Published in final edited form as:

Hum Mutat. 2009 November ; 30(11): 1535–1542. doi:10.1002/humu.21095.

SMC1A Expression and Mechanism of Pathogenicity in Probands with X-linked Cornelia de Lange Syndrome

Jinglan Liu¹, Rachel Feldman¹, Zhe Zhang², Matthew A. Deardorff^{1,3}, Eden V. Haverfield⁴, Maninder Kaur¹, Jennifer R. Li¹, Dinah Clark¹, Antonie D. Kline⁵, Darrel J. Waggoner⁴, Soma Das⁴, Laird G. Jackson⁶, and Ian D. Krantz^{1,3,*}

¹Division of Human and Molecular Genetics, The Children's Hospital of Philadelphia, PA 19104

²Center for Biomedical Informatics, The Children's Hospital of Philadelphia, PA 19104

³The University of Pennsylvania School of Medicine, Philadelphia, PA 19104

⁴Department of Human Genetics, The University of Chicago, Chicago, IL 60637

⁵Harvey Institute for Human Genetics, Department of Pediatrics, Greater Baltimore Medical Center, Baltimore, Maryland, MD 21204

⁶Department of Obstetrics and Gynecology, Drexel University School of medicine, Philadelphia, PA 19104

Abstract

Cornelia de Lange Syndrome (CdLS) is a dominantly inherited heterogeneous genetic disorder with multi-system abnormalities. 60% of probands with CdLS have heterozygous mutations in the *Nipped-B-like (NIPBL)* gene, 5% have mutations in the *SMC1A* gene, and 1 proband was found to have a mutation in the *SMC3* gene. Cohesin is a multi-subunit complex consisting of a SMC1A and SMC3 heterodimer and two non-SMC subunits. *SMC1A* is located on the human X chromosome and is reported to escape X inactivation. We show that 29 unrelated CdLS probands with 21 unique *SMC1A* mutations have been identified by our group and others including 7 males. All mutations identified to date are either missense or small deletions with all presumably preserving the protein open reading frame. Both wild type and mutant alleles are expressed. Females quantitatively express twice the amount of *SMC1A* mRNA as compared to males. The transcriptional profiling of 23 selected genes is different in *SMC1A* mutant probands, controls and *NIPBL* mutant probands. These results suggest that mechanistically *SMC1A*-related CdLS is not due to altered levels of the *SMC1A* transcript, but rather that the mutant proteins maintain a residual function in males and enact a dominant negative effect in females.

Keywords

SMC1A; CdLS; X-linked; Expression

* To whom correspondence should be addressed: Ian D. Krantz, M.D., Associate Professor of Human Genetics in Pediatrics, Director, Medical Genetics Training Program, University of Pennsylvania School of Medicine, The Children's Hospital of Philadelphia, 1007C Abramson Research Center, 3615 Civic Center Boulevard, Philadelphia, PA 19104-4318, Tel. (215) 590-2931, Fax. (215) 590-3850, ian2@mail.med.upenn.edu.

Supporting Information for this preprint is available from the *Human Mutation* editorial office upon request (humu@wiley.com)

Introduction

Cornelia de Lange Syndrome (CdLS; MIM#s 122470, 300590) is a dominantly inherited congenital disorder with multi-system anomalies including characteristic facial dysmorphism, limb defects (with the upper extremities being more severely affected), abnormalities of the gastrointestinal, cardiac, genitourinary, and other systems, growth retardation, and neurodevelopmental delays (Krantz, et al., 2004; Tonkin, et al., 2004). Recently the cohesin complex and its regulators have been etiologically implicated in the pathogenesis of CdLS (Krantz, et al., 2004; Musio, et al., 2006; Tonkin, et al., 2004). The cohesin complex consists of 4 major subunits, a SMC1A and SMC3 heterodimer, RAD21 (SCC1/MAD1), and SCC3 (STAG/SA). Cohesin controls sister chromatid cohesion during mitotic cell cycle with NIPBL facilitating its loading and unloading (Hirano, 2006). 60% of probands with CdLS have mutations in the *Nipped B-like* (*NIPBL*; MIM# 608667), 5% have mutations in *SMC1A* (MIM# 300040), and 1 proband was found to have a mutation in *SMC3* (MIM# 606062) (Deardorff, et al., 2007; Gillis, et al., 2004; Musio, et al., 2006) confirming genetic heterogeneity in CdLS. This also suggests the possibility of unknown additional genes or modifiers that contribute, either individually or in combination, to the CdLS phenotype as 35% of clinically diagnosed individuals with CdLS do not have identifiable mutations in these genes (Deardorff, et al., 2007). Little is known about how these structural cohesin components cause CdLS when mutated. The discovery that the cohesin complex also has a fundamental role in long-range regulation of transcription in *Drosophila* has shed light on the mechanism likely responsible for its role in development (Rollins, et al., 1999). Cohesin mediated gene regulation has been elucidated in several recent reports (Dorsett, 2007; Liu, et al., 2009; McNairn and Gerton, 2008; Wendt, et al., 2008), thus the CdLS phenotype is believed to be caused by gene dysregulation rather than chromosomal segregation effects, although DNA repair function has also been found to be impaired (Revenkova, et al., 2009; Vrouwe, et al., 2007).

