad Shaw, PhD1,2, Andrew H. Crosby, PhD7, Christine Eng, MD1,2, Yaping Yang, PhD1,2, James R. Lupski, MD, PhD1,3,25, Rui Xiao, PhD1,2 and Pengfei Liu, PhD1,2

**Purpose:** Defects in the cohesin pathway are associated with cohesinopathies, notably Cornelia de Lange syndrome (CdLS). We aimed to delineate pathogenic variants in known and candidate cohesinopathy genes from a clinical exome perspective.

**Methods:** We retrospectively studied patients referred for clinical exome sequencing (CES, N = 10,698). Patients with causative variants in novel or recently described cohesinopathy genes were enrolled for phenotypic characterization.

**Results:** Pathogenic or likely pathogenic single-nucleotide and insertion/deletion variants (SNVs/indels) were identified in established disease genes including *NIPBL* (N = 5), *SMC1A* (N = 14), *SMC3* (N = 4), *RAD21* (N = 2), and *HDAC8* (N = 8). The phenotypes in this genetically defined cohort skewed towards the mild end of the CdLS spectrum as compared with phenotype-driven cohorts. Candidate or recently reported cohesinopathy genes were supported by de novo SNVs/indels in *STAG1* (N = 3), *STAG2* (N = 5), *PDSSA* (N = 1), and *WAPL* (N = 1), and one inherited SNV in *PDSSA*. We also identified copy-number deletions affecting *STAG1* (two de novo, one of unknown inheritance) and *STAG2* (one of unknown inheritance). Patients with *STAG1* and *STAG2* variants presented with overlapping features yet without characteristic facial features of CdLS.

**Conclusion:** CES effectively identified disease-causing alleles at the mild end of the cohesinopathy spectrum and enabled characterization of candidate disease genes.

**Keywords:** Atypical cohesinopathies; Clinical exome sequencing (CES); Cohesin pathway; *STAG1*; *STAG2*

**INTRODUCTION**

The cohesin complex mediates sister chromatid cohesion and ensures accurate chromosome segregation, recombination-mediated DNA repair, and genomic stability during DNA replication and cell division. Accumulating evidence suggests that cohesin is also involved in regulating chromosomal loop/architecture and gene transcriptional regulation.1–3

Cohesin is a multisubunit protein complex composed of evolutionarily conserved core components encoded by *SMC1A* (MIM *300040*), *SMC3* (MIM *606062*), *RAD21* (MIM *606462*) and either *STAG1* (MIM *604358*) or *STAG2* (MIM *300826*) depending on the chromosomal location. Direct interaction between *SMC1A*, *SMC3*, and *RAD21* forms a tripartite ring structure that is used to entrap the replicated chromatid during sister chromatid cohesion (Fig. 1a). *STAG1/2* are the core structural component of functional cohesin and critical for the loading of cohesin onto chromatin during mitosis.1,2

---

© American College of Medical Genetics and Genomics

ARTICLE

GENETICS in MEDICINE | Volume 21 | Number 3 | March 2019

663
ARTICLE

In addition to the aforementioned structural components, cohesin also interacts with the regulatory factors of the cohesion cycle, including proteins encoded by NIPBL (MIM *608667), MAU2 (MIM *614560), PDSSA (MIM *613200) or PDSSB (MIM *605333), WAPL (MIM *610754), HDAC8 (MIM *300269), ESCO1 (MIM *609674), and ESCO2 (MIM *609353), to facilitate cohesion dynamics and function on chromatin (Fig. 1a).1,2

Precise orchestration of cohesin’s structural components and regulatory factors ensures faithful progression of the cohesion cycle (Fig. 1a). Defects of the structural or regulatory components of cohesin lead to various multisystem malformation syndromes described as “cohesinopathies”, a collection of syndromes with shared clinical findings such as distinctive facial features, growth retardation, developmental delay/intellectual disability (DD/ID), and limb abnormalities. Clinically, the most distinguishable type of cohesinopathy is the classic Cornelia de Lange syndrome (CdLS, MIM #122470), with the majority of cases explained by single-nucleotide and insertion/deletion variants (SNVs/indels) and exonic copy-number variants (CNVs) resulting in loss-of-function (LoF) alleles in NIPBL.4–6 The traditional phenotype-driven studies that included the mild end of the CdLS spectrum led to the discovery of SMC1A, SMC3, RAD21, and HDAC8 (MIM #300590, #610759, #614701, and #300882) as new cohesinopathy genes.4–11 The resultant CdLS phenotype is largely dependent on the genes being affected and pathogenic variant (PV) types.12 Although mild forms of CdLS present with less striking phenotypes and are more clinically challenging to recognize in comparison with the classic form, they have been found in an increasing number of patients with cohesinopathies.

Here, we used a genotype-driven approach to investigate the allelic series of genes encoding cohesin components based on a large cohort of patients (N = 10,698) with a variety of unselected clinical presentations who were referred for clinical exome sequencing (CES). We identified pathogenic or likely pathogenic variants in known CdLS genes (NIPBL, SMC1A, SMC3, RAD21, and HDAC8) in patients mostly without a clinical diagnosis of CdLS, representing a cohort on the mild end of the clinical presentation of cohesinopathies. By applying the same genotype-first approach in the CES cohort, we further established STAG1 and STAG2 as new cohesinopathy genes with variants that act by a putative LoF mechanism, corroborating recent reports of patients with developmental disorders carrying PV in these two genes.13–15 Additional studies of patients who had chromosome microarray analyses (CMA, N = 63,127) also identified deletion CNVs affecting STAG1 and STAG2, which further supports the human disease association of these two genes via a LoF
Table 1  Summary of variants in the known Cornelia de Lange syndrome (CdLS) genes identified by Baylor Genetics clinical exome sequencing

