Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes

Bo Yuan, Davut Pehlivan, Ender Karaca, Nisha Patel, Wu-Lin Charng, Tomasz Gambin, Claudia Gonzaga-Jauregui, V. Reid Sutton, Gozde Yesil, Sevcan Tug Bozdogan, Tulay Tos, Asuman Koparir, Erkan Koparir, Christine R. Beck, Shen Gu, Huseyin Aslan, Ozge Ozalp Yuregir, Khalid Al Rubeaan, Dhekra Alnaqeb, Muneera J. Alshammari, Yavuz Bayram, Mehmed M. Atik, Hatip Aydin, B. Bilge Geckinli, Mehmet Seven, Hakan Ulucan, Elif Fenercioglu, Mustafa Ozen, Shalini Jhangiani, Donna M. Muzny, Eric Boerwinkle, Beyhan Tuysuz, Fowzan S. Alkuraya, Richard A. Gibbs, James R. Lupski

Research Article **Genetics**

Cornelia de Lange syndrome (CdLS) is a genetically heterogeneous disorder that presents with extensive phenotypic variability, including facial dysmorphism, developmental delay/intellectual disability (DD/ID), abnormal extremities, and hirsutism. About 65% of patients harbor mutations in genes that encode subunits or regulators of the cohesin complex, including *NIPBL*, *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*. Wiedemann-Steiner syndrome (WDSTS), which shares CdLS phenotypic features, is caused by mutations in lysine-specific methyltransferase 2A (*KMT2A*). Here, we performed whole-exome sequencing (WES) of 2 male siblings clinically diagnosed with WDSTS; this revealed a hemizygous, missense mutation in *SMC1A* that was predicted to be deleterious. Extensive clinical evaluation and WES of 32 Turkish patients clinically diagnosed with CdLS revealed the presence of a de novo heterozygous nonsense *KMT2A* mutation in 1 patient without characteristic WDSTS features. We also identified de novo heterozygous mutations in *SMC3* or *SMC1A* that affected RNA splicing in 2 independent patients with combined CdLS and WDSTS features. Furthermore, in families from 2 separate world populations segregating an autosomal-recessive disorder with CdLS-like features, we identified homozygous mutations in *TAF6*, which encodes a core transcriptional regulatory pathway component. Together, our data, along with recent transcriptome studies, suggest that CdLS and related phenotypes may be "transcriptomopathies" rather than cohesinopathies.

Find the latest version:

Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes

Bo Yuan,[1] Davut Pehlivan,[1] Ender Karaca,[1] Nisha Patel,[2] Wu-Lin Charng,[1] Tomasz Gambin,[1] Claudia Gonzaga-Jauregui,[1] V. Reid Sutton,[1] Gozde Yesil,[3] Sevcan Tug Bozdogan,[4] Tulay Tos,[5] Asuman Koparir,[6] Erkan Koparir,[7] Christine R. Beck,[1] Shen Gu,[1] Huseyin Aslan,[8] Ozge Ozalp Yuregir,[8] Khalid Al Rubeaan,[9] Dhekra Alnaqeb,[10] Muneera J. Alshammari,[11] Yavuz Bayram,[1] Mehmed M. Atik,[1] Hatip Aydin,[12] B. Bilge Geckinli,[13] Mehmet Seven,[6] Hakan Ulucan,[6] Elif Fenercioglu,[6] Mustafa Ozen,[6] Shalini Jhangiani,[14] Donna M. Muzny,[14] Eric Boerwinkle,[14,15] Beyhan Tuysuz,[16] Fowzan S. Alkuraya,[2,17] Richard A. Gibbs,[1,14] and James R. Lupski[1,14,18,19]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine (BCM), Houston, Texas, USA. [2]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia. [3]Department of Medical Genetics, Bezmialem University, Istanbul, Turkey. [4]Department of Medical Genetics, Mersin University, Mersin, Turkey. [5]Department of Medical Genetics, Dr. Sami Ulus Research and Training Hospital of Women's and Children's Health and Diseases, Ankara, Turkey. [6]Department of Medical Genetics, Cerrahpasa Medical School of Istanbul University, Istanbul, Turkey. [7]Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey. [8]Department of Medical Genetics, Adana Numune Training and Education Hospital, Adana, Turkey. [9]University Diabetes Center, College of Medicine, [10]Research Department, University Diabetes Center, and [11]Department of Pediatrics, King Khalid University Hospital and College of Medicine, King Saud University, Riyadh, Saudi Arabia. [12]Center of Genetics Diagnosis, Zeynep Kamil Maternity and Children's Training and Research Hospital, Istanbul, Turkey. [13]Department of Medical Genetics, Marmara University School of Medicine, Istanbul, Turkey. [14]Human Genome Sequencing Center, BCM, Houston, Texas, USA. [15]University of Texas Health Science Center at Houston, Houston, Texas, USA. [16]Department of Pediatrics, Division of Medical Genetics, Cerrahpasa School of Medicine, Istanbul, Turkey. [17]Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. [18]Department of Pediatrics, BCM, Houston, Texas, USA. [19]Texas Children's Hospital, Houston, Texas, USA.

Cornelia de Lange syndrome (CdLS) is a genetically heterogeneous disorder that presents with extensive phenotypic variability, including facial dysmorphism, developmental delay/intellectual disability (DD/ID), abnormal extremities, and hirsutism. About 65% of patients harbor mutations in genes that encode subunits or regulators of the cohesin complex, including *NIPBL*, *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*. Wiedemann-Steiner syndrome (WDSTS), which shares CdLS phenotypic features, is caused by mutations in lysine-specific methyltransferase 2A (*KMT2A*). Here, we performed whole-exome sequencing (WES) of 2 male siblings clinically diagnosed with WDSTS; this revealed a hemizygous, missense mutation in *SMC1A* that was predicted to be deleterious. Extensive clinical evaluation and WES of 32 Turkish patients clinically diagnosed with CdLS revealed the presence of a de novo heterozygous nonsense *KMT2A* mutation in 1 patient without characteristic WDSTS features. We also identified de novo heterozygous mutations in *SMC3* or *SMC1A* that affected RNA splicing in 2 independent patients with combined CdLS and WDSTS features. Furthermore, in families from 2 separate world populations segregating an autosomal-recessive disorder with CdLS-like features, we identified homozygous mutations in *TAF6*, which encodes a core transcriptional regulatory pathway component. Together, our data, along with recent transcriptome studies, suggest that CdLS and related phenotypes may be "transcriptomopathies" rather than cohesinopathies.

Introduction

Cornelia de Lange syndrome (CdLS, MIM # 122470, MIM # 300590, MIM # 610759, MIM # 614701, and MIM # 300882) was initially described by Brachmann in 1916 and Cornelia de Lange in 1933 (1, 2). CdLS is a multisystem malformation syndrome characterized primarily by 4 major clinical features including distinctive facial dysmorphism, developmental delay/intellectual disability (DD/ID), extremity abnormality, and hirsutism. However, a wide

spectrum of abnormalities is observed among CdLS cases, which is partially attributed to underlying genetic heterogeneity. The causal genes identified to date consist exclusively of those encoding proteins seemingly related to the cohesin complex function. Nipped-B-like (*NIPBL*) and structural maintenance of chromosomes 1A (*SMC1A*) and 3 (*SMC3*) are the first 3 well-characterized causal genes for CdLS (3–6). Up to 60% of CdLS cases harbor mutations in *NIPBL*, which encodes a regulatory protein loading the cohesin complex onto the sister chromatids. Copy number variants (CNVs) in the *NIPBL* locus are found in approximately 5% of *NIPBL* point mutation–negative CdLS cases, while mosaic mutations in *NIPBL* are reported to account for an additional 23% (7–10). *SMC1A* and *SMC3* are core structural components of the cohesin complex. Approximately 5% of CdLS cases are found to have missense mutations or small in-frame deletions in *SMC1A* and *SMC3*. RAD21 homolog (*Schizosaccharomyces pombe*) (*RAD21*) and histone deacetylase 8 (*HDAC8*) are the 2 most recently identified genes

in which mutations are associated with a CdLS-like phenotype: *RAD21* encodes a key component of cohesin, providing a physical link between SMC1A and SMC3, while *HDAC8* encodes a lysine deacetylase that regulates SMC3 (11, 12).

Recently, it has become apparent that the cohesin complex not only functions in sister chromatid cohesion, but also plays a role in the regulation of transcription. Evidence suggests that *NIPBL*, *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*, the 5 genes associated with cohesin and CdLS, are each involved in transcriptional regulation (11–17). Variant alleles in genes encoding proteins important to cohesion function are also associated with other syndromes sharing phenotypic features with CdLS, such as Roberts–SC phocomelia syndrome (MIM # 268300, MIM # 269000) (18) and Warsaw breakage syndrome (MIM # 613398) (19). Due to the extensive clinical heterogeneity and imperfect genotype-phenotype correlations, it is important to investigate the genetic heterogeneity underlying CdLS and CdLS-like phenotypes and to enrich our understanding of the phenotypic spectrum caused by the corresponding gene mutations, i.e., phenotypic expansion.

Sharing clinical features with CdLS, Wiedemann-Steiner syndrome (WDSTS, MIM # 605130) is another heterogeneous disease, first described by Wiedemann et al. in 1989 (20). Also known as "hairy elbows syndrome," it is characterized by excessive hair growth around the distal third of the upper arm and the proximal third of the forearm and occasionally on other parts of the body (e.g., sacrum) along with short stature, distinctive facial features, and DD/ID. De novo deleterious mutations in lysine-specific methyltransferase 2A (*KMT2A*) were identified in patients with a clinical diagnosis of WDSTS (21). However, these patients show a moderately different spectrum of clinical findings, indicating that different mutations in *KMT2A*, or allelic heterogeneity, contribute to a spectrum of phenotypes. To date, a wide spectrum of abnormalities in WDSTS is observed, among which hairy elbows is found as the most prominent shared clinical feature (21–23).

Genetic studies of CdLS and the revelation of the various contributing cohesin proteins exemplify the use of a shared phenotypic definition to clinically derive a syndrome diagnosis, yet reveal a surprisingly complex genetic architecture that nonetheless has a unifying molecular role. To investigate a reciprocal approach interrogating the molecular basis of a disease phenotype, we initiated personal genome studies using whole-exome sequencing (WES), rather than interrogating variations at a specific locus, to determine whether patients with CdLS-like phenotypes, including WDSTS, harbor mutations in genes that are related to those known to underlie classical CdLS or perhaps reveal disease genes that may further broaden our understanding of the molecular perturbations leading to the CdLS phenotype. We report a previously unidentified, apparently recessive form of CdLS from 2 distinct world populations and provide evidence that CdLS could manifest as a result of a perturbed transcriptional profile and thus perhaps be better classified as a transcriptomopathy rather than a cohesinopathy.

Results

SMC1A variants in patients with WDSTS. Two brothers (WDSTS-1 and WDSTS-2) diagnosed with WDSTS were born to nonconsanguineous phenotypically normal Turkish parents. Clinical findings of the 8-year-old elder brother (WDSTS-1) included excessive hair growth primarily localized to the elbows and sacrum, cryptorchidism, hirsutism, fifth-finger clinodactyly, pectus excavatum, DD/ID, and facial dysmorphic features, with thick eyebrows, long eyelashes, long philtrum, synophrys, thin upper lip, and low-set ears. Short middle phalanges were also observed, indicating the hypoplastic third digit and partial syndactyly of the toes. A similar phenotype was also observed in the younger 6-year-old brother (WDSTS-2) (Figure 1 and Table 1).