SMC1A is an X-linked gene that is reported to escape X-inactivation in humans (Brown, et al., 1995; Tsuchiya and Willard, 2000), although it is subject to X inactivation in mouse (Sultana, et al., 1995). All human mutations identified in *SMC1A* to date are missense or small deletion mutations, preserving the open reading frame of the gene (Borck, et al., 2007; Deardorff, et al., 2007; Musio, et al., 2006). By comparison, the majority of *NIPBL* mutations in CdLS probands are frameshift or nonsense mutations that presumably lead to haploinsufficiency (Gillis, et al., 2004). A recent report suggests that the dynamic association between the cohesin complex and DNA is interfered by *SMC1A* mutations in CdLS probands (Revenkova, et al., 2009). In order to elucidate the biological role that mutant *SMC1A* may play in the pathogenesis of CdLS and identify a possible mechanism, we have summarized the *SMC1A* mutations identified in our laboratory and from other's reports (Borck, et al., 2007; Musio, et al., 2006) and performed targeted expression analyses on lymphoblastoid cell lines (LCLs) from *SMC1A* mutation positive CdLS probands.

Materials and Methods

DNA Isolation and Mutation Analysis

Genomic DNA was isolated from peripheral blood lymphocytes or LCLs (Gentra Systems). The entire *SMC1A* coding region (exons 1-25) as well as 5'-UTR was screened for mutations as described previously (Gillis, et al., 2004). PCR products were purified using QIAquick PCR purification kit (Qiagen) and were sequenced bidirectionally on an ABI 377 sequencer. GenBank NM_006306.2 and NP_006297.2 were used as *SMC1A* sequence references. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference (www.hgvs.org/mutnomen). The initiation codon is codon 1.

Sample Collection and Culture Conditions

All human subjects in this study were enrolled under an institutional review board–approved protocol of informed consent at The Children's Hospital of Philadelphia. All subjects were evaluated by one or more experienced clinicians. Lymphoblastoid cell lines (LCL) from 16 controls (8 females and 8 males) and 7 clinically affected CdLS probands with *SMC1A* mutations (6 females and 1 male) were grown uniformly in RPMI 1640 with 20% fetal bovine serum (FBS), 100 U penicillin/mL, 100 µg streptomycin/mL, and 1% L-glutamine. 5×10^6 exponentially growing cells were seeded in 15 ml media in 75 ml Falcon flasks, incubated at 37 °C and 5% CO₂. Cells were fed 24 hours later. After another 24 hours on day 3, 8 ml of the media was centrifuged and cell pellets were proceeded directly to DNA or RNA extraction.

Total RNA Isolation and cDNA Synthesis

Total RNA was extracted with the RNeasy Mini-Kit (Qiagen) and synthesis of double-stranded cDNA was performed using SuperScript First-Stranded cDNA Synthesis Kit (Invitrogen) according to the manufacturer's instruction. 8 µl total RNA (4 µg) and 4 µl random hexamer (200 ng) was used for the first strand cDNA synthesis in a 20 µl volume. Control samples were run concurrently according to the same protocol, with no SuperScript II RT added.

RT-PCR and Quantitative RT-PCR

RT-PCR reactions were set up in a 25µl volume containing 2 µl of cDNA products, 20 pmol of primers, 75µM of dNTPs, 1× PCR Buffer II (Applied Biosystems), 1 U AmpliTaq Gold (Applied Biosystems), and 2.0 mM MgCl₂ (Applied Biosystems). Samples were run under the following thermocycler conditions: 35 cycles at 95°C for 30 s, 58°C for 30 s, and 72°C for 30 s, followed by 72°C for 5 min. Products were then visualized by gel electrophoresis on a 1.5% Agarose gel. Quantitative RT-PCR was performed using the SYBR green master mix kit (Applied Scientific) in the ABI SDS-7900HT system (Applied Biosystems) in 384 well plates. All parameters were set up as default values. Relative gene expressions were calculated using the standard curve method or ΔC_t method (Roche Applied Science Technical Note No. LC 13/2001, <https://www.roche-applied-science.com/techresources>). 2 µl of first strand cDNA was added in 198 µl ddH₂O and called “100×”, “10×” and “1×” samples were then made by serial dilutions. “100×”, “10×” and “1×” dilutions were generated for each proband and each dilution was tested in triplicate. First strand cDNA synthesized without the addition of reverse transcriptase was used as reverse transcription negative control. GAPDH was used as an internal dosage control. For sequencing assay, RT-PCR products were subsequently run bidirectionally on an ABI 377 sequencer using the same pair of primers.

Western Blotting

Protein was extracted using lysis buffer (150 mM NaCl, 1% NP40, 50 mM Tris-HCl, pH 7.5) with protease inhibitor freshly added. The lysate containing 50 µg of cellular proteins was fractionated on the 10% pre-cast denaturing polyacrylamide gel (Bio-Rad Laboratories) then electro-transferred onto nitrocellulose membranes (Amersham Biosciences). Non-fat dried milk (Bio-Rad Laboratories) in TBS (Bio-Rad Laboratories) containing 0.1% Tween (Sigma) was used to make blocking solution. The membranes were incubated in 10% blocking solution overnight at 4°C. After that, the membranes were cut into 2 pieces, the upper piece was incubated either with a polyclonal rabbit anti-human SMC1 antibody (ab9262, Abcam) (1:200 dilution) or a polyclonal rabbit anti-human RAD21 (ab992, Abcam) (1:200); the lower piece was incubated with a monoclonal mouse anti-human ACTB (sc-8432, Santa Cruz) (1:1000 dilution). The above antibodies were diluted in 4%

blocking solution at room temperature. Following a 1 hour incubation, the membranes were washed for 5 minutes each time for six times with 0.1% Tween in TBS and then incubated with secondary antibodies, goat anti-rabbit IgG HRP (sc-2004, Santa Cruz) or goat anti-mouse IgG-HRP (sc-2005, Santa Cruz), both at 1:2000 dilution in 2.5% blocking solution at room temperature for 1 hour. The membrane was washed and signal was detected by using Chemiluminescence Reagent (Bio-Rad Laboratories) according to the manufacturer's protocol.