| Gene (transcript) | Genomic coordinates (hg19) | Exon /intron | Coding sequence change | Protein change | Zygosity | Inheritance | Novelty | Classification | CdLS as a differential diagnosis? | Dual molecular diagnosis? |
|-------------------|---------------------------|--------------|------------------------|----------------|----------|-------------|---------|----------------|--------------------------------|-------------------------|
| NIPBL (NM_133433.3) | Chr5: 36985760 exon10 | c.2479_2480del | p.R827Gfs*2 | Het | De novo | Reported | P | No (prenatal) | No |
| | Chr5: 37017173 exon24 | c.4829T>C | p.L1610P | Het | De novo | This cohort | LP | No |
| | Chr5: 37046239 exon38 | c.652T>C | p.L2176P | Het | De novo | This cohort | LP | Yes | No |
| | Chr5: 37052580 exon42 | c.7175G>A | p.C2392Y | Het | De novo | This cohort | LP | No | No |
| | Chr5: 36962223 intron5 | c.459-2A>G | Splicing | Het | De novo | This cohort | P | Yes, among others, POC | No |
| SMC1A (NM_006306.2) | ChrX: 53442118 exon2 | c.110T>C | p.G37V | Het | De novo | This cohort | LP | No | No |
| | ChrX: 53442112 exon2 | c.116C>G | p.S39* | Het | De novo | This cohort | P | Yes | MYH2 heterozygous c.1160C>T (p. A387V), de novo |
| | ChrX: 53442100 exon2 | c.128A>T | p.D43V | Het | De novo | This cohort | LP | No | No |
| | ChrX: 53442088 exon2 | c.140T>G | p.F47C | Het | De novo | This cohort | LP | No | CFTR homozygous c.1521_1523del (p. F508del) |
| | ChrX: 53441930 exon2 | c.298G>C | p.G100R | Het | De novo | This cohort | LP | No | EFHC1 heterozygous c.1612C>T (p.R538*), paternally inherited |
| | ChrX: 53440211 exon4 | c.586C>T | p.R196C | Hem | De novo | This cohort | LP | No | DMD hemizygous deletion exons 49-51, maternally inherited |
| | ChrX: 53440048 exon5 | c.655del | p.A219Lfs*45 | Het | De novo | This cohort | P | No | No |
| | ChrX: 53439899 exon5 | c.802_804del | p.K268del | Het | De novo | Reported | P | No | No |
| | ChrX: 53436051 exon9 | c.1487G>A | p.R496H | Het | De novo | Reported | P | No | No |
| | ChrX: 53430523 exon15 | c.2394dup | p.R799Tfs*4 | Het | De novo | This cohort | P | No | No |
| | ChrX: 53430498 exon15 | c.2420G>A | p.R807H | Het | De novo | This cohort | LP | No | No |
| | ChrX: 53426525 exon16 | c.2547del | p.I849Mfs*12 | Het | De novo | This cohort | P | No | No |
| | ChrX: 53423152 exon18 | c.2853_2856del | p.S951Rfs*12 | Het | De novo | Reported | P | No | No |
| | ChrX: 53438853 intron7 | c.1114-2A>G | Splicing | Hem | Maternal | This cohort | P | No | No |
| Gene (transcript) | Genomic coordinates (hg19) | Exon/intron | Coding sequence change | Protein change | Zygosity | Inheritance | Novelty | Classification | CdLS as a differential diagnosis? | Dual molecular diagnosis? |
|------------------|-----------------------------|-------------|------------------------|----------------|----------|-------------|---------|---------------|-------------------------------|-------------------------|
| SMC3 (NM_005445.3) | Chr10: 112341720 | exon9 | c.587T>C | p.I196T | Het | De novo | This cohort | LP | No | No |
|                  | Chr10: 112349688 | exon15 | c.1453_1455del | p.A485del | Het | De novo | This cohort | LP | No | No |
|                  | Chr10: 112356303 | exon19 | c.2111T>C | p.I704T | Het | De novo | This cohort | LP | No | No |
|                  | Chr10: 112362647 | exon27 | c.3362C>T | p.S1121F | Het | De novo | This cohort | LP | Yes | CREBBP heterozygous c.6137C>T (p. A2046V), de novo |
| RAD21 (NM_006265.2) | Chr8: 117862926 | exon12 | c.1550dupC | p.E518fs | Het | Paternal<sup>a</sup> | This cohort | P | No | No |
|                  | Chr8: 117866483 | intron10 | c.1161+1G>A | Splicing | Het | Maternal<sup>a</sup> | This cohort | P | No | No |
| HDAC8 (NM_018486.2) | ChrX: 71787758 | exon4 | c.418G>A | p.G140R | Het | De novo | This cohort | LP | No | No |
|                  | ChrX: 71715066 | exon5 | c.490C>T | p.R164*<sup>b</sup> | Het | De novo | Reported<sup>7</sup> | P | No | No |
|                  | ChrX: 71715066 | exon5 | c.490C>T | p.R164*<sup>b</sup> | Het | De novo | Reported<sup>7</sup> | P | No | No |
|                  | ChrX: 71715029 | exon5 | c.527A>G | p.D176G | Het | De novo | This cohort | LP | No | IRX5 compound heterozygous c.1362_1368delinsGT (p.K455fs) and c.240_242delCTC (p.S81del) |
|                  | ChrX: 71710823 | exon6 | c.584T>A | p.V195D | Het | De novo | This cohort | LP | No | No |
|                  | ChrX: 71684526 | exon8 | c.793G>A | p.G265R | Het | De novo | This cohort | LP | No | No |
|                  | ChrX: 71681927 | exon9 | c.932C>T | p.T311M | Het | De novo | Reported<sup>7</sup> | P | No | No |
|                  | ChrX: 71681922 | exon9 | c.937C>T | p.R313*<sup>b</sup> | Het | Not maternal | This cohort | P | No | No |

<sup>P</sup> pathogenic, <sup>LP</sup> likely pathogenic, <sup>POC</sup> product of conception

<sup>a</sup>Inherited variants from mildly affected parents, who were confirmed to be nonmosaic by Sanger sequencing (data not shown)

<sup>b</sup>Identical pathogenic variants in unrelated patients
mechanism. We also provide evidence supporting the candidacy of \textit{PDS5A} and \textit{WAPL} as cohesinopathy disease genes. Our findings emphasize the utility of CES to provide molecular diagnoses for disorders with extensive genetic and phenotypic heterogeneity, uncover the potential molecular etiologies of previously undiagnosed patients, and elucidate novel candidate cohesinopathy disease genes that potentially expand the genotype/phenotype characterizations of cohesinopathies.