DNA samples from each subject were tested for submicroscopic CNVs by both genome-wide BCM chromosomal microarray (CMA version 9.1.1) (24) and a high-resolution custom-designed CdLS array comparative genomic hybridization (aCGH) covering all cohesin-related genes, considering the retrospectively CdLS-like phenotypic features. No plausible potential pathogenic CNV was detected (data not shown). WES for the 2 affected children revealed an overlapping hemizygous missense substitution (ChrX: g.53442107 G>A [hg19]; c.C121T, p.Leu41Phe) in *SMC1A* (NM_006306) located on Xp11.22. Segregation analysis showed that the mother was a heterozygous carrier, while the father was hemizygous WT, confirming the X-linked inheritance of this variant (Figure 2A). The variant had not been previously identified and has not been observed in control populations, including the NHLBI Exome Sequencing Project (ESP, http://evs.gs.washington.edu/EVS/), the 1000 Genomes project (TGP, http://www.1000genomes.org/), and Atherosclerosis Risk in Communities (ARIC; https://www2.cscc.unc.edu/aric/), an internal control database used at BCM containing approximately 4,000 exomes (Table 2). The aa Leu41 in SMC1A is highly conserved from *Danio rerio* to *Homo sapiens* (Figure 2B).

Variants in the known CdLS genes. We undertook genomic studies in a cohort of Turkish subjects clinically diagnosed with CdLS. All studies were performed on the DNA samples extracted from the peripheral blood of the patients. We performed aCGH prior to WES studies for a cohort of 32 Turkish patients with CdLS, as CNVs in the *NIPBL* locus were reported in approximately 5% of *NIPBL* point mutation–negative CdLS cases (7). All 32 patients in our Turkish cohort were negative for plausible pathogenic CdLS-associated CNVs (data not shown); thus, we proceeded to perform WES. We first examined patient genomes for rare variants in known CdLS-associated genes, including *NIPBL*, *SMC1A*, *SMC3*, *RAD21*, and *HDAC8*, which were simultaneously interrogated by our WES approach. As anticipated, in 14 out of 32 unrelated subjects with CdLS, 14 deleterious rare variants were identified in the known disease-associated genes, including 10 in *NIPBL*, 2 in *SMC1A*, 1 in *SMC3*, and 1 in *HDAC8*, documenting *NIPBL* as a major contributory gene for clinically diagnosed CdLS subjects from the Turkish population (Table 3). Segregation analyses for the candidate variants confirmed 6 de novo *NIPBL* mutations in families in which both parental samples were available for study. The mutations in *SMC1A*, *SMC3*, and *HDAC8* were in all cases confirmed to be de novo (Table 3). In total, the identification rate of deleterious variants in the known genes in our Turkish CdLS cohort was 43.75% (31.25% in *NIPBL*, 6.25% in *SMC1A*, 3.125% in *SMC3*, and 3.125% in *HDAC8*), which was comparable with the findings in previously published world cohorts (25–28). These observations showed both the reliability of the clinical diagnosis in this Turkish cohort and the utility of WES to establish a molecular diagnosis in this genetically heterogeneous trait.
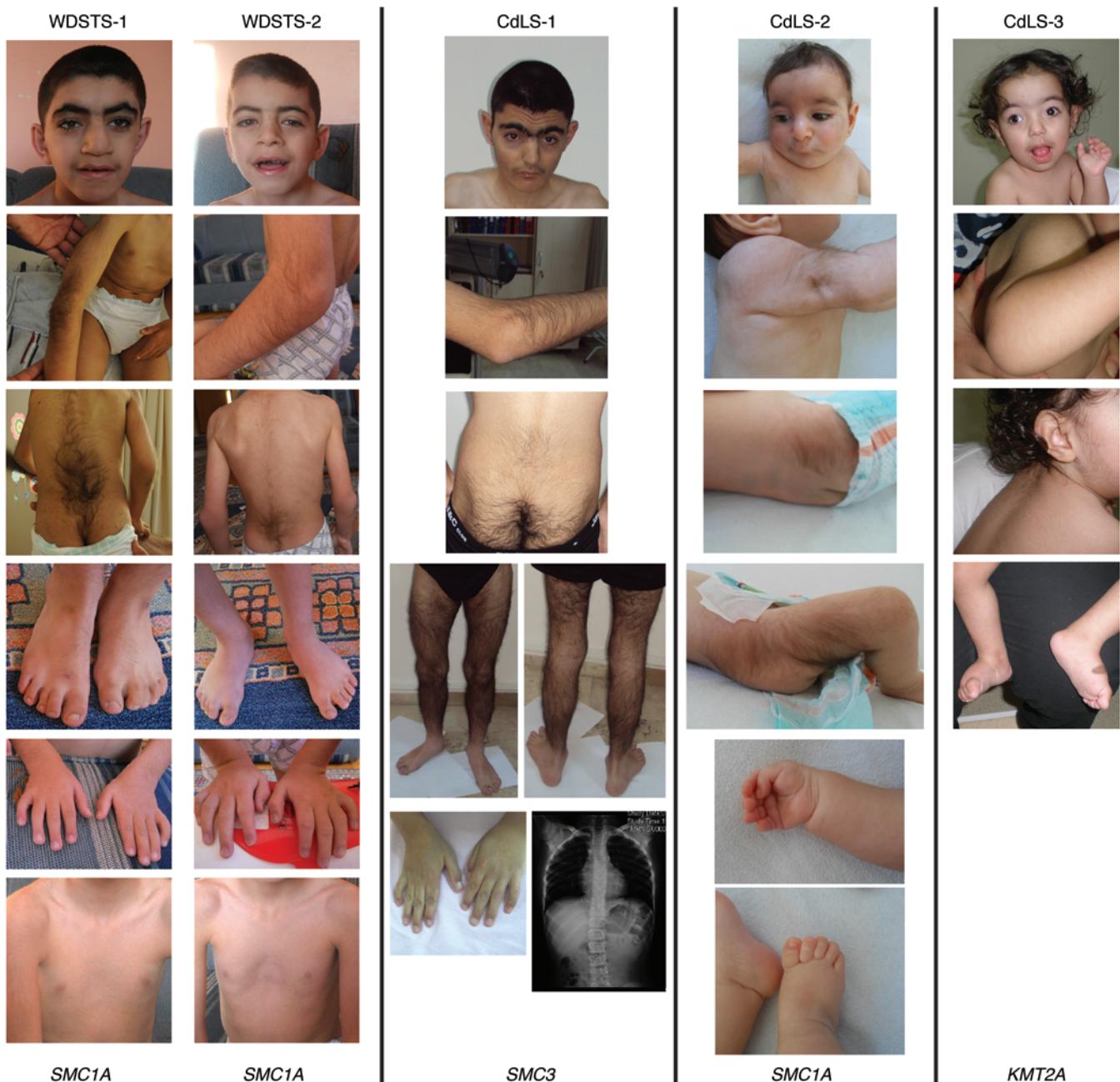
Figure 1. Photographs showing the representative clinical features of the patients WDSTS-1, WDSTS-2, CdLS-1, CdLS-2, and CdLS-3. Note the eyebrows of the WDSTS-1 and WDSTS-2 are bushy, whereas the eyebrows of CdLS-1 and CdLS-3 have the typical arched appearance of CdLS with synophrys. WDSTS-1 and WDSTS-2 also have square and bulbous nasal tips. The second row of images demonstrates hirsutism of elbows in WDSTS-1, WDSTS-2, CdLS-1, and CdLS-2. Also of note is hirsutism of lower back, buttocks, legs, and face in the subjects. All individuals have modest brachydactyly of fingers and/or toes, and WDSTS-1, WDSTS-2, and CdLS-1 have fifth finger clinodactyly. WDSTS-1 and WDSTS-2 have pectus excavatum, and in CdLS-1, mild mid-thoracic scoliosis can be seen on the plain radiograph of the chest.

Prior to our study, only 1 de novo variant in *SMC3* had been reported (6). Interestingly, in our cohort, a previously unidentified de novo 11 base-pair deletion variant encompassing an intron-exon junction (Chr10: g.112360773 GTTACAGGAACT>G [hg19]) in *SMC3* (NM_005445) was identified in a 17-year-old male patient (CdLS-1) (Tables 2 and 3, and Figure 3A). He presented with microcephaly, strabismus, hypermetropia, clinodactyly, restricted elbow movement, sandal gap of the feet, scoliosis, mild mixed-type hearing loss, hirsutism, DD/ID, and a dysmor-phic facial appearance, with micrognathia, long eyelashes, heavy eyebrows, synophrys, and large and prominent ears. Very striking excessive hair growth was also observed on the elbows, sacrum, and legs (Figure 1). In silico investigation suggested that the 11-bp deletion, spanning the 23rd intron-exon junction, might affect splicing by removing the splicing acceptor on the 3′ end of the intron. We used subject-derived blood samples to test these in silico predictions and our alternative splicing hypothesis. Total RNA was obtained from 4 family members (patient, unaffected

Table 1. Clinical summaries of the patients with clinical diagnoses of WDSTS or CdLS

Patients	BAB4135 (WDSTS-1)	BAB4136 (WDSTS-2)	BAB5593 (CdLS-1)	BAB5452 (CdLS-2)	BAB4964 (CdLS-3)	Deardorff et al., ref. 6[A]	Deardorff et al., ref. 6[A]
Clinical diagnosis	WDSTS	WDSTS	CdLS	CdLS	CdLS	CdLS	CdLS
Sex	Male	Male	Male	Female	Female	Female	Male
Genes	*SMC1A*	*SMC1A*	*SMC3*	*SMC1A*	*KMT2A*	*SMC1A*	*SMC1A*
Mutations	c.C121T/p.L41F	c.C121T/p.L41F	c.2536_2547del/p.E846_E849del	c.2794_2982del/p.D992_Q994del	c.C2233T/p.R745X	c.173del15/p.V58_R62del	c.T397G/p.F133V
Inheritance	Maternal	Maternal	De novo	De novo	De novo	De novo	De novo
Age	8 years	6 years	17 years	8 months	21 months	Unknown	Unknown
Prenatal growth retardation	Mild			+	+	+	
Short stature			+	+	+	+	+
Slim, muscular build	+	+		+			
Failure to thrive				+	+	+	+
Microcephaly	+	+	+	+	+		
Thick eyebrows	+	+	+	+			
Thin/arched eyebrows					+	+	+
Synophrys	+	+	+	+	+	+	+
Long curly eyelashes	+	+	+	+	+	+	+
Palpebral fissures		Downslanting		Downslanting	Narrow/downslanting		
Epicanthal folds				+	+		
Hypertelorism				+	+		
Strabismus			+		Exotropia		
Vision			Hypermetropia			Myopia	Astigmatism
Nasal bridge					Low	Low	High
Broad nose	+	+		+			
Long philtrum	+	+		+	+	+	+
Thin upper lip	+	+	+	+	+	+	+
Cupid's bow, exaggerated		+		+			
High arched palate				+	+		
Micrognathia			+	+	+		
Downturned corners of the mouth	+			+	+	+	+
Widely spaced teeth					+		
Low-set ears	+	+					
Dysmorphic ears	Hypoplastic antihelix				+		
Hearing loss			Mixed	Sensorineural		Sensorineural	
Low posterior hair line	+				+		
Congenital heart defect			+		PDA		
Cryptorchidism	+	+					
Extremity problems			Elbow limitation	Small hands and feet			Restriction of elbow movement
Brachydactyly	Moderate	Moderate	Moderate	Moderate	Moderate		
Fifth finger clinodactyly	+	+	+	+	+	+	
Hirsutism	+	+	+	+	+	+	
Hairy elbows	+	+	+	+			
DD/ID	+	mild	+	+	+	+	+
Behavioral problems			Nervous, forgetful				Anxiety