Custom Array Hybridization

Custom arrays designed in our laboratory were described previously (Liu, et al., 2009). In brief, 23 genes differentially expressed in severely affected CdLS probands with *NIPBL* mutations were selected. A custom Ziplex[®] TipChip microarray (Xceed Molecular) (<http://www.xceedmolecular.com/>) containing oligonucleotide probes of between 35 and 50 basepairs for these 23 genes was used to profile differences in gene expression between the LCL samples. Total RNA (500 ng) from 7 *SMC1A* mutant probands, 21 severely affected *NIPBL* mutant probands as well as 20 matched healthy controls was amplified and biotin labeled with the Illumina[®] Totalprep RNA amplification kit (Ambion). After hybridization, expression values were calculated by the Ziplex[®] Automated Workstation software. Statistical analysis was performed in the CHOP Bioinformatic Center.

Primer Design

SMC1A cDNA (GenBank NM_006306.2) was used as reference sequence. Primer 3 (<http://primer3.sourceforge.net/>) was used to design primers for RT-PCR, and Primer Express software v3.0 (Applied Biosystems) was used to design primers for quantitative RT-PCR and allele specific quantitative RT-PCR. To design primers for allele specific quantitative RT-PCR, the 3'-end nucleotide of a primer matches to either the wild-type or mutant allele, in addition, a second single mismatch was inserted to within 5 nucleotides near the 3'- end for each allele specific primer in order to enhance the specificity and sensitivity. A pair of primers that amplify a common region surrounding the mutation point was used for normalization. If no primer met the software criterion, we would manually select the primers to amplify the mutations. GAPDH was used as internal control for all quantitative RT-PCR assays. Primer sequences are available on request.

Results

SMC1A Mutation Identification

We have performed genomic *SMC1A* sequencing to screen CdLS probands without identifiable *NIPBL* mutations in our laboratory at The Children's Hospital of Philadelphia and in the molecular diagnostic laboratory at the University of Chicago. In addition to the 12 probands with 8 unique mutations in *SMC1A* genes from our previous report (Deardorff, et al., 2007), we have identified 11 additional unique *de novo* mutations in *SMC1A* from 17 CdLS probands (including 2 affected sisters). We further included 2 of the 4 mutations reported by Musio et al. (Musio, et al., 2006) and Brock et al. (Borck, et al., 2007) which were not seen in our cohort. All of the newly identified mutations in our group (c.421G>A, c.802_804del3, c.916_918del3, c.1193G>A, c.1478A>C, c.2046_2048delAGA, c.2077C>G, c.2342G>T, c.2446C>G, c.3146G>A, c.3367C>T) were *de novo* and not present in unaffected parents. In addition, all mutations were absent in more than 220 normal control alleles. Figure 1A and 1B depict *SMC1A* mutations identified to date among the 31 CdLS probands (29 unrelated probands plus 2 affected sisters). There are a total of 21 missense or small deletion mutations identified (Table 1). All lead to in-frame amino acid changes, and all the mutations are located at the amino acid sites that are conserved across multiple species

(http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=homologene&dopt=MultipleAlignment&list_uids=4597). Mutations are spread along the protein coding region, while no mutation was found in the 5'-UTR. We did not screen the 3'-UTR of *SMC1A* due to its big size (6,038 bp). There are 4 unrelated probands with mutations at c.1487G>A (p.R496H) in exon 9. There does not seem to be any gender specificity in the identified mutations, although c.587G>A (p.R196H) was identified in 2 male probands and not in any of the female probands (Fig. 1A). There are in total 22 females and 7 male unrelated probands. All are defined as mildly affected due to lack of major structural abnormalities and with mental retardation ranging from mild to moderate. The coding region of the *SMC1A* (NM_006306.2 and NP_006297.2) was searched for conserved domains using the online tool from NCBI (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Three conserved domains were recognized (an amino terminal NTPase domain from amino acid 4-148, a hinge domain from amino acid 515-629, and a carboxy terminal NTPase domain from amino acid 1,117-1,220). Mutations corresponding to each domain are illustrated (Fig. 1B). The *SMC1A* protein folds back on itself with the two NTPase domains forming a head domain. The head domain and the hinge domain of *SMC1A* therefore interact with *SMC3* to form a cohesin ring that, in proposed models, embraces the chromatin (Hirano, 2002; Hirano, 2006; Nasmyth, 2005). Most of the *SMC1A* mutations are scattered evenly outside the three conserved domains. No mutation was identified in the hinge domain. There are only a few mutations located within the NTPase domains (5 mutations), however, none of these mutations directly involves the functional motifs (ATP binding sites, ABC transporter signature motif, walker B, the walker/p-loop, Q-loop, D-loop and H-loop (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)), although one cannot rule out indirect actions on these motifs in the folded protein (Fig. 1B). Analyses of the *SMC1A* mutations among the 29 unrelated CdLS probands suggests that: 1) although more female probands have been identified the level of affection between male and female probands is equal in severity, 2) the mutant *SMC1A* protein is synthesized and has some functional activity since all mutations identified so far are in-frame alterations, 3) mutant *SMC1A* proteins, although impaired, seem able to maintain most of its biological capability as no mutation interferes with key motifs and mutation positive males (without a normal *SMC1A* allele) are viable.