\section*{MATERIALS AND METHODS}

\subsection*{Samples}
The study has been conducted through a collaborative effort between Baylor Genetics (BG) and Baylor-Hopkins Center for Mendelian Genomics (BHCMG), and has been approved by the Institutional Review Board of Baylor College of Medicine. Approved consents for publishing photos have been obtained. Please see Supplemental Appendix for detailed descriptions of samples in BG and BHCMG. Selected patients with \textit{STAG1}, \textit{STAG2}, or \textit{PDS5A} variants were enrolled after obtaining informed consent for further phenotypic characterization based on clinical notes submitted along with the CES order.

\subsection*{CES and variant interpretation}
CES was performed as previously described.\textsuperscript{16,17} The variant classification and interpretation were conducted by a clinical standard based on the American College of Medical Genetics and Genomics variant interpretation guidelines.\textsuperscript{18} Details of the CES experimental procedures and sample-wise quality control (QC) metrics can be found in Table S1. The possibility of mosaic variants in known CdLS genes\textsuperscript{19} was carefully evaluated. A variant is considered mosaic only if the variant read versus total read ratio is below 30\% and confirmatory Sanger sequencing demonstrates a comparable mosaic fraction.

The variants identified in this study have been deposited to ClinVar (accession numbers SCV000747051-SCV000747088 and SCV000747090-SCV000747093).

\subsection*{Chromosome microarray analysis}
The experimental design and data analysis of chromosome microarray analysis (CMA) were performed according to previously described procedures.\textsuperscript{20}

\subsection*{X-chromosome inactivation assay}
X-chromosome inactivation (XCI) studies were performed for the patient samples with \textit{STAG2} variants based on the protocol described by Allen et al.\textsuperscript{21} with modifications. Please see Supplemental Appendix for detailed protocols.

\subsection*{Estimation of pathogenic variant prevalence in somatic cancer samples}
The datasets from the COSMIC (http://cancer.sanger.ac.uk/cosmic/download) and ExAC (Exome Aggregation Consortium, http://exac.broadinstitute.org)\textsuperscript{22} databases were used for the calculation. The normalized PV abundance per gene in cancer samples is determined by the ratio between the PV frequencies of COSMIC versus the ExAC (y-axis in Fig. 1c). Please see Supplemental Appendix for details.

\section*{RESULTS}

\subsection*{Variants of established CdLS genes in the CES cohort}
Based on a genotype-driven selection approach, we identified 33 patients with pathogenic or likely pathogenic variants in the well-recognized CdLS genes from the CES cohort. Those variants include heterozygous or hemizygous SNVs/indels in \textit{NIPBL} (N = 5), \textit{SMC1A} (N = 14, X-linked), \textit{SMC3} (N = 4), \textit{RAD21} (N = 2), and \textit{HDAC8} (N = 8, X-linked) (Table 1). Genic variant distribution was calculated to show the per-gene contribution to molecular diagnosis among the five known CdLS genes (Fig. 1b). Of the 33 variants, 29 occurred de novo in the proband, 3 were inherited from a parent, and 1 was of unknown inheritance (not maternally inherited, paternal sample not available, Table 1). Among the inherited variants, one variant in \textit{SMC1A} was inherited from a symptomatic mother with a milder phenotype, demonstrating variable clinical presentation for X-linked dominant disorders; two variants in \textit{RAD21} were inherited from symptomatic parents with milder phenotypes, documenting variable expressivity of defects in \textit{RAD21}.

The CdLS patients in this cohort may be enriched for atypical or mild CdLS phenotypes, because those with classic CdLS presentation are more likely to be referred for specific single-gene or panel testing instead of CES. We retrospectively examined the clinical notes submitted by the referral clinicians for their differential diagnoses prior to CES. CdLS was not included in the initial differential diagnoses for 60\% of patients with a positive \textit{NIPBL} finding, 93\% with \textit{SMC1A}, and 75\% with \textit{SMC3} variants, and all those with \textit{RAD21} or \textit{HDAC8} variants (Table 1, Fig. 1b). These observations support the previous hypotheses that pathogenic variants in \textit{NIPBL} have a better correlation with classic CdLS, while \textit{SMC1A} and \textit{SMC3} pathogenic variants may contribute to milder CdLS features; the phenotypes caused by pathogenic variants in \textit{RAD21} and \textit{HDAC8} become more variable and sometimes present atypical CdLS features.\textsuperscript{12}

As a comparison with the genic distribution of our CES cohort, we analyzed the data from a phenotype-driven cohort of CdLS patients.\textsuperscript{19} Moreover, we re-examined the genic variant distribution on an independent phenotype-driven CdLS cohort (N = 41) from BHCMG, in which pathogenic or likely pathogenic variants in \textit{NIPBL} (N = 12), \textit{SMC1A} (N = 6), \textit{SMC3} (N = 2), and \textit{HDAC8} (N = 1) were identified (Table S2). The genic variant distribution of the BHCMG CdLS cohort is overall comparable with that calculated from the phenotype-driven cohort.\textsuperscript{13} However, both of these largely deviated from our CES cohort (Fig. 1b). The proportion of patients with \textit{NIPBL} pathogenic variants in our cohort was significantly lower in comparison with the aforementioned two phenotype-driven cohorts (chi-squared test, both with p < 0.001). The proportion of patients with \textit{SMC1A} pathogenic variants in our cohort and the BHCMG were significantly
higher than the other CdLS cohorts (chi-squared test, both with $p < 0.02$), indicating mild/atypical CdLS presentations in the BHCMG cohort. Therefore, the mutational spectrum in known CdLS genes in the CES cohort represent a distinct distortion and alternative perspective from phenotype-driven CdLS cohorts, where patients tend to present with classic phenotypes. 

Interestingly, 6/33 (18%) of the patients with positive findings from known CdLS genes carry a secondary diagnosis (Table 1), which is higher than the average observed fraction of patients with dual diagnoses from positive cases in the entire CES cohort (5%) (ref. 23). This is not unexpected because the predicted extent of multilocus diagnosis can be as high as 14% under a Poisson distribution model. 

The high representation of dual diagnosis and resultant blended phenotypes observed in this study may contribute to the complexity of the patients' phenotypes, further obscuring the underlying molecular causes, making clinical diagnosis challenging without the assistance from objective molecular testing.