[A]Two patients with mutations in the N-terminal NTPase domain of SMC1A previously reported by Deardorff et al. (ref. 6). PDA, patent ductus arteriosus.
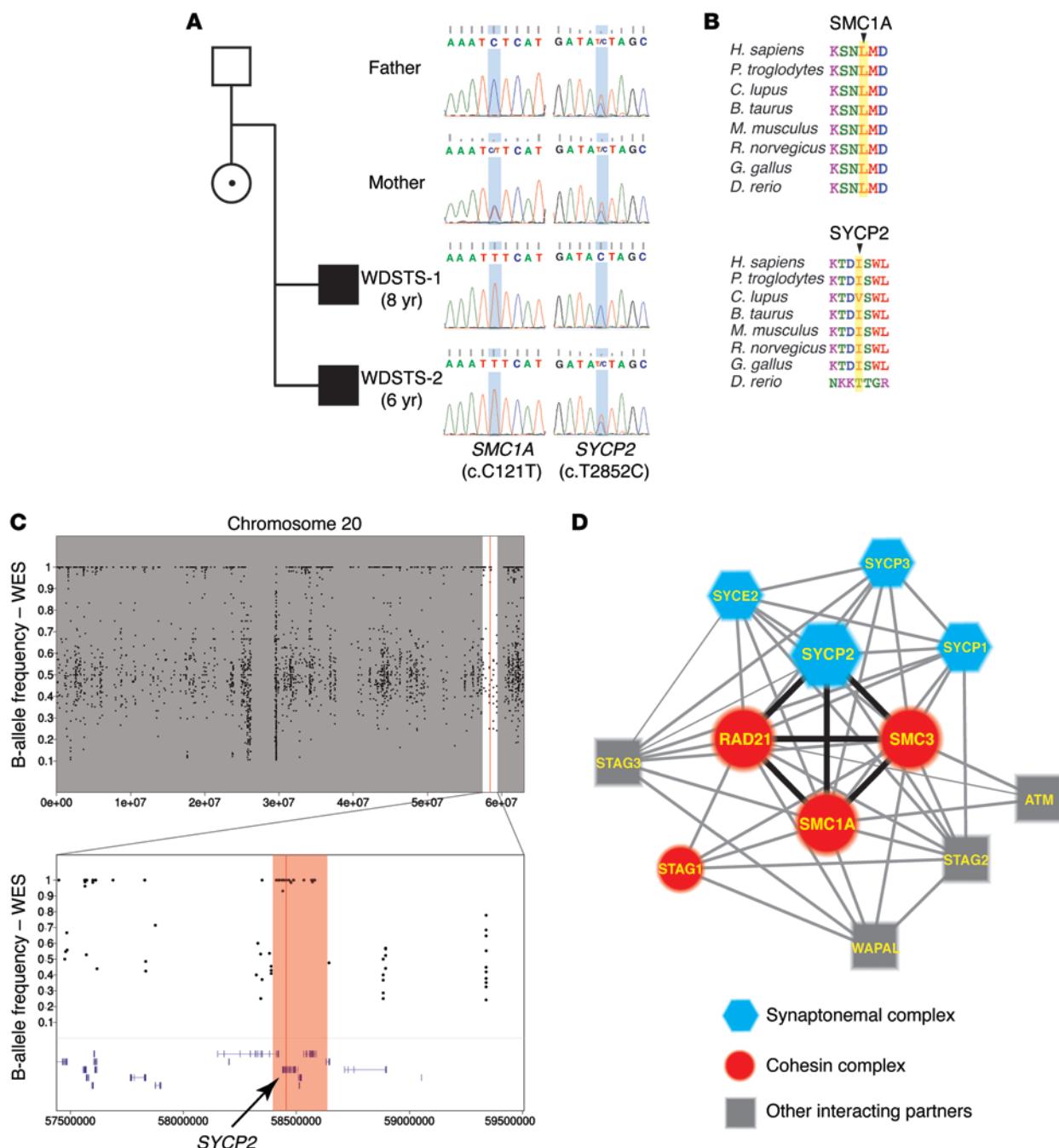
Figure 2. The variants in *SMC1A* and *SYCP2* identified in the patients with WDSTS (WDSTS-1 and WDSTS-2). (**A**) Pedigree of the family and the Sanger-sequencing chromatogram showing the segregation analysis of variants identified in *SMC1A* and *SYCP2*. The blue shading in the chromatograms marks the position of the variants. The genes and nt changes are shown underneath the chromatograms. (**B**) Peptide alignments showing the conservation of the affected aa across different species. First panel: Leu41 in SMC1A. Second panel: Ile951 in SYCP2. All these aa are highlighted by yellow shading, and they are highly conserved across the species. (**C**) The B-allele frequency plots from the WES data of WDSTS-1. The upper panel shows the overall B-allele frequencies of the entire chromosome 20. The lower panel shows the zoomed-in view of the region surrounding the variant (Chr20: g.58455447 A>G [hg19]). The region highlighted by red shading represents the AOH region, including the gene *SYCP2*, which is indicated by the black arrow. (**D**) The interaction network that includes both SC and the cohesin complex. Blue hexagons, major components of the SC; red circles, major components of the cohesin complex; gray squares, other interacting partners involved in this network. Heavy black lines show the strong interactions between the SYCP2 and the 3 major components of the cohesin complex: SMC1A, SMC3, and RAD21.

sister, and parents) and reverse-transcribed into cDNA, in which the alternative splicing product was validated through cDNA PCR followed by cloning and single-colony Sanger sequencing. Two different alternative splicing variants (SV1 and SV2) were

revealed: the SV1 utilized an alternative splicing acceptor inside exon 23 and resulted in an in-frame deletion (c.2536_2547del, p.E846_E849del), while the SV2 utilized the splicing acceptor adjacent to exon 24 and resulted in exon 23 skipping (Figure 3C).

Table 2. Disease-associated variants identified in the study

Patient ID	Clinical diagnosis	Genomic coordinates	nt change	aa change	Gene	Zygosity	Conserved	Predicated	Frequencies		
									ESP	TGP	ARIC
BAB4135 (WDSTS-1)	WDSTS	ChrX: 53442107	c.C121T	p.L41F	*SMC1A*	Hemizygous	+	Damaging	0	0	0
BAB4136 (WDSTS-2)	WDSTS	ChrX: 53442107	c.C121T	p.L41F	*SMC1A*	Hemizygous	+	Damaging	0	0	0
BAB5593 (CdLS-1)	CdLS/WDSTS	Chr10: 112360773	g.GTTACAGGAACT>G	p.E846_E849del	*SMC3*	Heterozygous	+	NA	0	0	0
BAB5452 (CdLS-2)	CdLS/WDSTS	ChrX: 53410176	c.2974-2 A>G	p.D992_Q994del	*SMC1A*	Heterozygous	+	NA	0	0	0
BAB4964 (CdLS-3)	CdLS	Chr11: 118344107	c.C2233T	p.R745X	*KMT2A*	Heterozygous	+	Damaging	0	0	0
BAB4037 (CdLS-4)	CdLS	Chr7: 99711697	c.C136T	p.R46C	*TAF6*	Homozygous	+	Damaging	0	0	0
DG932, DG933, DG936	Syndromic ID	Chr7: 99711522	c.T212C	p.I71T	*TAF6*	Homozygous	+	Damaging	0.008%	0	0

The SV2 could only be detected using the cloning strategy but not in the Sanger sequencing of the cDNA PCR product, indicating that the amount of the SV2 in the peripheral blood of the patients was much less than the SV1 (Figure 3, A and C).

Remarkably, in an 8-month-old female patient (CdLS-2) with a phenotype of excessive hair growth on the elbows, lower back, buttocks, legs, and face, as observed in CdLS-1, we identified a de novo splicing-site mutation (ChrX: g.53410176 T>C [hg19]; c.2974-2A>G) in *SMC1A* (Figure 3B). The assay of cDNA PCR followed by cloning and single-colony Sanger sequencing was also performed for the cDNA derived from the patient's peripheral blood. A splicing variant utilizing an alternative splicing acceptor inside exon 21 was detected, which resulted in an in-frame deletion of 3 aa (p.D992_Q994del) (Figure 3D). In both cases, CdLS-1 and CdLS-2, the splicing-site mutations of *SMC3* and *SMC1A* each

resulted in alternative splicing variants without leading to frameshift outcomes. Concordantly, the splicing machinery in these 2 cases found alternative splicing sites in the downstream exons, which resulted in in-frame deletions that might be less deleterious. It is possible that highly deleterious mutation (i.e., frameshift or nonsense) in *SMC3* and *SMC1A*, which may be incompatible with early embryo development, may be strongly selected against. This observation is also consistent with the fact that the reported mutations in *SMC1A* and *SMC3* are either missense mutations or small in-frame deletions. In addition to the prominent hairy elbows phenotype, the patient CdLS-2 also presented with short stature, prenatal growth retardation, failure to thrive, microcephaly, short neck, bilateral sensorineural hearing loss, hirsutism, mild extremity problems, DD/ID, and dysmorphic facial appearance, including micrognathia, long philtrum, long eyelashes, thick

Table 3. Rare variants in Turkish patients diagnosed with CdLS identified within genes previously reported as CdLS causing

Patient ID	Gene	nt change	aa change	Inheritance	Novel variants
BAB3616	*NIPBL*	c.5211_5212insT	p.Q1737fs	Unknown[A]	Yes
BAB3617	*NIPBL*	c.2147_2148del	p.Q716fs	De novo	Yes
BAB3621	*NIPBL*	c.64+1G>A	Splicing	Unknown[A]	ref. 64
BAB3623	*SMC1A*	c.802_804del	p.K268_K268del	De novo	Yes
BAB3967	*NIPBL*	c.1801delA	p.K601fs	Unknown[A]	Yes
BAB4008	*NIPBL*	c.771+1G>A	Splicing	De novo	Yes
BAB4141	*NIPBL*	c.65-3T>G	Splicing	De novo	Yes
BAB4695	*NIPBL*	c.G4321T	p.V1441L	De novo	ref. 27
BAB4709	*NIPBL*	c.C914A	p.S305X	De novo	Yes
BAB4797	*NIPBL*	c.7900_7937del	p.E2634fs	Unknown[A]	Yes
BAB4938	*HDAC8*	c.C356T	p.T119M	De novo	Yes
BAB5593 (CdLS-1)	*SMC3*	c.2536_2547del	Splicing/p.E846_E849del	De novo	Yes
BAB5452 (CdLS-2)	*SMC1A*	c.2794-2A>G	Splicing/p.D992_Q994del	De novo	Yes
BAB6475	*NIPBL*	c.T7637C	p.L2546P	De novo	Yes