***SMC1A* Escapes X Inactivation and Is Biallelically Expressed in Human Lymphoblastoid Cells**

SMC1A is located on human chromosome Xp11.21-11.22 in an area that was reported to escape X inactivation more than a decade ago, and there is no *SMC1A* homolog on the Y chromosome identified (Brown, et al., 1995; Carrel and Willard, 2005). A literature search has retrieved limited direct evidence of the biallelic expression of *SMC1A*, probably due to the lack of information on exonic single nucleotide polymorphisms (SNPs) to differentiate the expression of the two alleles. The cDNA sequencing of probands with mutations in *SMC1A* coding regions will confirm *SMC1A* biallelic expression and further demonstrate whether mutant alleles are also expressed or not. Lymphoblastoid cell lines (LCLs) are available in our laboratory for 7 CdLS probands (6 females and 1 male) with *SMC1A* mutations therefore the 6 female probands were selected for this study. After RNA extraction, first strand cDNA was synthesized from each of the 6 samples and RT-PCR was performed. cDNA sequencing was examined by both forward and reverse primers (Supp. Fig. S1). Both of the wild type and mutant type alleles of *SMC1A* are expressed in all of the 6 probands, indicating transcription from the mutant allele is not impaired seriously (sequences from reverse primers are not shown).

Transcriptional Analysis of *SMC1A*

Gene expression was analyzed on lymphoblastoid cell lines (LCLs) from 7 CdLS probands (6 females and 1 male) with *SMC1A* mutations. In addition, 8 healthy females and 8 healthy males were chosen as controls. RT-PCR was first conducted on all of the 23 samples. After reverse transcription, using primers that amplified the region between exons 19 and 24, we have found that all the samples were able to produce the expected 628 bp amplicons indicating the existence of the processed *SMC1A* mRNA in each of them. RT-PCR amplification was also seen in the male proband 017P indicating that the missense mutation (c.397T>G, p.F133V) in him does not disrupt *SMC1A* transcription. The RT-PCR results of 1 female proband, 1 female control, 1 male proband, and 1 male control are showed in Figure 2A.

Quantitative RT-PCR was performed to examine possible dosage changes of *SMC1A* transcripts between males and females, controls and patients. *SMC1A* is X-linked and is reported to escape X inactivation (Brown, et al., 1995). If *SMC1A* escapes X inactivation we would anticipate that normal females would express twice the amount of *SMC1A* transcripts as compared to normal males (assuming autoregulation of transcript levels does not occur). In order to avoid biases while doing reverse transcription, we have designed 2 pairs of primers against either the 5'-end or the 3'-end of *SMC1A* cDNA. Female probands did not show reduced *SMC1A* mRNA levels with female probands and female controls demonstrating similar levels of *SMC1A* transcripts ($p=0.26$ and $p=0.10$ for the 2 primer pairs respectively). This was also true for the male controls and the proband, although statistical analysis could not be performed as there was only one male proband tested. However, the expression level of females is almost twice as much as that of males in both control and proband categories ($p=0.0004$ for comparison between control females and control males) (Fig. 2B and Supp. Fig. S2A). Combining and dividing all samples according to their gender, re-analysis demonstrated a clearer picture with females expressing two fold more *SMC1A* mRNA than males ($p=0.0008$ and $p=0.003$ for the 2 primer pairs respectively) (Fig. 2C and Supp. Fig. S2B). If we arranged our samples into 5 families and compared *SMC1A* mRNA levels between members of the same family as well as across different families, we have found large variations between both related and un-related individuals. In general, within the same family, females generally express twice as much *SMC1A* transcripts as the males, with no expression difference between healthy females and the CdLS females. However, female probands in one family could have higher *SMC1A* expression than the healthy females in another family, and the probands could have higher expression levels than their parents (Fig. 2D and Supp. Fig. S2C). These results strongly support that *SMC1A* escapes X-inactivation, and the total amount of mRNA produced in the CdLS probands is not decreased. However, the amount of *SMC1A* expressed is quite variable in humans. Dosage sensitivity for *SMC1A* levels may not be critical in CdLS, but rather the dominant acting effects of the mutations are likely the driving force in causing the phenotypic effects.

Although the overall transcription did not change in *SMC1A* mutant individuals, the wild type and the mutant alleles may be transcribed differently. For example, the mutant allele could produce less mRNA and the wild type allele produce more mRNA due to a potential feedback mechanism. Conversely it is possible that the wild type allele is transcribed at twice the normal level while the mutant allele is completely shut down. In order to assay for this possibility we performed allele specific quantitative RT-PCR. Out of the 6 female cell lines available on the CdLS probands, there were 5 different mutations in *SMC1A*. Primer Express software v3.0 failed to design any allele specific primer for either the wild type or the mutant allele for these 5 probands, primers were then manually selected but unfortunately failed all the quantitative tests due to lack of specificity. In the end, we were able to design primers for both wild type and mutant alleles for proband 074P (c. 3364 T>C, p.F1122L). We have also designed primers to amplify a common region surrounding the

mutation which was used to normalize the expression levels of the wild type and the mutant alleles by ΔC_t methods. The mutant allele has a slightly reduced expression level as compared to the wild type allele, and it is statistically significant ($p=0.01$) (Fig. 3, Supp. Fig. S3). Although with data from only one sample, this minor change between the two alleles may explain why we could not detect the overall expression alterations between controls and probands.