### Candidate disease genes in the cohesin structural and regulatory components

STAG1, STAG2, PDS5A, PDS5B, WAPL, and MAU2 encode close interacting factors of NIPBL, SMC3, SMC1A, RAD21, and HDAC8 in the cohesin pathway, and thus may potentially supplement the locus heterogeneity of cohesinopathies. According to the ExAC database, NIPBL, SMC3, SMC1A, and RAD21 have probability of LoF intolerance (pLI) scores of 1.00, while HDAC8 has a pLI of 0.92. Similarly, STAG1, STAG2, PDS5A, PDS5B, WAPL, and MAU2 all have pLI scores of 1.00, suggesting their intolerance to LoF variants (Table S3). In our CES cohort, we identified putative LoF (truncating/splicing) or de novo missense variants in STAG1 (3), STAG2 (2), PDS5A (2), and WAPL (1). Through collaboration with the Deciphering Developmental Disorder (DDD) study and BHCMG, three additional de novo variants in STAG2 were identified.

De novo heterozygous SNVs/indels in STAG1 (NM_005862.2), including one frameshift variant (c.2009_2012del [p.N670Ifs*25]) and one missense variant (c.1129C>T [p.R377C]), were identified in patients 1 and 2, respectively (Fig. 2a). Both patients had common clinical findings that included DD/ID, hypotonia, seizures, mild dysmorphic features, and skeletal abnormalities (Table 2, Table S4). In addition, one heterozygous de novo SNV (c.654+5G>C) in STAG1 was identified in patient 3 (Fig. 2a) along with a heterozygous de novo c.1720-2A>G SNV (observed twice in ExAC including one potentially being mosaic) in ASXL1 (Bohring–Opitz syndrome; MIM #605039). Patient 3 presented with global developmental delay, dysmorphic facial features, seizures, optic atrophy, mild hypotonia, skin hypopigmentation, hirsutism, possible autism spectrum disorder, and structural brain abnormalities (Table 2, Table S4). The concurrent de novo variants in STAG1 and ASXL1 could possibly contribute to a dual molecular diagnosis of this patient.

De novo heterozygous/hemizygous SNVs/indels in STAG2 (X-linked, NM_006603.4), including two stopgain variants, two missense variants, and one frameshift variant, were identified in four females (patients 7–10; patient 7, c.418C>T [p.Q140*]; patient 8, c.1605T>A [p.C535*]; patient 9, c.1811G>A [p.R604Q]; patient 10, c.1658_1660delinsT[p.K553Ifs*6]); and one male (patient 11 [hemizygous], c.476A>G [p.Y159C]) (Fig. 2b). These patients shared common clinical findings of DD/ID, hypotonia, microcephaly, dysmorphic features, and skeletal abnormalities (Table 2, Table S4). Skewed X-inactivation (XCI) was observed in patient 8, whereas XCI was noninformative for patient 7 due to homozygosity of the marker being used for the XCI study (data not shown). In our study, truncating variants were identified in 3/4 female patients, but not in males. Although this observation is based on a limited number of patients, it is consistent with the hypothesis that truncating variants of X-linked genes may impose more severe pathogenic effect on males than females.

One heterozygous SNV, c.2275G>T (p.E759*), in PDS5A (NM_001100399.1) was identified in patient 13 with severe developmental delay, marked hypotonia, failure to thrive, dysmorphic features, hyperextensible knees, eye anomalies, and skeletal abnormalities (Table 2, Table S4). Interestingly, this patient also had a concurrent heterozygous de novo SNV, c.3325A>T (p.K1109*), in ASXL3 (Bainbridge–Ropers syndrome, MIM #615485), which presumably explains the major phenotypes. This PDS5A variant is predicted to introduce a premature stop codon in PDS5A in the longer transcript (NM_001100399.1) but does not affect the shorter transcript (NM_001100400.1), suggesting a potential mild defect caused by this variant. However, the role of different isoforms of PDS5A in the cohesin complex is not well established in the literature. Notably, the father of patient 12, who shared the PDS5A p.E759* variant, had speech impediment. Although the pathogenicity of the p.E759* variant in PDS5A remains to be investigated, it may modulate the patient's phenotype and constitute a dual diagnosis together with ASXL3. In addition, one heterozygous de novo SNV (c.654+5G>C) in PDS5A was identified in another patient with neurodevelopmental disorders. This intronic PDS5A variant was predicted to affect splicing of the major messenger RNA (mRNA) transcript of PDS5A by prediction programs including SpliceSiteFinder-like and MaxEntScan (http://www.interactive-biosoftware.com/doc/alamut-visual/2.6/splicing.html).

Finally, one de novo heterozygous SNV in WAPL (NM_015045.3), c.2192G>A (p.R731H), was identified in one patient with neurodevelopmental disorders. This observation corroborates a previous report in which a partial duplication involving WAPL was identified in a patient from a phenotype-driven CdLS cohort, providing further evidence for WAPL as a candidate disease gene.
STAG1 (SNVs/indels, CNV)

- p.R69*
- p.R604Q
- p.W489Vfs*10
- p.N670Ifs*25
- p.S580Vfs*21

STAG2 (SNVs/indels, CNV)

- p.R69*
- p.Q140*
- p.A638Vfs*10
- p.K553Ifs*6
- p.C535* intragenic deletion predicted to result in p.L473_L1198del

Scale

- Patient 6
- Patient 4
- Patient 5

Patient 9, STAG2 c.1605T>A (p.C535*)

Patient 8, STAG2 c.1811G>A (p.R604Q)
Each of the variants in \textit{STAG1}, \textit{STAG2}, \textit{PDSS5A}, and \textit{WAPL} described above were not observed in the control population databases including ExAC and ESP5400 (National Heart, Lung, and Blood Institute [NHBLI] Exome Sequencing Project, \url{http://evs.gs.washington.edu/EVS/}). The interpretation of deleterious effects of the de novo missense SNVs identified in this study was supported by multiple prediction algorithms (Table S5).

