From the Institute of Human Genetics
at the University of Lübeck
Provisional director: Prof. Dr. Frank Kaiser

Functional and genetic analyses of coding and non-coding variants causing Cornelia de Lange Syndrome (CdLS)

Dissertation

in fulfilment of the requirements
for the Doctoral Degree
at the University of Lübeck

from the Department of Natural Sciences

Submitted by
Jelena Pozojević
from Zrenjanin, Serbia
Lübeck, 2018

First referee: Prof. Dr. Frank Kaiser

Second referee: Prof. Dr. Jens Mittag

Date of oral examination: 23.7.2018.

Approved for printing: 26.7.2018.

Table of contents

Abstract

Cornelia de Lange Syndrome (CdLS) is a dominantly inherited malformation syndrome. The phenotype is very broad and heterogeneous including characteristic facial features, intellectual disability, upper limb anomalies, growth disturbances and a large variety of other signs and symptoms. In approximately 75% of patients clinical diagnosis can be confirmed on genetic level. Before the beginning of my PhD studies, disease-causing variants in five genes encoding structural components (*SMC1A*, *SMC3*, *RAD21*) or regulators (*NIPBL*, *HDAC8*) of the cohesin complex have been identified as genetic cause of CdLS. At least 65% of CdLS patients carry variants in *NIPBL* and these patients are often more severely affected, presenting with a 'classical' CdLS phenotype. The remaining four genes account for approximately 10% of cases. However, genetic causes of the remaining ~25% of CdLS patients were unknown. Within my PhD time, we were able to augment the spectrum of mutations in the five previously described CdLS genes. Therefore, we have identified different mosaic variants that have escaped routine diagnostic by Sanger sequencing approaches on lymphocyte DNA, by using DNA from other tissues and more sensitive sequencing technologies such as pyrosequencing or SNaPshot primer extension analysis. By this, we have identified patients with somatic mutations in *NIPBL* as well as an unaffected mother of two siblings with a variant in *HDAC8*. We have shown that somatic mosaicism further contributes to the broad phenotypical variation in patients with CdLS and it is of utmost importance for patient counselling and risk prediction. Based on our data, we recommend the use of more sensitive sequencing technologies even in testing for a disease-causing variant in obviously unaffected parents, in order to exclude low-level mosaicism.

In addition to variants in the five CdLS genes, we have identified variants in *ARID1B*, *SETD5*, *KMT2A* and *ANKRD11* encoding different components involved in transcriptional regulation or chromatin organization. Interestingly, one of the *ANKRD11* variants was also described as somatic mosaicism.

In addition to genetic and functional characterization of protein-coding regions that represent only ~2% of the entire human genome, we have started to investigate non-coding regulatory elements relevant for expression of the 'main CdLS gene' *NIPBL*. By use of various technologies such as chromatin conformation capture, genome editing and reporter gene assays, we have identified an enhancer element upstream of *NIPBL* as well as the long non-coding RNA '*NIPBL*-

AS1' transcribed antisense from *NIPBL* by a common bidirectional promoter. Identification of the functional elements that drive *NIPBL* expression is of extreme importance not only for understanding the complex regulation mechanisms but also because they represent putative target regions for future therapies (e.g. modification of dosage-sensitive *NIPBL* transcript levels).

Furthermore, we have identified a putative enhancer element within a 60 kb-spanning genomic region by the use of different *in silico* analyses heterozygously deleted in three members of a family, all diagnosed as CdLS-like. Genome editing in different cell lines, expression analyses and various functional investigations have shown that this enhancer regulates expression of the *SMC2* gene, which is 1 Mb distant and encodes for a structural component of the condensin complex. Condensin has structural and functional similarities with cohesin, which might explain the phenotypical overlaps of these patients with CdLS. Very recently we have identified two additional cases with a frameshift variant in the protein-coding region of *SMC2* and overlapping clinical features. Further investigations revealed a common misregulation of several condensin genes in patient cells as well as model cells generated by CRISPR/Cas9 genome editing. Our results highlight the role of condensin in transcriptional regulation processes and further show functional relationships between cohesin and condensin, providing a putative explanation for phenotypical overlaps found in these patients.

Within this thesis, a combination of various genetic and functional approaches was used to decipher the molecular mechanisms of CdLS and CdLS-overlapping phenotypes. Expanding the spectrum of mutations in the known and new genes, including a high frequency of somatic mosaicism, is of a high relevance for improving molecular diagnostics and genetic counselling. In addition, our data delineate the importance of functional regulatory elements within non-coding regions of the genome. While research in human genetics focused mainly on the small protein-coding part of the genome, clinical relevance of the remaining 98% non-protein coding genome has been poorly understood. With the rapid development of new sequencing technologies, we have to be prepared to understand this large part of the genome, not only to identify the genetic causes of human disease but also to understand phenotypical variability and allow better prediction of the phenotype and/or disease progression based on genetic information.