Translational Analysis of *SMC1A* by Western Blot

We next tested if the mutations in *SMC1A* could interfere with protein synthesis. We have chosen 2 families (017 and 174) for Western blot analysis, each family included the proband (P), the father (Fa), and the mother (Ma). 017P is a male proband and 174P is a female. We have also included 2 sporadic probands: 060P and 203P. All probands, including the male proband, demonstrated synthesis of the *SMC1A* protein, but quantifiable differences in the amount of protein were not obvious between probands and their parents, nor between probands and unrelated controls (Fig. 4). Although quantitative RT-PCR has revealed that females transcribe twice as much *SMC1A* mRNA as males, the Western blot does not allow us to effectively examine protein with 2 fold changes quantitatively. Another core subunit of the cohesin complex, RAD21, was used as an additional internal protein control and was present at comparable levels in both proband and control samples.

Gene Expression is Distinct Between CdLS Probands with *SMC1A* Mutations, CdLS Probands with *NIPBL* Mutations and Unaffected Individuals

CdLS probands with *SMC1A* mutations are characterized more by cognitive impairments while their other clinical manifestations are milder than the classic *NIPBL* mutant probands (Deardorff, et al., 2007). In our previous studies (Liu, et al., 2009), we have identified a specific gene expression profile for classic CdLS probands with *NIPBL* protein truncating mutations. A custom built array platform with a cohort of 23 targeted genes from the genome wide expression profile that were significantly mis-expressed in severely affected CdLS probands was created to demonstrate if the phenotypical diversity between probands with *NIPBL* and *SMC1A* mutations can be explained by different gene expression patterns by analyzing the transcription of the 23 genes using our custom arrays on severely affected probands with *NIPBL* mutations (21 individuals), *SMC1A* mutant probands (7 individuals as we have analyzed above) as well as matched healthy controls (20 individuals). The average expression level of each gene is different between the 3 groups, with the level of misexpression in the *SMC1A* mutant group being intermediate between the *NIPBL* mutant and control samples with only one exception, although the expression difference is not always statistically significant (Supp. Table S1). Principle Component Analysis (PCA) based on the expression levels of all the 23 genes was able to differentiate *NIPBL* mutant probands from healthy controls with *SMC1A* mutant probands aligned between these groups. This would indicate that individuals with *SMC1A* mutations are similar but not quite the same as the severely affected individuals based on gene expression, which might explain the phenotypical variations between the 2 CdLS subgroups (Fig. 5).

Discussion

During the mitotic cell cycle, sister chromatid cohesion is controlled by the cohesin complex which is formed by subunits *SMC1A*, *SMC3*, *SCC1*, and *SCC3* (Hirano, 2006; Nasmyth, 2005). *NIPBL* facilitates cohesin loading and unloading. Approximately 60% of CdLS probands have mutations in *NIPBL*, 5% have mutations in the *SMC1A* gene and one proband with an *SMC3* mutation has been reported (Deardorff, et al., 2007). Previous SMC protein biochemistry, electron microscopy, and x-ray crystallography have shown that SMC proteins consist of globular N- and C-terminal domains that contain ATP-binding Walker A

and Walker B motifs, respectively (Hirano, 2006; Nasmyth, 2005). These motifs are juxtaposed in the globular head domain of the SMCs. It has been shown that ATP binding promotes the association of the two intermolecular heads and that ATP hydrolysis drives them apart. Both of these events facilitate the ability of cohesin to encircle chromatin. At the opposite side of the ring, a central domain of SMC1A forms a globular hinge domain that interacts with a similar motif of SMC3 to form a heterodimer. It has been shown that mutations at the *Bacillus* SMC hinge domain disrupts DNA binding and ATP hydrolysis, leading to disruption of the DNA loading process (Hirano, 2002; Hirano, 2006; Nasmyth, 2005). By matching the 21 SMC1A mutations (19 from our group and 2 from published reports) from 29 unrelated CdLS probands to the SMC1A protein domains and motifs, we have revealed that most of the mutations are located outside the head and hinge domains. A few mutations are within the head domains, however, none of the mutations directly involves the functional motifs (Fig. 1), although we cannot rule out interference with these domains in the mature fully folded protein. Cohesin formation and chromatin segregation does not seem to play a major role in the pathogenesis of CdLS.

Prior studies (Deardorff, et al., 2007) have shown that CdLS probands with *SMC3* and *SMC1A* mutations present with a milder clinical picture than the classic form of CdLS typically associated with truncating mutations in *NIPBL*. In the initial study of 12 patients with *SMC1A* or *SMC3* mutations, our group found that these patients are more likely to have normal birth weight and growth patterns, to walk, and to acquire speech (Deardorff, et al., 2007). All patients studied had mild facial features consistent with CdLS and had varying degrees of mental retardation, but none had major structural anomalies typically seen in classic CdLS. This is in stark contrast to individuals with truncating *NIPBL* mutations who are generally very severely affected with multiple structural anomalies and severe growth and cognitive retardation. The diversity of the 2 groups of CdLS probands could possibly be explained by their dissimilar gene expression patterns (Fig. 5). Since *NIPBL* is a regulator of the cohesin complex and both *SMC3* and *SMC1A* are actual structural components of cohesion complex, the exact mechanisms by which mutations in these proteins manifest their effects on development and gene regulation are likely to be quite different.