We identified CNV deletions affecting \textit{STAG1} and \textit{STAG2} in our clinical CMA cohort, supporting LoF as the presumed disease-contributing mechanism; no putative LoF CNVs of \textit{PDSS5A}, \textit{PDSS5B}, \textit{WAPL}, or \textit{MAU2} were identified. In total, we identified three CNV deletions affecting \textit{STAG1} (two de novo, one of unknown inheritance) in patients with developmental disorders (Fig. 2c, Table S6). In the literature, six CNV deletions overlapping \textit{STAG1} were reported, with the smallest two deletions being intragenic (exons 2–5 and exons 13–18, respectively).\textsuperscript{13} Moreover, eight cases with neurodevelopmental disorders were reported in the DECIPHER database harboring relatively small-sized deletions (<5 Mb) affecting \textit{STAG1} (https://decipher.sanger.ac.uk/\textsuperscript{15}) (Fig. 2c, Table S6). These \textit{STAG1}-overlapping deletions identified in affected patients strongly indicate that haploinsufficiency is likely to be the disease-contributing mechanism for \textit{STAG1}. In addition, a 33.9-Kb CNV deletion with unknown inheritance encompassing exons 15–32 of \textit{STAG2} (predicted to result in an in-frame deletion p.L473_L1198del), was identified in patient 12 with dysmorphic features, microcephaly, and seizures (Fig. 2b, Table S6). This female patient showed skewed XCI, consistent with the observation in patient 8.

\textbf{Patients with \textit{STAG1} and \textit{STAG2} variants have phenotypes overlapping the CdLS spectrum}

We evaluated the clinical phenotypes for patients 1–2 (\textit{STAG1}) and patients 7–11 (\textit{STAG2}). Patient 3 (\textit{STAG1}) was excluded from the evaluation because the identification of concurrent de novo variants in ASXL1 together with \textit{STAG1} may largely complicate the \textit{STAG1}-alone phenotypes.

Patients described in this paper presented for genetic evaluation due to developmental delay and/or congenital anomalies but not with classic distinctive facial features or a recognizable pattern of malformation suggestive for a particular syndrome such as CdLS (Fig. 2d). The most common features among these patients with \textit{STAG1} and \textit{STAG2} variants were DD/ID, behavioral problems, hypotonia, seizures, microcephaly, failure to thrive, short stature, mild dysmorphic features, and 2–3 toe syndactyly (Table 2).

Clinical profiling suggested many overlapping features with CdLS, which include DD/ID, growth failure including short stature and microcephaly, hearing loss, synophrys, micrognathia, limb anomalies, and hypoplastic male genitalia. Some other less common features of CdLS, such as cutis marmorata, myopia, congenital diaphragmatic hernia (CDH), and renal anomalies, among others, were also observed in several of these patients. A more detailed characterization is described in Table 2 and Table S4.

Among the distinctive craniofacial features present in over 95% of the patients with a clinical diagnosis of CdLS,\textsuperscript{13} our patients collectively had microbrachycephaly, low-set ears, synophrys, long curly eyelashes, broad nasal bridge, anteverted nares, long and smooth philtrum, thin upper lip, and micrognathia; however, these features were not present concurrently in a single patient. Interestingly while microcephaly is one of the most characteristic features in CdLS, only 4/7 patients (one \textit{STAG1} and three \textit{STAG2}) had microcephaly. Although the numbers are small, a higher percentage of microcephaly was observed in patients with a \textit{STAG2} variant (3/5) in comparison with \textit{STAG1} (1/2). In contrast to CdLS, where mild to severe limb anomalies are common and are usually helpful to establish a clinical diagnosis, the patients in this study had common but more subtle findings in their extremities, such as fifth finger clinodactyly and syndactyly. Skeletal anomalies including scoliosis (3/7), vertebral anomalies (3/7), and rib fusion (2/7) were observed in our patients, all with variants in \textit{STAG2}. Even though these skeletal anomalies can be observed in patients with classic CdLS, vertebral and rib anomalies would be considered as rare or atypical features for CdLS.

Comparing patients with \textit{STAG1} or \textit{STAG2} variants, DD/ID and mild dysmorphic features have been consistently observed, which is in line with the previous reports\textsuperscript{13–15} (Table 2). Despite the small cohort size, it seems that patients with \textit{STAG2} variants have more multisystem congenital anomalies such as CDH, congenital heart disease, and vertebral anomalies. Growth failure was observed as well, but apparently more in the postnatal period than prenatally. Patients with a \textit{STAG2} variant appear to have more severe growth failure especially in weight and length parameters compared with those with \textit{STAG1} variants.

Although \textit{STAG1} and \textit{STAG2} have been implicated in cancers due to their function in the cohesin pathway and the observation of chromosomal segregation defects in defective cell lines (e.g., \textit{STAG2} as an indicator for myeloid neoplasms), onset of tumors has not been observed in our study nor in the patients reported in the literature with developmental...
Table 2 Genotypes and phenotypes of patients with SNVs/indels in STAG1, STAG2, and PDSSA identified in current study