[A]Unknown: parental samples were not available for segregation analyses.
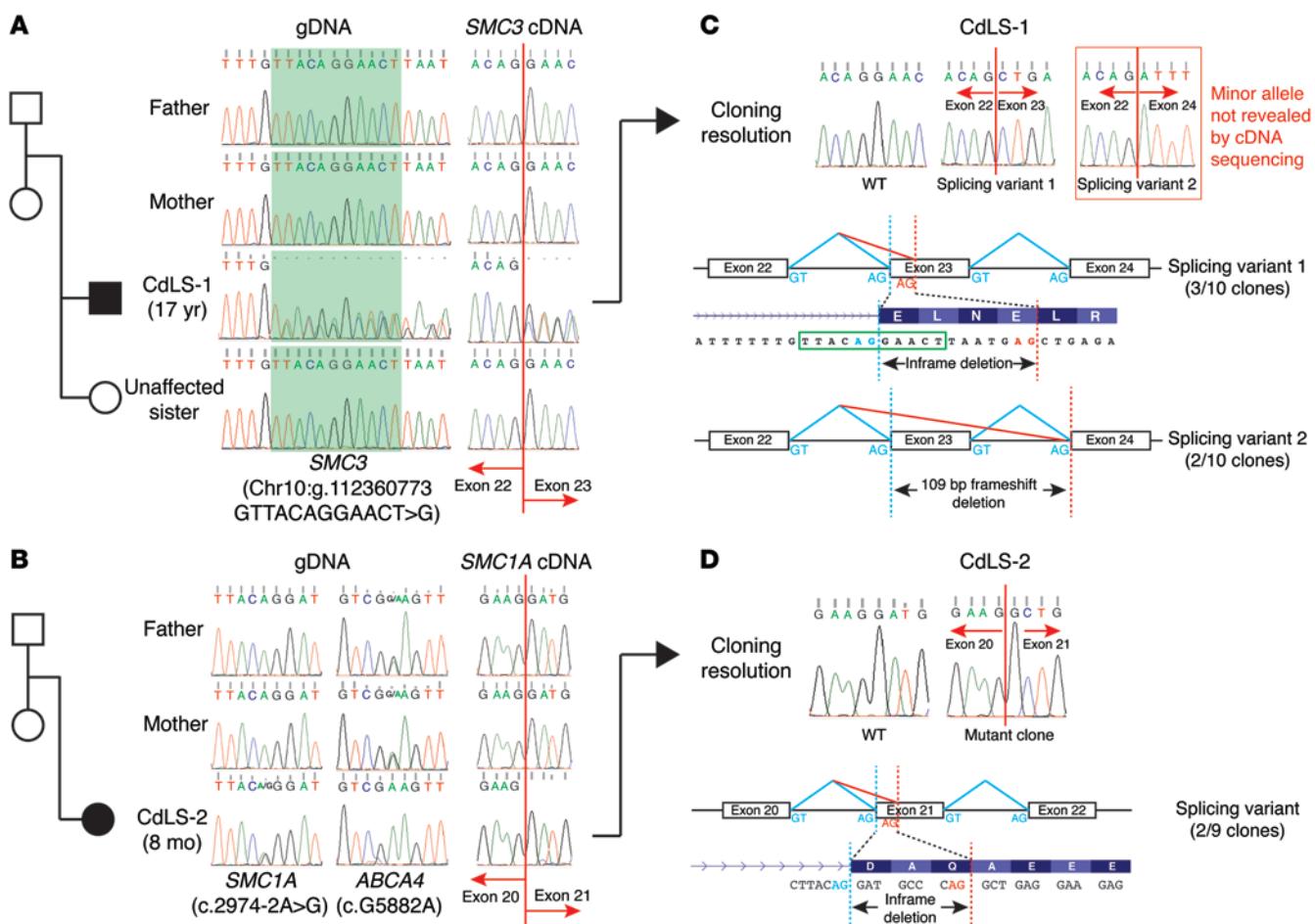
Figure 3. Variants in *SMC3* and *SMC1A* identified in patients with CdLS (CdLS-1 and CdLS-2). (**A** and **B**) Pedigree of the families of CdLS-1 and CdLS-2 and their corresponding Sanger-sequencing chromatograms showing the segregation analyses. The chromatograms on the left show the segregation of the variants in each family. The gene and nt changes are shown underneath the chromatograms. The chromatograms on the right show the outcomes of alternative splicing introduced by the splicing-site variants and their segregations in the family. Red lines in the middle split the 2 adjacent exons. (**C** and **D**) Cloning method resolves the compound alleles in patient-derived cDNA. Different kinds of alternative spliced alleles are shown in the top panel in forms of Sanger-sequencing chromatograms. The outcomes of alternative splicing are shown in the lower panels. Blue dashed lines, site of normal splicing acceptor; red dashed lines, site of alternative splicing acceptors.

eyebrows, synophrys, thin upper lip, and high arched palate (Table 1). She also had retinitis pigmentosa, which could potentially be explained by a homozygous mutation (Chr1: 94473807 C>T [hg19]; c.G5882A, p.G1961E) in *ABCA4* (NM_000350), the most frequently observed disease-associated mutation (~11%) of European ancestry (Figure 3B and ref. 29).

KMT2A and TAF6 variants in patients with CdLS-like phenotypes. The extensive overlapping clinical findings between CdLS and WDSTS together with the finding of mutation in *SMC1A* in WDSTS patients prompted us to postulate that CdLS and WDSTS also share underlying biological mechanisms and rare variants in genes encoding proteins involved in similar genetic pathways underpinning these 2 syndromes. We analyzed the patients in the Turkish CdLS cohort in whom molecular diagnosis was inconclusive from known CdLS-associated genes (total number = 18/32 with no molecular diagnosis) in order to identify potential disease genes. Remarkably, in 1 female patient (CdLS-3) we identified a heterozygous nonsense mutation (Chr11: g.118344107 C>T [hg19]; c.C2233T, p.Arg745Ter) in *KMT2A* (NM_005933), the

causal gene for WDSTS (21). Segregation analysis in the family confirmed that this mutation was not present in the parents or the healthy elder brother, indicating it to be a de novo mutation in the patient (Figure 4). This mutation was not observed in ESP, TGP, and ARIC and was predicted to be damaging by SIFT (30), PolyPhen-2 (31), LRT (32), and MutationTaster (refs. 32, 33, and Table 2).

CdlS-3 was a 21-month-old Turkish female with a healthy 7-year-old brother and was born to nonconsanguineous healthy parents. She was referred to the clinic due to developmental and growth delay. Physical examination was notable for short stature, microcephaly, clinodactyly, hirsutism, DD/ID, and facial dysmorphism, with long philtrum, thin and arched eyebrows, synophrys, long eyelashes, thin upper lip, and high arched palate (Figure 1 and Table 1). Although a variant in the gene associated with WDSTS was identified, she did not appear to have a hairy elbows phenotype, the most prominent and distinctive clinical feature among patients clinically diagnosed with WDSTS; however, she had a facial appearance more consistent with CdLS.
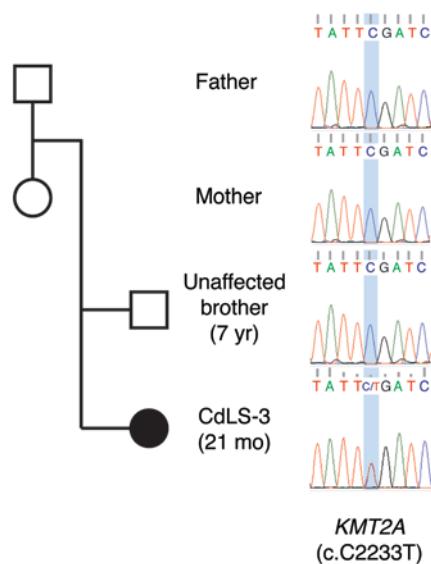
Figure 4. Segregation analysis of the *KMT2A* mutation in the family of CdLS-3. The gene and nt changes are shown underneath the Sanger-sequencing chromatograms.

Among the 32 Turkish patients with a clinical diagnosis of CdLS, 7 patients (21.875%) were born to consanguineous parents, indicating a potential contribution of recessive alleles to CdLS phenotypes. We investigated the personal genomes of these patients for recessive rare and potential pathogenic variants. In 1 patient (CdLS-4), we identified a deleterious homozygous variant (Chr7: g.99711697 G>A [hg19]; c.C136T, p.Arg46Cys) in *TAF6* (TBP-associated factor, where TBP indicates TATA-box binding protein, 80 KD, NM_005641), which was not observed in any control databases and was predicted to be damaging by SIFT, Poly-Phen-2, LRT, and MutationTaster (Table 2).

This subject was a 4-year, 6-month-old male born to consanguineous parents and characterized with typical CdLS clinical features including short stature, microcephaly, narrow shoulders, single transverse palmar crease on the right hand, undescended testicle on the left, hirsutism, DD/ID, and dysmorphic facial appearance, including arched eyebrows with synophrys, long curly eyelashes, a widow's peak, long philtrum, prominent nose, microstomia, and thin upper lip (Figure 5A). Four other children were reported in the family, including an unaffected brother, an unaffected 17-year-old sister with seizures and ID, a sister who died at 9 months of age, and a brother who died at 5 months of age. According to the parents, both the deceased brother and sister were small for gestational age and had a phenotype similar to our subject. Segregation analysis showed that the parents and the unaffected sister were heterozygous carriers for the variant identified, while the unaffected younger brother was homozygous WT (Figure 5B). The variant also occurred in 1 of the regions of absence of heterozygosity (AOH), which likely resulted from identity by descent from the consanguineous marriage (Figure 5B).

To provide further evidence for a potential recessive CdLS or CdLS-like phenotype, we studied exomes of patients in recessive kindreds from a Saudi Arabian clinical population. Independently, a variant in *TAF6* (Chr7: g.99711522 A>G [hg19]; c.T212C,

p.Ile71Thr) was identified in a cohort of multiplex consanguineous Saudi families with syndromic ID. The family consists of first-cousin, once-removed healthy parents with 5 children and 1 miscarriage (Figure 5C). Three of their children (2 sons, DG932/DG936, and 1 daughter, DG933, as shown in Figure 5C) were referred for evaluation because of ID and dysmorphic features. The 3 siblings had hirsutism, broad great toes and Y-shaped 2-3 toe syndactyly, and mild facial dysmorphism in the form of thick bushy eyebrows with synophrys, low frontal hairline, a prominent nose, and long columella. These clinical features are reminiscent of a CdLS-like phenotype. Homozygosity mapping revealed several shared blocks of AOH (Figure 5D). Filtering of the exome variants by the coordinates of these homozygous intervals revealed a single variant in *TAF6* that segregated with the phenotype within the family: the 3 affected individuals were found to be homozygous for the mutation. Consistent with recessive Mendelian inheritance, the unaffected siblings were WT reference, while the parents were both heterozygous carriers (Figure 5D). The variant was absent in a collection of 516 Saudi exomes from King Faisal Specialist Hospital and Research Center and was listed in ESP with an allele frequency of 1:13,000, a very low frequency compatible with a rare variant causing disease in the homozygous state. PolyPhen-2 and SIFT scores also revealed likely pathogenicity (Table 2).