We used RT-PCR to qualitatively confirm that *SMC1A* is expressed in lymphoblastoid cell lines grown from CdLS probands and controls. Subsequently we looked quantitatively at the expression level of *SMC1A* in male versus female controls. Quantitative RT-PCR showed that females express twice as much *SMC1A* mRNA as males, indicating that both X chromosomes are likely responsible for the production of *SMC1A* mRNA. As allele specific primers for each individual mutation could not be designed successfully due to the repetitive sequences surrounding the mutation alleles, we were only able to examine one proband. By looking at allele specific expression in this female, we found that both the mutant and wild type alleles are expressed, although at slightly different levels. Therefore, we suspect that the CdLS phenotype in female patients is due to a dominant negative effect of the altered protein. *SMC1A* has been reported to escape X inactivation in human (Brown, et al., 1995) but subject to X inactivation in mouse (Sultana, et al., 1995). Biallelic expression data of *SMC1A* could not be found in the literature except for one study using human mouse somatic cell hybrids (Carrel and Willard, 2005). Our study confirms the biallelic expression of the wild type and mutant alleles of *SMC1A* both by q-RT-PCR and cDNA sequencing.

Quantitative RT-PCR and Western blot analyses have also shown that female probands compared to female controls and male probands compared to male controls express equal levels of *SMC1A*, both at the transcriptional and translational levels. The mutations do not have an influence on overall expression level, indicating that the mutant mRNA is not being actively degraded and it can be translated into mutant protein, this result is consistent with previous reports (Musio, et al., 2006; Revenkova, et al., 2009). Prior sequencing studies

have shown that all identified *SMC1A* mutations are protein sparing, in that they are either missense or small in frame deletions. While the mutated proteins may have slightly altered function, they are suspected to be produced in similar amounts as normal *SMC1A*. We presume that “non-protein sparing” mutations, such as frameshift mutations, are incompatible with life in males and would result in no phenotype in females as long as they had one normal *SMC1A* allele present. Since *SMC1A* is a component of the cohesin ring, mutations that prevent the production of the *SMC1A* protein in males (who have only one allele) would prevent sister chromatid cohesion and disrupt normal cell division. Additionally, since *SMC1A* and cohesin complex is thought to be a “master switch” that controls various other genes along the chromosome, the implications for non-functional *SMC1A* could be severe disruption of gene regulation (Liu, et al., 2009). We believe that the protein-sparing mutations are consistent with the milder phenotype of probands with *SMC1A* mutations, as residual function of the complex is dependent on survival. We have previously found that probands with *NIPBL* mutations had more severe characteristics of CdLS and a greater variety of frameshift and non-sense mutations. It appears that there is a greater variability in *NIPBL* mutations among CdLS probands (Deardorff, et al., 2007). Because the *NIPBL* protein is a regulator of cohesin and is not a structural part of the cohesin ring (it is thought to assist with loading and unloading of chromosomes from the ring), there may be more room for variability in terms of the functionality of *NIPBL*. It has been demonstrated in our lab that haploinsufficiency of *NIPBL* is tolerated by humans, as the other allele is up-regulated and actual mRNA levels of *NIPBL* in probands with truncating mutations of one allele is actually 70% of unaffected individuals (Liu, et al., 2009). Similar levels of *NIPBL* expression in probands was also reported by another group (Borck, et al., 2006).

One limitation of the current study is that Western blot cannot be used to quantify the exact protein levels for a measure of gene expression. It is possible that probands with *SMC1A* mutations undergo normal transcription and produce normal levels of mRNA, but the mutated transcript is degraded. The detection of *SMC1A* protein in the male proband confirms that mutated proteins are processed and produced. The lack of precocious sister chromatid separation (PSCS) in the studied probands (data not shown) suggests that sister chromatid cohesion is not sensitive to mutant cohesin proteins, which is consistent with other reports that have shown negative PSCS in different organisms with mutations in cohesin subunits or *NIPBL* (Dorsett, et al., 2005; Losada, et al., 1998; Uhlmann and Nasmyth, 1998; Zhang, et al., 2007). However, more *SMC1A* mutant probands should be studied and more sensitive assays that could reflect chromosome segregation defects with a better resolution need to be performed. Recently an increasing number of studies have suggested *SMC1A* and *NIPBL* are “master switches” that control downstream gene regulation. Genome-wide expression arrays to analyze CdLS probands with *SMC1A* or *NIPBL* mutations reveals that probands with *SMC1A* mutations show a lesser degrees of dysregulation of gene expression than those individuals with *NIPBL* mutations and are closer to unaffected individuals in their expression profiles (Fig. 5, Supp. Table S1) (Liu, et al., 2009). This implies that *SMC1A* mutations exert a weaker block on transcriptional control at G1 phase, which is more consistent with the milder phenotype seen in these patients, although defects in double strand DNA repairs at G2 phase may also be involved in CdLS pathogenesis (Revenkova, et al., 2009; Vrouwe, et al., 2007). The work presented here represents the first steps towards understanding how alterations in this important pathway, newly implicated in human development, affects its control of gene expression. Understanding the exact mechanism by which mutant *SMC1A* proteins affect downstream dysregulation of gene expression will help lead to the identification of downstream targets of cohesin function and hopefully towards the development of effective treatments in the future.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful for the participation of the children and families with CdLS and to the CdLS Foundation for their support.

Funding: This work was supported by a CdLS Foundation Fellowship Grant (to J.L.); and the National Institute of Child Health and Human Development (PO1 HD052860 and R21 HD050538 to I.D.K., KO8 HD055488 to M.A.D.).