| Genes | STAG1 (NM_005862.2) | STAG2 (NM_006603.4) | PDSSA (NM_001100399.1) |
|-------|---------------------|---------------------|------------------------|
| **Patients** | | | |
| | Patient 1 | Patient 2 | Patient 3 | Reported in Ref. 13 (n = 17) | Patient 7 | Patient 8 | Patient 9 | Patient 10 | Patient 11 | Reported in Ref. 14 (n = 11) | Patient 13 |
| | | | | | | | | | | | | |
| **Age at last exam** | | | | | | | | | | | | |
| | 11 years | 4 years | 4 years | 30 months to 33 years (median 7 years) | 3 years | 4 years | 1 year | 1 year | 5 years | 8 years | 2 years 6 months |
| | 5 months | 8 months | 2 months | | 8 months | 6 months | | | 3 months | | |
| **Variant** | c.2009_2012del (p.N670fs*25) | c.1129C>T | c.2275G>T (p.E759*) | c.418C>T (p.Q140*) | c.473G>A (p.A158*) | c.3325A>T (p.K1109*) | c.418C>T (p.Q140*) | c.473G>A (p.A158*) | | |
| **Critical gene(s)** | STAG1, STAG2, and PDS5A | STAG1 | STAG1 | STAG1 | STAG1 | STAG1 | STAG1 | STAG1 | STAG1 | AXL3, PDSSA |
| **Gender** | F | M | F | 9 M/8 F | F | F | M | F | F | F |
| **Inheritance** | De novo | De novo | Both de novo | De novo or inherited | De novo | De novo | De novo | De novo | De novo | De novo or inherited |
| **Growth** | IUGR | Failure to thrive | Short stature | Microcephaly | Intellectual disability | Developmental delay | Autism spectrum disorder | IUGR | Failure to thrive | Short stature |
| | | | | | | | | | | | |
| **Development** | + | NR | + | 1/17 | NR | + | + | + | + | + |
| **Neurobehavioral** | | | | | | | | | | | |
| **Behavioral problems** | NR | + | + | NR | NR | + | NR | NR | + | NR |
| **Seizures** | NR | +, during infancy | + | 7/17 with epilepsy | NR | + | NR | NR | + | NR |
| **Hypertonia** | NR | - | - | NR | NR | + | + | NR | + | NR |
| **Hypotonia** | NR | + | + | NR | NR | + | + | NR | + | NR |
| **Craniofacial features** | Long curly eyelashes | Synophrys | Antverted nares | Depressed broad nasal bridge | Bulbous nasal tip | Low-set ears | Dysmorphic ears | Long philtrum | High arched palate | Thin upper lip |
| | | | | | | | | | | | |
| | NR | + | - | See footnotesb | NR | + | - | See footnotesb | NR | + | - | See footnotesb |
| **Skin, nails, hair** | Hypoplastic nails | Hairline | Nails | Cuts marmorata | Ocular | Otoaryngologic | Cardiovascular | | | |
| | | | | | | | | | | | |
| | NR | - | - | NR | NR | - | NR | - | NR | - | |
| **Hairline** | | | | | | | | | | | |
| | | | | | | | | | | |
| **Cardiovascular** | Congenital heart defect | | | | | | | | | | |
| | - | PDA | - | 1/17 | + | NR | NR | - | NR | - | |
| **Respiratory/thorax** | | | | | | | | | | | |
| | | | | | | | | | | |

---

**Footnotes:**

a NR, no record
b NR, no record

---

**Note:** The table includes genotypes and phenotypes of patients with SNVs/indels in STAG1, STAG2, and PDSSA identified in the current study.
| Genes                  | STAG1 (NM_005862.2) | STAG2 (NM_006603.4) | PDSSA (NM_001100399.1) |
|-----------------------|---------------------|---------------------|-----------------------|
| Gastrointestinal      | Gastroesophageal    | MN                 | +, Nissen and G-tube  |
| Genitourinary/ Renal  | reflux              | reflux              |                       |
| Anomaly               |                     |                     |                       |
| Hypoplastic male      | NA                  | NA                  | NA                    |
| genitalia             | NA                  | NA                  |                       |
| Cryptorchidism        | +, left             | +, vertebrae clefts|                       |
| Structural anomalies  | NA, horseshoe kidney| NR, not examined    |                       |
| of the renal tract    |                     |                     |                       |
| Musculoskeletal/      | Scoliosis           | +, T4-5, T10-11, BL|                       |
| extremities           | Rib fusion          | +, vertebral clefts|                       |
| Vertebral anomalies   | -                   | -                   |                       |
| Arm/hand anomalies    | +                   | -                   |                       |
| Limited elbow extension| +                   | -                   |                       |
| Fifth finger          | +                   | +                   |                       |
| clinodactylly         | -                   | -                   |                       |
| Single transverse      | NR                  | NR                  |                       |
| palmar crease         | +                   | +                   |                       |
| 2–3 toe syndactyly    | +                   | -                   |                       |
| Studies and imaging   | Abnormal brain MRI | -                   | +, mild               |
| Abnormal              | -                   | NR                  | +, ectopic posterior  |
|                       |                     | -                   | pituitary stalk       |

BL bilateral, CA coarctation of the aorta, IUGR intrauterine growth retardation, NA not applicable, NR no record, PDA patent ductus arteriosus, PS pulmonic stenosis, VSD ventricular septal defect, PFO patent foramen ovale, CNV copy-number variant, SNV, single-nucleotide variant, MRI magnetic resonance image

*STAG1 was affected by both CNV deletion and SNVs/indels. The deletions included three de novo and one with unknown parent of origin, which encompassed STAG1 and PCCB; one intragenic, which was absent in the mother; and two intragenic, which were maternally inherited; the SNVs/indels included eight de novo missense and two de novo frameshift variants of STAG1. Facial features included 14/17 with deep-set eyes, 13/17 with wide mouth, 7/17 with high nasal bridge, 8/17 with thin eyebrows, 4/17 with widely spaced central incisors, micrognathia, ear abnormalities, wide-set eyes, beaked or prominent nose, arched eyebrows, or low-set ears, cleft/arched palate.
discorders caused by constitutional pathogenic variants in\nSTAG1 and STAG2 (refs.13–15). Moreover, no obvious\nincreased risk of cancer is reported in patients with other\ncohesinopathies caused by defects in genes such as NIPBL,\nSMC1A, and SMC3 (ref. 1). Consistent with this observation,\nour chromosome analysis of one patient (patient 7) did not\nreveal any evidence for chromosomal segregation defects\n(data not shown).

**DISCUSSION**

In this study, we applied a genotype-driven approach to\ndecipher the genetic causes of cohesinopathy from a CES\nperspective. We describe a series of disease-contributing\nvariants in known cohesinopathy genes, and also provide\nmolecular evidence supporting the candidacy of recently\ndescribed or new disease genes.

*NIPBL* defects are underrepresented in this cohort likely\ndue to ascertainment bias associated with its more clinically\nrecognizable presentations. The scarcity of putative LoF\nvariants for certain cohesin genes including *PDS5B* and\n*MAU2* in this cohort indicates that LoF variants in these genes\nmay exert strong pathogenic effects on early development\nleading to incompatibility with life. Alternatively, the lack of\nevidence supporting the pathogenicity of variants in *PDS5B*\nand *MAU2* could reflect limitations of interpreting missense\nvariants based on proband-only CES. HDAC8 and *SMC1A*\nare the only two well-studied X-linked genes among the cohesin\ncomponents. They seem to be relatively spared from the\nstrong selection in human development possibly due to\nprotection of pathogenic alleles in the gene pool by XCI in\nfemales. Consistently, variants in these two genes are highly\nrepresented in the CES cohort as compared with cohorts\nassembled by phenotypic characterization (Fig. 1b).