Autosomal-recessive inheritance of CdLS or CdLS-like phenotypes is a previously unreported trait. To investigate the properties of the 2 disease-associated cosegregating mutations identified in the patients, we performed functional assays in *Drosophila* S2 cell lines. *TAF6* is well conserved between human and *Drosophila*, and it stabilizes the core promoter-recognition TFIID complex (34). In addition, the N-terminal region of TAF6 (TAF6N) is sufficient for TFIID stability in S2 cells (34). Given that both mutations locate in the histone fold of the N-terminal region, we tested the interaction ability of the hTAF6N mutant proteins with 4 components in the TFIID complex, including dTAF1, dTAF4, dTAF9, and dTBP. We transiently knocked down the endogenous *Drosophila* TAF6 (*dTaf6*) in S2 cells using dsRNAs targeting the C-terminal region of dTAF6 (*dTaf6C*) and overexpressed either WT human TAF6N (*hTAF6N^{WT}*) or the 2 human TAF6N mutants (*hTAF6N^{R46C}* and *hTAF6N^{I71T}*) individually. Coimmunoprecipitation (co-IP) experiments revealed that the binding affinities between mutated hTAF6N and TFIID components were reduced: hTAF6N^{R46C} moderately lost its interaction with most of the components we tested, while hTAF6N^{I71T} largely reduced the binding ability with all the components. Nevertheless, both hTAF6N mutants lost their interaction with dTBP. Therefore, these 2 mutations disrupted the interactions between TAF6 and other TFIID components, which may lead to the instability of this complex (Figure 5E). To further investigate the potential functional consequences of these 2 *TAF6* mutations, we examined their effect on cell survival/growth in *Saccharomyces cerevisiae*. We transformed the expressing plasmids of *ytaf6^{WT}*, *ytaf6mt1* (corresponds to *hTAF6N^{R46C}*), and *ytaf6mt2* (corresponds to *hTAF6N^{I71T}*) into the yeast *taf6* temperature-sensitive strain. Yeast spot assays were performed at both permissive and restrictive temperatures. There was a slight but not obvious reduction in the growth patterns in mutant transformants (Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI77435DS1). The result was not very surprising,
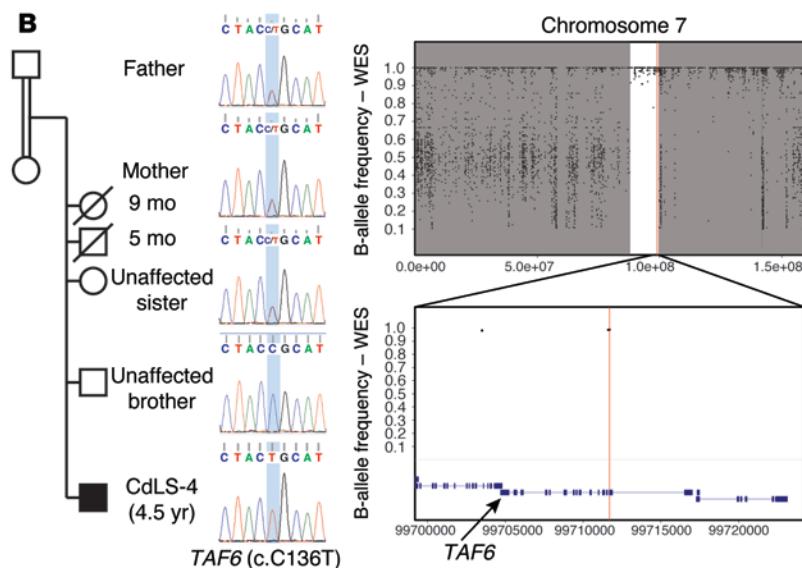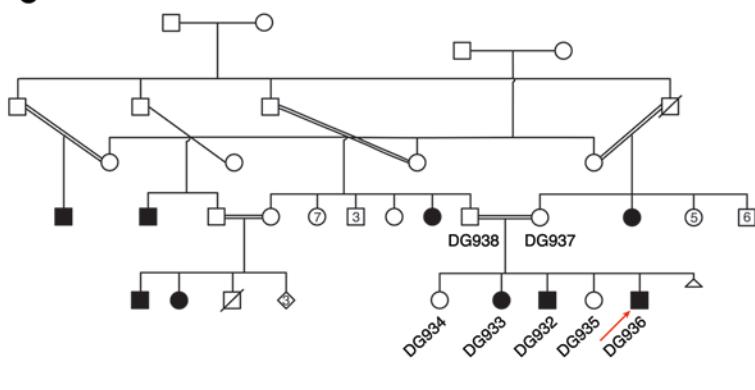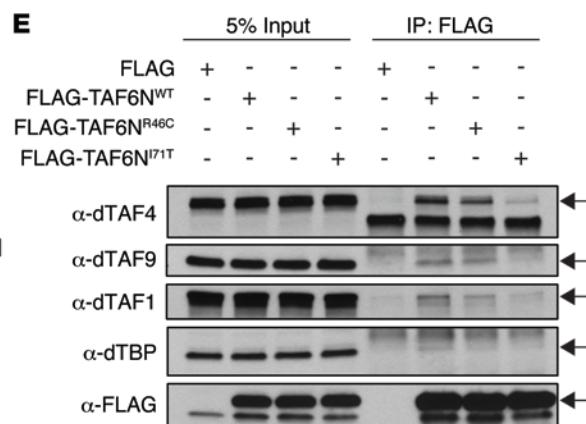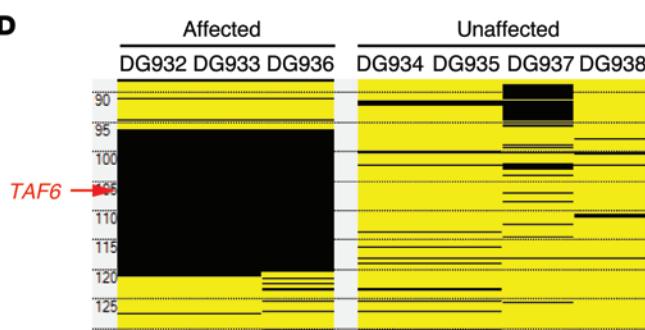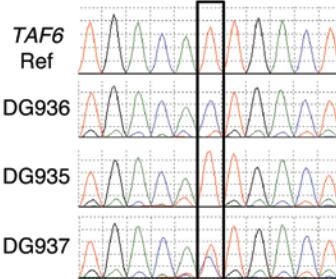
Figure 5. *TAF6* variants identified in the Turkish family of CdLS-4 and the Saudi family. (A) Patient photographs showing the clinical features of CdLS-4 and 2 patients (DG932 and DG936) from the Saudi family. **(B)** The pedigree of the CdLS-4 family and the segregation of the *TAF6* variant. B-allele frequency plots from the WES data for CdLS-4 are shown on the right. The region of AOH is shown as a white region between the flanking gray areas. Top panel: B-allele frequency plot of the entire chromosome 7; bottom panel: zoomed-in view of the region encompassing *TAF6* indicated by the black arrow. **(C)** The pedigree of the Saudi family. Red arrow indicates the patient that underwent WES. **(D)** The homozygosity mapping on the left shows the region of AOH (black blocks) encompassing *TAF6* (indicated by red arrow) in DG932, DG933 and DG936. The Sanger-sequencing chromatograms on the right show homozygous WT in the control and DG935, homozygous mutation in DG936, and heterozygous mutation in the mother, DG937. **(E)** Co-IP in *Drosophila* S2 cells testing TAF6-binding affinity. S2 cells were treated with dsRNA targeting *dTAF6C*. FLAG-TAF6N^WT, FLAG-TAF6N^R46C, and FLAG-TAF6N^I71T correspond to the S2 cells transfected with WT or mutated *dTAF6N* as indicated. FLAG, negative control. Four TFIID components (dTAF1, dTAF4, dTAF9, and dTBP) were detected with Western blot using antibodies against each indicated component. Left 4 columns, input; right 4 columns, proteins co-IP with FLAG-TAF6N. The lanes were run on the same gel, but were noncontiguous.

since the patients with the homozygous variants, although having clinically apparent developmental defects, still survived through development and beyond. It indicates that these mutations may be compatible with life. Thus, this yeast growth assay may not be an optimal and conclusive approach to studying the downstream effects of these 2 perhaps more subtle mutations.

Discussion

SMC1A and SMC3 contribute to phenotypes resembling WDSTS. *SMC1A* is the causal gene for the X-linked CdLS cases presenting with milder phenotypes (5, 6). The identification of an *SMC1A* mutation in a family segregating WDSTS could reflect a potential role for cohesin genes/proteins in the molecular etiology of WDSTS. As one of the key components in the cohesin complex, SMC1A forms a tripartite ring structure with SMC3, RAD21, and stromal antigens (STAGs) to hold the sister chromatids together by trapping them inside the ring (35). SMC1A contains 2 NTPase head domains in the NH$_2$ and COOH termini, respectively, and 1 hinge domain in the middle linked to the 2 terminal head domains by the coiled-coil domains. To date, all the causative mutations reported in CdLS are missense mutations or small in-frame deletions in the NTPase and coiled-coil domains, and the missense mutation found in WDSTS-1 and WDSTS-2 is in the N-terminal NTPase domain (Figure 6A). WDSTS-1 and WDSTS-2 share clinical phenotypes similar to those of other patients with *SMC1A* mutations within the same domain but without the hairy elbows feature (Table 1).

Sharing a structure similar to that of SMC1A, SMC3 forms a V-shaped SMC1A/SMC3 heterodimer in the tripartite ring structure via the interaction between the hinge domains (6). The structural and functional similarities between the gene products of *SMC1A* and *SMC3* imply that genetic variation in the 2 genes may result in similar phenotypes. The striking overlapping WDSTS clinical findings between patient CdLS-1, with an *SMC3* mutation, and patients WDSTS-1/ WDSTS-2, with an identical familial X-linked hemizygous *SMC1A* mutation, not only substantiate this implication, but also strongly correlate the WDSTS phenotypes with CdLS, as both can be caused by defects in cohesin genes.

The phenotypic/genotypic similarity between WDSTS and CdLS is further consolidated by the identification of a de novo *SMC1A* splicing-site mutation in the patient CdLS-2 with combined CdLS and WDSTS clinical features.

It is interesting that we observed different disease expressivity between the 2 brothers harboring the identical mutation in *SMC1A*. The elder brother presented more severe clinical WDSTS features than the younger brother. Specifically, the elder appeared to have more prominent findings of hairy elbows and sacrum (Figure 1). The difference may potentially be due to underlying genetic modifiers that can alter the phenotype caused by the *SMC1A* mutation. We capitalized on the availability of personal genome sequences to examine for a potential modifier gene by comparing the rare variants uniquely observed in each brother. Interestingly, a homozygous missense mutation (Chr20: g.58455447 A>G [hg19]; c.T2852C, p.Ile951Thr) was found in synaptonemal complex protein 2 (*SYCP2*, NM_014258) in the more severely affected elder brother, while it was heterozygous in the mildly affected younger brother and both unaffected parents (Figure 2, A and B, and Supplemental Table 1). The variant is at a low frequency (0.23% in ESP, 0.05% in TGP, and 0.5% in ARIC), highly conserved, and predicted to be deleterious. The homozygous variant also mapped within an AOH region in the elder brother (Figure 2C), indicating a potential distant relationship between the parents. Our in silico investigation of *SYCP2*'s functions and interactions revealed that *SYCP2* encoded a key component of the synaptonemal complex (SC) involved in chromosomal synapsis during meiosis. This component was also shown to act as a close interacting partner with the cohesin complex by strongly interacting with *SMC1A*, *SMC3*, and *RAD21* gene products according to evidence from STRING v9.1 (ref. 36 and Figure 2D). We speculate that defects in *SYCP2* might contribute to the additional and more severe phenotypic findings in the elder brother by epistasis.