References

- Borck G, Zarhrate M, Bonnefont JP, Munnich A, Cormier-Daire V, Colleaux L. Incidence and clinical features of X-linked Cornelia de Lange syndrome due to SMC1L1 mutations. *Hum Mutat.* 2007; 28(2):205–6. [PubMed: 17221863]
- Borck G, Zarhrate M, Cluzeau C, Bal E, Bonnefont JP, Munnich A, Cormier-Daire V, Colleaux L. Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5' untranslated region of the NIPBL Gene. *Hum Mutat.* 2006; 27(8):731–5. [PubMed: 16799922]
- Brown CJ, Miller AP, Carrel L, Rupert JL, Davies KE, Willard HF. The DXS423E gene in Xp11.21 escapes X chromosome inactivation. *Hum Mol Genet.* 1995; 4(2):251–5. [PubMed: 7757075]
- Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature.* 2005; 434(7031):400–4. [PubMed: 15772666]
- Deardorff MA, Kaur M, Yaeger D, Rampuria A, Korolev S, Pie J, Gil-Rodriguez C, Arnedo M, Loeys B, Kline AD, 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):485–94. [PubMed: 17273969]
- Dorsett D. Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. *Chromosoma.* 2007; 116(1):1–13. [PubMed: 16819604]
- Dorsett D, Eissenberg JC, Misulovin Z, Martens A, Redding B, McKim K. Effects of sister chromatid cohesion proteins on cut gene expression during wing development in *Drosophila*. *Development.* 2005; 132(21):4743–53. [PubMed: 16207752]
- Gillis LA, McCallum J, Kaur M, DeScipio C, Yaeger D, Mariani A, Kline AD, Li HH, Devoto M, Jackson LG, 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):610–23. [PubMed: 15318302]
- Hirano T. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev.* 2002; 16(4):399–414. [PubMed: 11850403]
- Hirano T. At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol.* 2006; 7(5):311–22. [PubMed: 16633335]
- Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, Yaeger D, Jukofsky L, Wasserman N, Bottani A, Morris CA, 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):631–5. [PubMed: 15146186]
- Liu J, Zhang Z, Bando M, Itoh T, Deardorff MA, Clark D, Kaur M, Tandy S, Kondoh T, Rappaport E, et al. Transcriptional Dysregulation in NIPBL and Cohesin Mutant Human Cells. *PLoS Biol.* 2009; 7(5):e1000119. [PubMed: 19468298]
- Losada A, Hirano M, Hirano T. Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. *Genes Dev.* 1998; 12(13):1986–97. [PubMed: 9649503]
- McNairn AJ, Gerton JL. Cohesinopathies: One ring, many obligations. *Mutat Res.* 2008; 647(1-2):103–11. [PubMed: 18786550]

- Musio A, Selicorni A, Focarelli ML, Gervasini C, Milani D, Russo S, Vezzoni P, Larizza L. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat Genet.* 2006; 38(5):528–30. [PubMed: 16604071]
- Nasmyth K. How might cohesin hold sister chromatids together? *Philos Trans R Soc Lond B Biol Sci.* 2005; 360(1455):483–96. [PubMed: 15897174]
- Revenkova E, Focarelli ML, Susani L, Paulis M, Bassi MT, Mannini L, Frattini A, Delia D, Krantz I, Vezzoni P, et al. Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. *Hum Mol Genet.* 2009; 18(3):418–27. [PubMed: 18996922]
- Rollins RA, Morcillo P, Dorsett D. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics.* 1999; 152(2):577–93. [PubMed: 10353901]
- Sultana R, Adler DA, Edelhoff S, Carrel L, Lee KH, Chapman VC, Willard HF, Disteche CM. The mouse Sb1.8 gene located at the distal end of the X chromosome is subject to X inactivation. *Hum Mol Genet.* 1995; 4(2):257–63. [PubMed: 7757076]
- 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):636–41. [PubMed: 15146185]
- Tsuchiya KD, Willard HF. Chromosomal domains and escape from X inactivation: comparative X inactivation analysis in mouse and human. *Mamm Genome.* 2000; 11(10):849–54. [PubMed: 11003698]
- Uhlmann F, Nasmyth K. Cohesion between sister chromatids must be established during DNA replication. *Curr Biol.* 1998; 8(20):1095–101. [PubMed: 9778527]
- Vrouwe MG, Elghalbzouri-Maghrani E, Meijers M, Schouten P, Godthelp BC, Bhuiyan ZA, Redeker EJ, Mannens MM, Mullenders LH, Pastink A, et al. Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. *Hum Mol Genet.* 2007; 16(12):1478–87. [PubMed: 17468178]
- Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature.* 2008; 451(7180):796–801. [PubMed: 18235444]
- Zhang B, Jain S, Song H, Fu M, Heuckeroth RO, Erlich JM, Jay PY, Milbrandt J. Mice lacking sister chromatid cohesion protein PDS5B exhibit developmental abnormalities reminiscent of Cornelia de Lange syndrome. *Development.* 2007; 134(17):3191–201. [PubMed: 17652350]

Figure 1.