Patients harboring *STAG1* or *STAG2* variants seem to share\nmany of the clinical features seen in the well-described CdLS\nphenotype. Apparently affected patients in our cohort are\ndevelopmentally and intellectually impaired as those with\nCdLS. However, their spectrum of growth, craniofacial, and\nmusculoskeletal features are not as severe as the spectrum of\nCdLS. Overall, only one patient (patient 3 [STAG1]) fulfills\nthe diagnostic criteria for CdLS by meeting the CdLS\ncharacteristic facial features.26 Note that the concurrent de\nnovo variant in *ASXL1* may largely contribute to the\ndifferential diagnosis of CdLS for patient 3 (Table S7).\nAlthough the currently available clinical information we had\nmight not be as sufficient for a diagnosis of CdLS or other\ncohesinopathies, a “CdLS-like” syndrome started to emerge.\nThe *STAG1/STAG2*-related disorders seem to be at the mild\nend of the CdLS spectrum, making the clinical diagnosis for\nthese two genes more challenging for physicians. Putting\ntogether the constellation of clinical features might help to\nend the diagnostic odyssey earlier, and with this series of cases\nawareness can be extended. Given the challenges, compre-\nhensive genomic analysis, such as CES, should be offered to\nefficiently provide a molecular diagnosis for these cohesino-\npathy conditions.

Notably, the LoF *PDS5A* variant (patient 13) was inherited\nfrom a father with speech impediment. Although the\nphenotypic consequence of this variant remains unclear (as\ndiscussed in Results), its potential contribution cannot be\ncompletely ruled out. Unfortunately, samples from the\npaternal grandparents or other relatives are not available for\ntesting. Defects in the cohesin complex, as demonstrated in\nthe CdLS genes, are likely to be detrimental to proper\norganisal development, and milder phenotypic conse-\nquences have been observed.11 With our experience of known\nCdLS gene variants among 10,698 individuals, two distinct\nnovel pathogenic variants in *RAD21* as well as one novel\npathogenic variant in *SMC1A* (X-linked) were identified in\nthree unrelated patients with neurodevelopmental disorders,\nall inherited from affected parents with milder phenotypes\n(Table 1). Moreover, transmission of a pathogenic variant\nbetween generations has been reported in *STAG1* (ref. 13).\nTherefore, with the reported variable expressivity of the\ncohesin defects, it is plausible that the reproductive potential,\ngeneetic transmission, and severity of phenotype may be\ndependent on various factors, including the components\nbeing affected, the PV types, the inheritance mode (e.g., X-\nlinded or autosomal dominant), and the downstream\npaths disrupted by defects in a particular component. Thus,\nadditional genotype–phenotype correlation studies are war-\nranted to further delineate the spectrum of cohesinopathies.

The mutational landscape of cohesin genes in somatic\ncancer may represent an alternative view to reflect contribu-\ntion of these genes to biological processes, with minimum\nselection as compared with that imposed during early human\ndevelopment. Among cancer samples deposited to the\nCMOC database subjected to genome-wide screening,\ntruncating variants were observed in all cohesin genes. While\nmisense variants did not show any substantive difference\nbetween cohesin genes, putative LoF variants in *STAG2* were\nhighly represented in the somatic cancer cohort (Fig. 1c).\nLoF variants in *STAG2* have been significantly associated with\nseveral cancers,27,28 suggesting a likely pleiotropic effect of\n*STAG2*, possibly with strong involvement in tumorigenesis.\nInterestingly, we have observed a patient with mosaic *STAG2*\nLoF variant in the CES cohort. The patient does not have\nneurodevelopmental problems, but instead presented with\nhematological malignancy. Therefore, we considered the\n*STAG2* defect in this patient as not being causal for a\ncohesinopathy. Consequently, caution should be taken when\ninterpreting variants in cohesin genes by considering the\npossibility that they may arise as somatic changes after the\ncritical period of early human development.

Accumulating evidence suggests that cohesin contributes to\nthe topological organization of the genome, regulates DNA\nreplication, and facilitates long-range gene transcription\nregulation.29,30 In addition, the interactions between cohesin\nand other transcription machinery and chromatin remodeling\ncomplexes to recognize specific genomic loci and regulate\ngene transcription have aggregated these complexes into\nthe same pathways of transcription regulation.30–33 Notably,
genes encoding components of chromosome remodeling and transcription regulation machineries, such as ANKRD11, AFF4, KMT2A, TAF1, and TAF6, have been associated with phenotypes reminiscent of CdLS. Such findings expand the molecular mechanism underlying cohesinopathies into transcriptional regulation. Interestingly, gene expression studies of patients with elevated dosage of STAG2 reveal a dysregulated transcriptome and pinpoint altered expression levels of developmentally important genes. Therefore, the versatility of cohesin in cohesion and transcription regulation warrants a further investigation of its downstream effectors.

In summary, the genotype-first approach focusing on a specific pathway enabled us to investigate patients with nonclassic cohesinopathy phenotypes; this approach also allowed us to discover patients with variants in new or recently reported disease genes, namely STAG1, STAG2, and potentially PDS5A and WAPL, which may further expand the genetic heterogeneity underlying cohesinopathies. Future studies of cellular phenotypes, with regard to functional studies of DNA repair and transcriptome analysis, are warranted to further elucidate the mechanistic consequences due to defects in specific cohesin components, which may shed light on precision medicine efforts targeting distinct molecular pathways.

**ELECTRONIC SUPPLEMENTARY MATERIAL**

The online version of this article (https://doi.org/10.1038/s41436-018-0085-6) contains supplementary material, which is available to authorized users.

**ACKNOWLEDGEMENTS**

This study was supported in part by the National Human Genome Research Institute/National Heart, Lung, and Blood Institute (NHGRI/NHLBI) grant UM1HG006542 to the BHCMG; and National Institutes of Neurological Disorders and Stroke (NINDS) grant R35 NS105078-01 to JRL. JEP was supported by the NHGRI grant K08 HG008986. AHC, ELB and LR are supported by Newlife (Ref:16-17/12). We acknowledge Dr. Suren V. Mulegama for critical comments on this manuscript. This study makes use of data generated by the DECIPHER community. A full list of centers that contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via e-mail from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003), a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute (grant number WT098051). The views expressed in this publication are those of the author(s) and not necessarily those of the Wellcome Trust or the Department of Health. The study has UK Research Ethics Committee (REC) approval (10/H0305/83, granted by the Cambridge South REC, and GEN284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network.