This finding remarkably underscores that mutations in potential modifier genes may alter the expressivity of the disease phenotypes caused by the same driver mutation. Mutations in a modifier may either aggravate or mitigate the symptoms attributable to the driver mutation. In the case of the brothers with WDSTS, we speculate that homozygous mutation in *SYCP2* may lead to the more severe phenotype observed in WDSTS-1. The known interactions of SYCP2 with the protein products of CdLS-associating genes are consistent with this epistasis hypothesis (Figure 2D). Moreover, these findings also support the hypothesis that mutational load in an individual may disrupt the biological balance and lead to the occurrence or aggravation of a disease phenotype.

KMT2A potentially expands the genes associated with CdLS. *KMT2A* encodes a histone methyltransferase catalyzing histone H3 lysine 4 methylation (H3K4me), which plays an important role in chromatin-mediated transcriptional regulation during embryogenesis, cell fate determination, and cell-cycle progression as well as stem cell function. To date, all mutations reported in the WDSTS patients are exclusively de novo nonsense/frameshift alleles leading to immediate/subsequent gains of stop codons in the downstream sequences (ref. 21 and Figure 6B). The de novo nonsense mutation identified in CdLS-3 was found in a patient diagnosed with CdLS rather than WDSTS. In this patient, we observed overlapping features of CdLS and WDSTS as well as several unique features of WDSTS. However,

Table 4. Summary of the similarity of gene function and shared phenotypic features among the syndromes discussed in the manuscript

Syndromes	Gene (function)	Hirsutism	Long eyelashes	Eyebrows	Extremities	Congenital heart disease	Structural renal anomalies	Fifth finger clinodactyly	Short stature
CdLS	*NIPBL, HDAC8* (regulator of cohesin complex), *SMC1A, SMC3* and *RAD21* (subunits of cohesin complex)	+	+	Thin, arched	Ectrodactyly, oligodactyly, phocomelia	+	+	+	Prenatal
WDSTS	*KMT2A* (histone H3K4 methyltransferase, regulation of gene expression)	+	+	Thick	brachydactyly	–	–	+	Prenatal
FLHS	*SRCAP* (catalytic component of the multiprotein chromatin-remodeling SRCAP complex, transcriptional coactivator)	+	+	–	Brachydactyly	+/–	+/–	+	Prenatal
RSTS	*CREBBP* (CREB-binding protein, histone acetyltransferase, transcriptional coactivation of gene expression)	+	+	Thick, arched	Broad thumbs and toes	+	–	+	Postnatal
CSS	*SMARCB1, SMARCA4, SMARCA2, ARID1A*, and *ARID1B* (genes encoding subunits of the SWI/SNF, chromatin remodeling)	+	+	Thick	–	+	+	+	Prenatal
KS	*KMT2D* (histone H3K4 methyltransferase, regulation of gene expression), *KDM6A* (histone H3K4me2/3 demethylase, regulation of gene expression)	+	+	Sparse	–	+	+	+	Postnatal
KBGS	*ANKRD11* (inhibits ligand-dependent transcriptional activation by recruiting HDACs to the coactivator/nuclear receptor complex)	–	–	Broad, bushy	Decreased hand length, syndactyly	–	–	+	Postnatal
ATRX	*ATRX* (helicase, regulation of gene expression)	–	–	Medial flare	–	+	+	+	Postnatal

although she had hirsutism, she did not present with hairy elbows and sacrum. CdLS-specific phenotypes were also observed in the patient (Table 1). The blended clinical findings of CdLS with WDSTS in this patient suggest that *KMT2A*, the known gene for WDSTS, may potentially contribute to the phenotypic expansion of CdLS.

Phenotypic features and mutations in KMT2A and TAF6 connect CdLS to transcription machinery defects. It has been noted that there may be phenotypic overlap between WDSTS and Floating-Harbor syndrome (FLHS, MIM # 136140). A report of individuals in a family with hairy elbows but with facial features more consistent with FLHS has been published (37). Additionally, there is a recent case of a boy with a CdLS phenotype and a mutation in *EP300* (38), a gene mutated in about 3% of individuals with Rubinstein-Taybi syndrome (RSTS, MIM # 180849), which is associated with milder abnormalities of the digits. Finally, the phenotypic overlap between FLHS and RSTS has long been noted (39, 40), which is not surprising, given that the gene that causes FLHS, *SRCAP*, is a coactivator of *CREBBP*, the gene mutated in the majority of individuals with RSTS. When these disorders are thought of together, shared phenotypic features begin to emerge, including long eyelashes, hirsutism, abnormal eyebrows, limb and digit abnormalities, and growth deficiency (Table 4). These

shared and relatively unique phenotypic features would suggest a shared pathophysiology among these disorders.

Our identification of potential causative mutations in *KMT2A* and *TAF6* emphasizes that genes involved in epigenetic modification, chromatin remodeling, and transcriptional regulation pathways may be associated with CdLS. X-linked α-thalassemia/mental retardation syndrome (ATRX, MIM # 301040), KBG syndrome (KBGS, MIM # 148050), Kabuki syndrome (KS, MIM # 147920, MIM # 300867), RSTS, FLHS, Coffin-Siris syndrome (CSS, MIM # 135900), and WDSTS all share phenotypic features overlapping with those observed in CdLS (Table 4). Interestingly, the genes associated with these syndromes are all involved in epigenetic modification, chromatin remodeling, and transcriptional regulation pathways: ATRX-associated gene *ATRX* encodes a helicase modifying chromatin structure and regulating the transcription of downstream genes (41, 42); *ANKRD11*, mutations of which can result in KBGS, encodes a negative regulator of ligand-dependent transcriptional activation (43); KS-associated genes *KMT2D* and *KDM6A* encode interacting histone H3K4 methyltransferase and H3K27 demethylase, respectively (44, 45); *CREBBP* (associated with RSTS) and *SRCAP* (associated with FLHS) encode interact-

Figure 6. Summaries of the damaging, potentially disease-associated mutations identified in this study and those reported previously. (**A**) Summary of *SMC1A* mutations reported in patients with CdLS (in black) and mutations reported in this study (in red). The domains of SMC1A are illustrated and annotated in the diagram. (**B**) Summary of *KMT2A* mutations reported in patients with WDSTS (in black) and the mutation reported in this study (in red). The domains of KMT2A are illustrated, with each annotated underneath the diagram.

ing partners, both of which are involved in transcriptional coactivation (46, 47); mutations in genes encoding subunits of SWI/SNF complex, which encompasses the protein product of *SRCAP* (48), have also been identified in patients with CSS (49). More interestingly, in a recently published study of a cohort of patients with CdLS or CdLS-like phenotypes, de novo mutations in the KBGS-associated gene *ANKRD11* were identified, demonstrating the phenotypic overlap between KBGS and CdLS (50). From the comparisons among different genes, gene functions, and clinical features (Table 4), a picture of perturbed transcriptional regulation–associated phenotypes, perhaps a "transcriptomopathy," begins to emerge. In aggregate, the evidence above suggest that different syndromes that share phenotypic overlaps may harbor highly penetrant Mendelizing genetic variants involved in similar molecular pathways.

Defects in components of cohesin complex, especially those in NIPBL, are to date the major causes of CdLS. However, in addition to its crucial role in chromosome segregation, cohesin is also strongly suggested as regulating transcription. NIPBL is reported to be a regulator of gene expression in either cohesin-dependent or -independent manners (15). Moreover, by using FISH in cells from patients with CdLS harboring *NIPBL* mutations, Nolen et al. have shown that defects in NIPBL result in the regional decompaction of higher-order chromatin architecture occurring at gene-rich genomic intervals, suggesting NIPBL's association with the transcription machinery (13). A recent study by Lopez-Serra et al. shows in yeast that Scc2 (yeast orthologue of human NIPBL) is intimately

associated with the RSC complex (mutations of its components in human may result in CSS) in nucleosome positioning, chromatin remodeling, and transcriptional regulation (51). These functional molecular studies correlate with our clinical and genomic analysis in patients with CdLS and CdLS-like features, where we identified variants in proteins involved in transcriptional regulation.

Biochemical data obtained from studies in *Drosophila* reveal interactions between RAD21, one of the major structural components of cohesin complex, and the trithorax-group proteins; these link the cohesin components and chromatin-remodeling proteins into the same pathway of transcriptional regulation (refs. 52, 53, and Figure 7). In addition, genome-wide studies suggest the association between cohesin and RNA polymerase II (Pol II), which is also the interacting partner of the TFIID complex (ref. 17 and Figure 7). TFIID is a megadalton-sized multiprotein complex necessary for RNA Pol II–initiated transcription, which reads epigenetic markers or interacts with transcriptional activators. It consists of TBP and TAFs (54), which include TAF6 as a major component. TAF6 is conserved from yeast to human. In *Drosophila*, TAF6 together with other TAFs (TAF4, TAF5, TAF9, and TAF12) form the functional core of TFIID, and RNAi knockdown targeting TAF6 shows its role in stabilizing the TFIID complex (55). It is not immediately apparent why perturbation of a component of a ubiquitously expressed general transcription factor complex may result in a tissue-specific phenotype, such as CdLS. However, we do note that knockdown of TAF6 in *Drosophila* resulted in a phenotype restricted to the wing area, suggesting a morphogenetic function of TAF6

Figure 7. Diagram summarizing interaction network potentially encompassing all the proteins encoded by the genes identified in the study. Three categories of cellular functions (sister chromatid cohesion, chromatin remodeling, and transcriptional regulation) are shown. The proteins are illustrated in colored spheres. NIPBL, SMC1A, SMC3, and RAD21 are the 4 components in cohesin, which is involved in sister chromatid cohesion, chromatin remodeling, and transcriptional regulation. The interaction between trithorax-group proteins and the RAD21 links KMT2A into this interaction network. RNA Pol II interacts with both TFIID and cohesin and then potentially links TAF6, a core component of TFIID complex, to the cohesin.

(56). The defects in TAF6 may result in perturbed transcriptions initiation by RNA Pol II. From this perspective of transcriptional regulation, it is intriguing that we identified recessive *TAF6* mutations in 2 separate world populations enriched for recessive disease–associated variants due to a high rate of consanguinity.

As the cohesin may participate in the same pathway as the genes involved in transcriptional regulation, the identification of deleterious mutations in *TAF6* strongly supports the hypothesis that defects of genes involved in the transcription machinery, such as complexes with function of chromatin structure remodeling and transcriptional regulation, may result in a spectrum of disorders resembling CdLS. The aggregate evidence above suggest that these genes may contribute to phenotypes with characteristic core clinical features (Table 4) that may reflect a global disturbance in transcriptional regulation or a class of disorders that could constitute a "transcriptomopathy."