Twenty-one mutations in *SMC1A* identified among 29 unrelated CdLS probands. (A) Schematic illustration of *SMC1A* gene structure and localization of mutations in CdLS probands. Mutations are uniformly distributed along *SMC1A* coding regions. Labeled blocks represent exons. Mutations of probands are described in both nucleotide and amino acid changes. Probands from our group are marked with “CDL” or “UC”, probands reported previously by Deardroff *et al.* are underlined with previous labels in the brackets, probands reported by Musio *et al.* and Borck *et al.* are marked as “Musio” and “Borck” in *italics*, respectively, affected males are labeled with “♂”. (B) *SMC1A* mutations are matched to the corresponding protein domains. Most of the mutations are located outside the conserved protein domains and no mutation is within functional motifs. The hinge domain is mutation free. The arrows point to the 7 probands with expression analyzed in this report. Amino terminal P-loop NTPase domain: aa 4-148; SMC proteins Flexible Hinge Domain: aa 515-629; carboxy terminal P-loop NTPase domain: aa 1,117-1,220. Motifs within the amino terminal P-loop NTPase domain are: walker/P-loop (aa 32-39), ATP binding sites (aa 35-36 and aa 38-40), Q-loop/lid (aa 134-137). Motifs within the carboxy terminal P-loop NTPase domain are: ABC transporter signature motif (aa 1,128-1,137), Walker B (aa 1,152-1,157), ATP binding site (aa 1,156-1,157), D-loop (aa 1,160-1,163) and H-loop/switch region (aa 1,185-1,191) (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). GenBank NM_006306.2 and NP_006297.2 were used as *SMC1A* sequence references. Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence. The initiation codon is codon 1.

Figure 2.

Transcriptional analysis of CdLS probands with *SMC1A* mutations. **(A)** RT-PCR shows that both male and female probands express *SMC1A* mRNA. P: proband; Ma: mother; Fa: father. 107P is a female proband and 017P is a male proband. **(B)** Quantitative RT-PCR reveals females expressed twice as much *SMC1A* mRNA as males, and there is no difference between controls and probands. One male proband (017P) and 6 female probands (060P, 074P, 107P, 144P, 174P and 203P) were tested. **(C)** Quantitative RT-PCR analyzing the 23 samples by their genders further confirms that *SMC1A* escapes X-inactivation. **(D)** Quantitative RT-PCR analysis on 5 different families demonstrated remarkable variations between related and un-related individuals. Families 017, 144, 174, 107 and 074 were analyzed. Primer pairs amplifying the 5' end were used for this analysis. Reverse transcribed cDNAs were tested in triplicates at each dilution for 3 serial dilutions.

Figure 3.

Wild type allele is transcribed at a slightly higher level than the mutant allele of *SMC1A* as demonstrated by allele specific quantitative RT-PCR in proband 074P. Wild type *SMC1A* allele has a small but significant elevated expression level ($p=0.01$) as compared to the mutant allele in 074P.

Figure 4.

Western blot analysis of SMC1A protein in CdLS probands with *SMC1A* mutations. RAD21 protein was also tested as an internal protein control. All examined individual expressed comparable amounts of either SMC1A or RAD21. No difference at the protein level is observed between probands (017P, 174P, 60P and 203P) and healthy controls (017Fa, 017Ma, 174Fa and 174Ma). The male proband 017P does not show less SMC1A or RAD21 proteins than others. Similarly, it is hard to identify the dosage alterations between males (017P, 017Fa, 174Fa) and females (017Ma, 174P, 174Ma, 60P and 203P). SMC1A and RAD21 are represented by a ~160KD and a ~130 KD protein band, respectively. ACTB is the loading control and appears as a protein band around 40 KD. P: proband; Ma: mother; Fa: father.

Figure 5.

Classifications of 3 groups of samples based on a 23 gene expression profile. Unsupervised clustering of 21 healthy controls (green dots), 20 severe CdLS probands with *NIPBL* mutations (red dots) and 7 CdLS probands with *SMC1A* mutations (blue dots) by Principle Component Analysis (PCA). The separation between the control and severe proband groups indicates that they have different gene expression patterns. Individuals with *SMC1A* mutations have expression patterns intermediate to these two distinct groups.

Table 1

Twenty-one in-frame mutations identified in *SMC1A* gene among 29 unrelated CdLS probands.

Base Change	Amino Acid Change	Exon Location	Type of Mutation	No. of Probands
c.173del15	p.V58_R62del	1-2	Deletion	1
c.397T>G	p.F133V	3	Missense	1
c.421G>A	p.E141K	3	Missense	1
c.587G>A	p.R196H	4	Missense	2
c.802_804del3	p.K268del	5	Deletion	2
c.916_918del3	p.S306del	5	Deletion	1
c.1193G>A	p.R398Q	7	Missense	3
c.1478A>C	p.E493A	9	Missense	2
c.1486 C>T	p.R496C	9	Missense	1
c.1487G>A	p.R496H	9	Missense	3 [♫]
c.2046_2048delAGA	p.E683del	12	Deletion	1
c.2077C >G	p.R693G	12	Missense	1
c.2131C>T	p.R711W	13	Missense	2
c.2342G>T	p.C781F	14	Missense	1
c.2369G>A	p.R790Q	14	Missense	1
c.2446C>G	p.R816G	15	Missense	1
c.2493_2495del3	p.Q831E/832del	16	Deletion	1 [♫] , ^ψ
c.3146G>A	p.R1049Q	20	Missense	1
c.3254A>G	p.Y1085C	21	Missense	1 ^ψ
c.3364T>C	p.F1122L	22	Missense	1
c.3367C>T	p.R1123W	22	Missense	1

GenBank NM_006306.2 and NP_006297.2 was used as *SMC1A* reference sequence.

Nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, according to journal guidelines (www.hgvs.org/mutnomen). The initiation codon is codon 1.

[♫] Affected family members are excluded.

^ψ Reported from the literature.