**DISCLOSURE**

Bayor College of Medicine (BCM) and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of Baylor Genetics (BG), formerly the Baylor Miraca Genetics Laboratories (BMGL), which performs chromosomal microarray analysis and clinical exome sequencing. JR, VP, WJ, CS, WB, SWC, AMB, JLS, CE, YY, RX, and PL are employees of BCM and derive support through a professional services agreement with the BG. JRL serves on the Scientific Advisory Board of the BG. JRL has stock ownership in 23andMe, is a paid consultant for Regeneron Pharmaceuticals, has stock options in Lasergen, Inc., and is a coinventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting. The other authors declare no conflict of interest. All authors read and approved the final manuscript.

**REFERENCES**

1. Liu J, Krantz ID. Cornelia de Lange syndrome, cohesin, and beyond. Clin Genet. 2009;76:303–14.
2. Losada A. Cohesin in cancer: chromosome segregation and beyond. Nat Rev Cancer. 2014;14:389–93.
3. Yuan B, Pehlivan D, Karaca E, et al. Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. J Clin Invest. 2015;125:636–51.
4. Krantz ID, McCallum J, DeScipio C, et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nat Genet. 2004;36:631–5.
5. Pehlivan D, Hullings M, Carvalho CM, et al. NIPBL rearrangements in Cornelia de Lange syndrome: evidence for replicative mechanism and genotype-phenotype correlation. Genet Med. 2012;14:313–22.
6. Tonkin ET, Wang TI, Liso S, et al. NIPBL, encoding a homolog of fungal Scsc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. Nat Genet. 2004;36:636–41.
7. Deardorf MA, Bando M, Nakato R, et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature. 2012;489:313–7.
8. Deardorf MA, Kaur M, Yaeger D, et al. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. Am J Hum Genet. 2007;80:485–94.
9. Deardorf MA, Wilde JJ, Albrecht M, et al. RAD21 mutations cause a human cohesinopathy. Am J Hum Genet. 2012;90:1014–27.
10. Musio A, Selicorini A, Focarelli ML, et al. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat Genet. 2006;38:528–30.
11. Deardorf MA, Noon SE, Krantz ID. Cornelia de Lange syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews. Seattle, WA: University of Washington; 2005.
12. Mannini L, Cucco F, Quaratonti V, et al. Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. Hum Mutat. 2013;34:1589–96.
13. Lehalle D, Mosca-Boidron AL, Beugtrup A, et al. STAG1 mutations cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. J Med Genet. 2017;54:479–88.
14. Mulegama SV, Klein SD, Mulatinho MV, et al. De novo loss-of-function variants in STAG2 are associated with developmental delay, microcephaly, and congenital anomalies. Am J Med Genet A. 2017;173:1319–27.
15. Soardi FC, Machado-Silva A, Linhares ND, et al. Familial STAG2 germline mutation defines a new human cohesinopathy. NPI Genom Med. 2017;2:7.
16. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369:1502–11.
17. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014;312:1870–9.
18. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation
of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17:405–24.

19. Ansari M, Poke G, Ferry Q, et al. Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. *J Med Genet.* 2014;51:659–68.

20. Gambin T, Yuan B, Bi W, et al. Identification of novel candidate disease genes from de novo exonic copy number variants. *Genome Med.* 2017;9:83.

21. Allen RC, Zoghbi HY, Moseley AB, et al. Methylation of Hpall and Hhal sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet.* 1992;51:1229–39.

22. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.

23. Posey JE, Harell T, Liu P, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med.* 2017;376:21–31.

24. Pehlivan D, Erdin S, Carvalho CBM, et al. Evidence implicating cohesin/condensin gene noncoding CNVs in the Cornelia de Lange. (Meeting abstract) *Genomic Disorders 2012: The Genomics of Rare Diseases.* 2012.

25. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. *Am J Hum Genet.* 2009;84:524–33.

26. Kline AD, Krantz ID, Sommer A, et al. Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. *Am J Med Genet A.* 2006;143A:1287–96.

27. Solomon DA, Kim JS, Bondaruk J, et al. Frequent truncating mutations of STAG2 in bladder cancer. *Nat Genet.* 2013;45:1428–30.

28. Solomon DA, Kim T, Diaz-Martinez LA, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science. 2011;333:1039–43.

29. Sofuova S, Yaffe E, Chan WC, et al. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 2013;32:3119–29.

30. Kagey MH, Newman JJ, Bilodeau S, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature.* 2010;467:430–5.

31. Rubio ED, Reiss DJ, Walsh P, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A.* 2008;105:8309–14.

32. Allen BL, Taatjes DJ. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol.* 2015;16:155–66.

33. Strubbe G, Popp C, Schmidt A, et al. Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. *Proc Natl Acad Sci U S A.* 2011;108:5572–7.

34. Izumi K, Nakato R, Zhang Z, et al. Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. *Nat Genet.* 2015;47:338–44.

35. O’Rave JA, Wu Y, Dorfel MJ, et al. TAF1 variants are associated with dysmorphic features, intellectual disability, and neurological manifestations. *Am J Hum Genet.* 2015;97:922–32.

36. Parenti I, Gervasini C, Pozojevic J, et al. Broadening of cohesinopathies: exome sequencing identifies mutations in ANKRD11 in two patients with Cornelia de Lange-overlapping phenotype. *Clin Genet.* 2016;89:74–81.

37. Kumar R, Corbett MA, Van Bon BW, et al. Increased STAG2 dosage defines a novel cohesinopathy with intellectual disability and behavioral problems. *Hum Mol Genet.* 2015;24:7171–81.

38. Gillis LA, McCallum J, Kaur M, et al. NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. *Am J Hum Genet.* 2004;75:610–23.

39. Liu J, Feldman R, Zhang Z, et al. SMC1A expression and mechanism of pathogenicity in probands with X-linked Cornelia de Lange syndrome. *Hum Mutat.* 2009;30:1535–42.

40. Goldstein JH, Tim-Aroon T, Shieh J, et al. Novel SMC1A frameshift mutations in children with developmental delay and epilepsy. *Eur J Med Genet.* 2015;58:562–8.