In summary, we identified a deleterious mutation in *SMC1A*, a causal gene of CdLS, in 2 related Turkish patients with a clinical diagnosis of WDSTS. Reciprocally, from a cohort of 32 Turkish patients with clinical diagnoses of CdLS, we also identified a deleterious mutation in *KMT2A*, the causal gene of WDSTS, in 1 CdLS patient. Moreover, in the CdLS cohort, we identified 1 deleterious mutation in *SMC3* and 1 in *SMC1A* in 2 independent patients with clinical features resembling those seen in both CdLS and WDSTS. Since CdLS is a heterogeneous disorder with limited molecular etiologies, additional genetic variations await to be uncovered. In this study, we emphasize the clinical overlap of CdLS not only with

other cohesinopathies, but also with disorders such as RSTS and FLHS that are caused by mutations in genes related to transcription. We therefore assert the importance of the contribution of transcription machinery to the phenotypes and molecular etiologies of CdLS-like phenotypes. Given a selected patient population with a high rate of consanguinity, potential recessive genes may also be identified. We propose that *TAF6* may act as a potential responsible gene for autosomal-recessive CdLS.

Methods

Human subjects. This study was approved by the Institutional Review Board at BCM and King Faisal Specialist Hospital and Research Center. Informed consent was obtained from all subjects prior to enrollment in the project. All subjects were evaluated by one or more clinical geneticists. Genomic DNA was extracted from whole blood using the Gentra Puregene Blood Extraction Kit per the manufacturer's protocol (QIAGEN). All the genomic studies, including aCGH and WES, were performed on DNA samples extracted from whole blood.

CdLS array CGH. We initially performed our in-house CdLS high-density aCGH targeting of cohesin/condensin genes (7) in all patients with clinical diagnoses of CdLS and WDSTS prior to submission to the WES pipeline to exclude potential pathogenic CNVs. To summarize the array design, an Agilent 8X60K format was used to include 46 genes (Supplemental Table 2) that belong to cohesin/condensin complex families and their upstream/downstream 50 kb genomic regions. The array was designed using the Agilent eArray website (https://earray.chem.agilent.com/earray/), with average probe density of about 5 probes per 1 kb. All array procedures including DNA fragmentation and labeling, array hybridization, washing, and scanning were performed according to the manufacturer's instructions and previously described protocols (57) with minor modifications. Sex-matched female control (NA15510) DNA and male control (NA10851) DNA were obtained from Coriell cell repositories (Coriell Institute for Medical Research; http://ccr.coriell.org). Data processing and analyses were done using Agilent Feature Extraction Software (version 11.5, Agilent Technologies) and Agilent Genomic Workbench (edition 7.0, Agilent Technologies).

CMA. Agilent customized whole-genome array (BCM CMA version 9.1.1) was performed on patients WDSTS-1 and WDSTS-2 (24, 58, 59). Since *KMT2A* is not targeted on the CdLS aCGH as a candidate gene, the CMA could investigate any potential pathogenic CNVs in the *KMT2A* locus and the entire genome. The array design and experimental procedures were described previously (24, 57).

WES and data analysis. Thirty-two Turkish patients with CdLS and 2 patients with WDSTS (WDSTS-1 and WDSTS-2) underwent WES at the BCM Human Genome Sequencing Center (BCM-HGSC) through the Baylor-Hopkins Center for Mendelian Genomics research initiative.

Genomic DNA samples were constructed into Illumina paired-end precapture libraries according to the manufacturer's protocol (Illumina Multiplexing Sample Preparation Guide 1005361 D) with modifications as described in the BCM-HGSC protocol (https://www.hgsc.bcm.edu/content/protocols-sequencing-library-construction). Libraries were prepared using Beckman robotic workstations (Biomek NXp and FXp models). Briefly, from 0.5 to 1 μg of DNA was sheared into fragments of approximately 300-400 base pairs with the Covaris E210 system following end-repair, A-tailing, and ligation of the Illumi-

na multiplexing PE adaptors. Precapture ligation-mediated PCR (LM-PCR) was performed for 6 to 9 cycles of amplification using either the 2X SOLiD Library High Fidelity Amplification Mix (a custom product manufactured by Invitrogen), the Phusion High-Fidelity PCR Master Mix (NEB, M0531L) or KAPA HiFi DNA Polymerase (Kapa Biosystems Inc., KM2602). Purification was performed with Agencourt AMPure XP beads (Beckman, A63882) after enzymatic reactions, and, following the final purification, quantification and size distribution of the precapture LM-PCR product were determined using the LabChip GX electrophoresis system (PerkinElmer).

In preparation for cocapture, the uniquely barcoded precapture libraries were pooled in equimolar amounts into 4- or 6-plexes using 1 μg per pool (for libraries prepared with the Invitrogen and KAPA enzymes) or 2 μg per pool (for libraries prepared with the Phusion enzyme). Each 4-plex library pool was then hybridized in solution to the HGSC-designed core capture reagent (52 Mb, NimbleGen) according to the manufacturer's protocol (NimbleGen SeqCap EZ Exome Library SR User's Guide) with minor revisions, while 6-plex library pools used the custom VCRome 2.1 capture reagent (42 Mb, NimbleGen). Human COT1 DNA and full-length blocking oligonucleotides were added into the hybridization to block repetitive genomic sequences and adaptor sequences. Postcapture LM-PCR amplification was performed for 12 to 14 cycles of amplification with the enzyme used in precapture library preparation. After the final AMPure XP bead purification, quantity and size of the capture library were analyzed using the Agilent Bioanalyzer 2100 DNA Chip 7500. The efficiency of the capture was evaluated by performing a qPCR-based quality assay on the 4 standard NimbleGen internal controls. Successful enrichment of the capture libraries was estimated to range from a 6 to 9 of ΔCt value over the nonenriched samples.

Library templates were prepared for sequencing using Illumina's cBot Cluster Generation System with TruSeq PE Cluster Generation Kits (PE-401-3001) according to the manufacturer's protocol. Briefly, these libraries were denatured and diluted in hybridization buffer in order to achieve a load density of approximately 800 K clusters/mm. Each library pool was loaded in a single lane of a HiSeq flow cell, and each lane was spiked with 2% phiX control library for run quality control. The sample libraries then underwent bridge amplification to form clonal clusters, followed by hybridization with the sequencing primer. Sequencing runs were performed in paired-end mode using the Illumina HiSeq 2000 platform. Using the TruSeq SBS Kits (FC-401-3001), sequencing-by-synthesis reactions were extended for 101 cycles from each end, with an additional 7 cycles for the index read. With the sequencing run yielding an average of approximately 9.6 Gb per sample, samples achieved an average of 92% of the targeted exome bases covered to a depth of 20× or greater.

Initial sequence analysis was performed using the HGSC Mercury analysis pipeline (https://www.hgsc.bcm.edu/software/mercury) (60). In summary, the bcl files produced on-instrument were first transferred into the HGSC analysis infrastructure by the HiSeq Real-time Analysis Module. Mercury then ran the vendor's primary analysis software (CASAVA) to de-multiplex pooled samples and generate sequence reads and base-call confidence values (qualities), followed by the mapping of reads to the GRCh37 Human reference genome (Genome Reference Consortium; http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/) using the Burrows-Wheeler aligner (BWA; http://bio-bwa.sourceforge.net/). The resulting binary

alignment/map (BAM) file underwent quality recalibration using GATK (http://www.broadinstitute.org/gatk/) and, where necessary, the merging of separate sequence-event BAMs into a single sample-level BAM. BAM sorting, duplicate read marking, and realignment to improve insertion/deletion (indel) discovery all occur at this step.

WES for the Saudi family was performed using the TruSeq Exome Enrichment kit (Illumina) following the manufacturer's protocol. Samples were prepared for the Illumina sequencing library, with enrichment for the desired targeting using the Illumina Exome Enrichment protocol. The captured libraries were sequenced using Illumina HiSeq 2000 Sequencer, and reads were mapped against UCSC hg19 by BWA. Single-nt polymorphisms and indels were detected by SAMTOOLS. The resulting 77814 variants were filtered according to homozygosity, predicted pathogenicity and location within the autozygome (61).

The findings of WES studies were deposited into the NCBI's ClinVar archive (SCV000195835, SCV000195836, SCV000195837, SCV000195838, SCV000195839, SCV000195840, SCV000195841, SCV000195842, SCV000195843, SCV000195844, SCV000195845, SCV000195846, SCV000195847, SCV000195848, SCV000195849, SCV000195850, SCV000195851, SCV000195852).

PCR validation and segregation analyses. The variants of interest were confirmed by Sanger sequencing following PCR amplification, and segregation analyses were performed on all available members of the families. Same pairs of primers are utilized for both PCR amplification and Sanger sequencing. Primer sets and their corresponding families are listed in Supplemental Table 3. PCR reactions are performed following the protocols for HotStarTaq DNA Polymerase (QIAGEN).

B-allele frequency calculation and homozygosity mapping. To detect regions with AOH, we used B-allele frequency data obtained from both cSNP arrays and WES. B-allele frequency from WES was determined by computing the variant/total reads ratio for each SNP present in VCF file. The regions with AOH were identified using in-house scripts in R language (www.r-project.org) and the circular binary segmentation algorithm (62). Alternatively, genotyping data from the Axiom SNP Chip platform (Affymetrix) was utilized to calculate AOH using AutoSNPa (63).

Splicing-site mutation analysis. Blood samples for splicing-site mutation analysis were collected using the PAXgene blood RNA tube (PreAnalytiX). Total RNA was extracted using PAXgene blood RNA kit (PreAnalytiX) by following the manufacturer's protocol (http://www.qiagen.com/resources/resourcedetail?id=dc649975-b6ba-43e7-a3e9-e9986c612d2e&lang=en). cDNA was obtained using SuperScript III First-Strand Synthesis System for RT-PCR kit (Invitrogen) by following the manufacturer's protocol (http://tools.lifetechnologies.com/content/sfs/manuals/superscriptIIIfirststrand_pps.pdf). Primers used for PCR amplification of the normal alleles and alternatively spliced alleles are listed in Supplemental Table 4. Sanger sequencings were performed to verify the alternative splicing site.

Cloning was performed to resolve the heterozygous alleles from PCR amplification. PCR amplification product was ligated into *pCR4-TOPO* vector using the TOPO TA Cloning Kit (Invitrogen) and transformed into One Shot TOP10 competent cells (Invitrogen). Cells after transformation were plated onto LB-agar plates supplemented with 50 μg/ml of kanamycin. Single colonies were picked from the plates after 37°C overnight culture and inoculated into LB-liquid medium with kanamycin for 37°C overnight culture in suspension. The plasmids

were extracted from each cell culture using QIAprep Spin Miniprep Kit (QIAGEN), and M12-forward and M12-reverse primers (Invitrogen) were used for Sanger sequencing.

Constructs for TAF6 mutation used in functional assays. For constructs used in co-IP experiments, WT or mutated N-terminal regions of hTAF6 (*hTAF6N^{WT}*, *hTAF6N^{R46C}*, and *hTAF6N^{I71T}*) were cloned into *pUASTattB* vectors with an N-terminal FLAG tag. For making *dTaf6* dsRNA to knock down endogenous *Drosophila Taf6*, the C-terminal region of *dTaf6* was cloned into the *pUASTattB* vector and used as a template for in vitro transcription using the MEGAscript RNAi Kit (Life Technologies). For constructs used in yeast experiments, 1 kb upstream and downstream sequences of *ytaf6* were cloned into the *pRS316* vector as regulatory regions. Full-length transcripts of WT *ytaf6* and mutant *ytaf6* with the corresponding mutations in humans (*ytaf6mt1* and *ytaf6mt2*) were also cloned into *pRS316* vectors. Cloning and DNA purification were performed based on standard protocols. Enzymes were from NEB, and DNA purification kits were from Invitrogen and QIAGEN. All constructs were Sanger sequenced before performing experiments.

Cell culture, co-IP, and Western blot. The *Drosophila* S2 cell line was cultured in Schneider's media (Lonza) with 10% FBS (Thermo Scientific) and antibiotics mix (Gibco; Life Technologies) at room temperature. The transfection was performed using Effectene (QIAGEN) according to the manufacturer's instructions. Proteins were expressed using the Actin-GAL4/UAS system. In co-IP experiments, cells were treated with dsRNA against the C-terminal region of *dTAF6* and then transfected with *pUASTattB-hTAF6N* constructs. Cells were harvested 48 hours after transfection and lysed on ice in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5% glycerol, 0.5% Triton X-100, and EDTA-free cocktail complete protease inhibitor; Roche). Lysate supernatant was incubated with 1% BSA preblocked EZ red α-HA beads (Sigma-Aldrich) overnight. Beads were washed with lysis buffer and boiled in 2× Laemmli buffer. SDS-PAGE, transfer, and Western blot were performed according to standard protocols. Primary antibodies were used at the following dilutions: rabbit α-FLAG antibody, 1:2000 (Abcam, ab1162); mouse α-dTAF4 antibody, 1:1000 (34); mouse α-dTAF9 antibody, 1:100 (34); mouse α-dTAF1 antibody, 1:100 (34); and mouse α-dTBP antibody, 1:100 (34). Goat HRP-conjugated secondary antibodies were used at 1:3000 dilution (Jackson ImmunoResearch, 115-035-166, 111-035-144). Membranes were developed using Western Lightning Plus-ECL (PerkinElmer) followed by x-ray film (Genemate) detection.

Yeast spot assay. A *pRS316-taf6* series of constructs was transformed into *S. cerevisiae* strain *YSB 552* (*MATa his3D200 leu2D1 or leu2-3,112 ura3-52 taf6D::LEU2* [*pRS313* (*CEN HIS3 taf6-19*)]) according to standard protocols. Nutrition restrictions on uracil, histidine, and leucine were used to maintain the transformed plasmid and the original *ytaf6* temperature-sensitive construct in *YSB 552*. For dilution spot assays, yeast cells grown to logarithmic phase at room temperature were initially diluted to 7 OD/ml. Subsequent 10-fold dilutions were made and spotted using the Replica plater. Plates were incubated at room temperature and at 37°C. In this experiment, *pRS316* empty vector was used as a negative control.

Study approval. This study was approved by the Institutional Review Board at BCM (H-29697) and the King Faisal Specialist Hospital and Research Center (RAC# 2080006). Informed consent was obtained from all subjects prior to enrollment in the project. Written informed consent to use the photographs of patients appearing in the figures was also obtained.

Address correspondence to: James R. Lupski, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 604B, Houston, Texas 77030, USA. Phone: 713.798.3723; E-mail: jlupski@bcm.edu.

1. Brachmann W. Ein fall von symmetrischer monodaktylie durch Ulnadefekt, mit symmetrischer flughautbildung in den ellenbeugen, sowie anderen abnormitaten (zwerghaftigkeit, halsrippen, behaarung). *Jarb Kinder Phys Erzie*. 1916;84:225–235.

2. de Lange C. Sur un type nouveau de degenerescence (typus Amstelodamensis). *Arch Med Enfants*. 1933;36:713–719.

3. Tonkin ET, Wang TJ, Lisgo S, Bamshad MJ, Strachan T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat Genet*. 2004;36(6):636–641.

4. Krantz ID, et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat Genet*. 2004;36(6):631–635.

5. Musio A, et al. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat Genet*. 2006;38(5):528–530.

6. Deardorff MA, 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(3):485–494.

7. Pehlivan D, et al. NIPBL rearrangements in Cornelia de Lange syndrome: evidence for replicative mechanism and genotype-phenotype correlation. *Genet Med*. 2012;14(3):313–322.

8. Russo S, et al. Intragenic and large NIPBL rearrangements revealed by MLPA in Cornelia de Lange patients. *Eur J Hum Genet*. 2012;20(7):734–741.

9. Cheng YW, et al. Copy number analysis of NIPBL in a cohort of 510 patients reveals rare copy number variants and a mosaic deletion. *Mol Genet Genomic Med*. 2014;2(2):115–123.

10. Huisman SA, Redeker EJ, Maas SM, Mannens MM, Hennekam RC. High rate of mosaicism in individuals with Cornelia de Lange syndrome. *J Med Genet*. 2013;50(5):339–344.

11. Deardorff MA, et al. RAD21 mutations cause a human cohesinopathy. *Am J Hum Genet*. 2012;90(6):1014–1027.

12. Deardorff MA, et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature*. 2012;489(7415):313–317.

13. Nolen LD, Boyle S, Ansari M, Pritchard E, Bickmore WA. Regional chromatin decompaction in Cornelia de Lange syndrome associ-

ated with NIPBL disruption can be uncoupled from cohesin and CTCF. *Hum Mol Genet*. 2013;22(20):4180–4193.

14. Dorsett D. Cohesin: genomic insights into controlling gene transcription and development. *Curr Opin Genet Dev*. 2011;21(2):199–206.

15. Zuin J, et al. A cohesin-independent role for NIPBL at promoters provides insights in CdLS. *PLoS Genet*. 2014;10(2):e1004153.

16. Dowen JM, et al. Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. *Stem Cell Reports*. 2013;1(5):371–378.

17. Schaaf CA, et al. Genome-wide control of RNA polymerase II activity by cohesin. *PLoS Genet*. 2013;9(3):e1003382.

18. Vega H, et al. Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. *Nat Genet*. 2005;37(5):468–470.

19. van der Lelij P, et al. Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1. *Am J Hum Genet*. 2010;86(2):262–266.

20. Wiedemann HR, Kunze J, Grosse FR, Dibbern H. A syndrome of abnormal facies, short stature, and psychomotor retardation. In: *Atlas of Clinical Syndromes: A Visual Aid to Diagnosis for Clinicians and Practicing Physicians*. 2nd ed. London, United Kingdom: Wolfe Publishing Ltd; 1989:198–199.

21. Jones WD, et al. De novo mutations in MLL cause Wiedemann-Steiner syndrome. *Am J Hum Genet*. 2012;91(2):358–364.

22. Polizzi A, Pavone P, Ciancio E, La Rosa C, Sorge G, Ruggieri M. Hypertrichosis cubiti (hairy elbow syndrome): a clue to a malformation syndrome. *J Pediatr Endocrinol Metab*. 2005;18(10):1019–1025.

23. Visser R, Beemer FA, Veenhoven RH, De Nef JJ. Hypertrichosis cubiti: two new cases and a review of the literature. *Genet Couns*. 2002;13(4):397–403.

24. Wiszniewska J, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet*. 2014;22(1):79–87.

25. Gillis LA, 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(4):610–623.

26. Yan J, et al. Mutational and genotype-phenotype correlation analyses in 28 Polish patients with Cornelia de Lange syndrome. *Am J Med Genet A*. 2006;140(14):1531–1541.

27. Pie J, et al. Mutations and variants in the cohesion factor genes NIPBL, SMC1A, and SMC3 in a cohort of 30 unrelated patients with Cornelia de Lange syndrome. *Am J Med Genet A*. 2010;152A(4):924–929.

28. Bhuiyan ZA, et al. Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. *J Med Genet*. 2006;43(7):568–575.

29. Zernant J, et al. Genetic and clinical analysis of ABCA4-associated disease in African American patients. *Hum Mutat*. 2014;35(10):1187–1194.

30. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–1081.

31. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter*. 2013;Chapter 7:Unit7.20.

32. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res*. 2009;19(9):1553–1561.

33. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575–576.

34. Wright KJ, Marr MT, Marr MT 2nd, Tjian R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A*. 2006;103(33):12347–12352.

35. Haering CH, Farcas AM, Arumugam P, Metson J, Nasmyth K. The cohesin ring concatenates sister DNA molecules. *Nature*. 2008;454(7202):297–301.

36. Franceschini A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013;41:D808–D815.

37. Koc A, Karaer K, Ergun MA, Cinaz P, Percin EF. A new case of hairy elbows syndrome (hypertrichosis cubiti). *Genet Couns*. 2007;18(3):325–330.

38. Woods SA, Robinson HB, Kohler LJ, Agamanolis D, Sterbenz G, Khalifa M. Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap Cornelia de Lange syndrome. *Am J Med Genet A*. 2014;164A(1):251–258.

39. Stevens CA. Rubinstein-Taybi syndrome. In: Pagon RA, et al., eds. *GeneReviews* [Internet]. Seattle, Washington, USA; 1993.

40. Nowaczyk MJM, Nikkel SM, White SM. Floating-Harbor syndrome. In: Pagon RA, et al., eds. *GeneReviews* [Internet]. Seattle, Washington, USA; 1993.

41. Gibbons RJ, Picketts DJ, Villard L, Higgs DR. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell*. 1995;80(6):837–845.

42. Law MJ, et al. ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell*. 2010;143(3):367–378.

43. Zhang A, et al. Identification of a novel family of ankyrin repeats containing cofactors for p160 nuclear receptor coactivators. *J Biol Chem*. 2004;279(32):33799–33805.

44. Ng SB, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010;42(9):790–793.

45. Lederer D, et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *Am J Hum Genet*. 2012;90(1):119–124.

46. Petrij F, et al. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature*. 1995;376(6538):348–351.

47. Hood RL, et al. Mutations in SRCAP, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome. *Am J Hum Genet*. 2012;90(2):308–313.

48. Monroy MA, Schott NM, Cox L, Chen JD, Ruh M, Chrivia JC. SNF2-related CBP activator protein (SRCAP) functions as a coactivator of steroid receptor-mediated transcription through synergistic interactions with CARM-1 and GRIP-1. *Mol Endocrinol*. 2003;17(12):2519–2528.

49. Tsurusaki Y, et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. *Nat Genet*. 2012;44(4):376–378.

50. Ansari M, 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(10):659–668.

51. Lopez-Serra L, Kelly G, Patel H, Stewart A, Uhlmann F. The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. *Nat Genet*. 2014;46(10):1147–1151.

52. Hallson G, et al. The Drosophila cohesin subunit Rad21 is a trithorax group (trxG) protein. *Proc Natl Acad Sci U S A*. 2008;105(34):12405–12410.

53. Strubbe G, 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(14):5572–5577.

54. Bieniossek C, et al. The architecture of human general transcription factor TFIID core complex. *Nature*. 2013;493(7434):699–702.

55. Wright KJ, Marr MT, Tjian R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A*. 2006;103(33):12347–12352.

56. Umemori M, et al. RNAi-mediated knockdown showing impaired cell survival in Drosophila wing imaginal disc. *Gene Regul Syst Bio*. 2009;3:11–20.

57. Shinawi M, et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet*. 2010;47(5):332–341.

58. Cheung SW, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med*. 2005;7(6):422–432.

59. Boone PM, et al. Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat*. 2010;31(12):1326–1342.

60. Reid JG, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. *BMC Bioinformatics*. 2014;15:30.

61. Alkuraya FS. The application of next-generation sequencing in the autozygosity mapping of human recessive diseases. *Hum Genet*. 2013;132(11):1197–1211.

62. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004;5(4):557–572.

63. Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT. Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. *Hum Mutat*. 2006;27(10):1041–1046.

64. Borck G, et al. NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. *J Med Genet*. 2004;41(12):e128.