Zusammenfassung

Das Cornelia de Lange-Syndrom (CdLS) gehört zur Gruppe der seltenen Erkrankungen. Obwohl das klinische Erscheinungsbild der Patienten mit CdLS sehr heterogen ist, zeigen Patienten mit CdLS charakteristische faziale Merkmale, eine unterschiedlich stark ausgeprägte Intelligenzminderung, Wachstumsstörungen und oftmals Fehlbildungen der Arme und Hände. Bei etwa 75 % der Patienten kann die meist initiale klinische Diagnose genetisch bestätigt werden. Zum Zeitpunkt des Beginns meiner Promotionszeit waren genetische Varianten in fünf verschiedenen Genen als Ursache für das CdLS beschrieben. Alle fünf Gene codieren für Strukturkomponenten (*SMC1A, SMC3, RAD21*) oder Regulatoren (*NIPBL, HDAC8*) des Cohesin-Komplexes. Mindestens 65 % der CdLS-Patienten tragen Varianten im *NIPBL*-Gen. Diese Gruppe der Patienten zeigen häufig einen eher stark ausgeprägten Phänotyp, der auch als „charakteristischer CdLS-Phänotyp“ bezeichnet wird. Genetische Varianten in den anderen vier Genen sind als genetische Ursache für circa weitere 10 % der Patienten beschrieben und für 25 % der Patienten blieb die genetische Ursache ungeklärt. Während meiner Promotionszeit konnten wir das Spektrum und die Anzahl genetischer Varianten innerhalb der fünf CdLS-Gene erweitern. Besonders hervorzuheben sind hierbei sogenannte somatische Mosaik-Varianten, die oftmals nicht innerhalb der bis dato etablierten Routine-Diagnostik unter Anwendung von Sanger-Sequenzierung von Lymphozyten-DNA identifiziert werden konnten. Durch die Verwendung sensitiver Sequenzierungstechnologien wie der Pyrosequenzierung oder SNaPshot Primer Extension-Analyse und der Analyse von DNA-Proben aus Urin, Hautbiopsien und Abstrichen der Mundschleimhaut konnten im Besonderen somatische Varianten in *NIPBL* aber auch im *HDAC8*-Gen nachgewiesen werden. Durch die im Rahmen meiner Promotionsarbeit erzielten Ergebnisse konnten nicht nur neue Erkenntnisse über die Bedeutung somatischer Mosaikmutationen bei der phänotypischen Variation in Patienten mit CdLS generiert werden, sondern es konnte auch aufgezeigt werden, dass ein geringes Maß eines somatischen Mosaiks bereits in nicht-betroffenen Elternteilen nachweisbar sein kann, was direkte Implikationen für die Beratung der Familien und die Errechnung des Wiederholungsrisikos hat. Zusätzlich zur Erweiterung des Mutationsspektrums in den fünf bekannten CdLS-Genen konnten genetische Varianten in *ARID1B*, *SETD5*, *KMT2A* und *ANKRD11*, die für verschiedene Regulatoren der Transkription

bzw. der Chromatinorganisation codieren, identifiziert werden. Interessanterweise konnte eine der Varianten im *ANKRD11*-Gen ebenfalls als somatisches Mosaik beschrieben werden.

Zusätzlich zu Analysen genetischer Varianten innerhalb des Protein-codierenden Anteils des humanen Genoms fokussierten sich ein großer Teil meiner Arbeiten auf funktionelle Analysen des nicht-codierenden Anteils des Genoms. Hierbei wurden genetische Varianten innerhalb regulatorische Elemente analysiert, die einen direkten Einfluss auf die Expression spezifischer Gene haben. Durch die Anwendung einer Vielzahl molekularbiologischer Techniken wie *Chromatin Conformation Capture*, *Genome-Editing*, Reportergen-Analysen konnten wir ein *Enhancer*-Element identifizieren, welches die Aktivität des *NIPBL*-Promotors reguliert. Zusätzlich konnten wir zeigen, dass der *NIPBL*-Promoter die Expression einer nicht codierenden (long non-coding) RNA reguliert, welche bidirektional zum *NIPBL*-Gen transkribiert wird. Die Identifizierung funktioneller Elemente, welche die Expression des *NIPBL*-Gens regulieren, ist nicht nur wichtige Grundlage für das Verständnis komplexer Regulationsmechanismen, sondern ermöglicht möglicherweise die Entwicklung neuer Therapieansätze zur Modulation der *NIPBL*-Expressionslevel.

In einem weiteren Projekt wurde in einem Patienten mit einem CdLS-ähnlichen Phänotyp mittels Array-CGH eine Deletion auf Chromosom 9 identifiziert. Diese Deletion wurde in zwei weiteren Betroffenen innerhalb einer Familie und dem Indexpatienten mittels Bruchpunkt-überspannender PCR bestätigt. Durch verschiedene *in silico* Analysen konnte innerhalb dieser Region ein putatives regulatives Element identifiziert werden. In sich daran anschließenden Untersuchungen wie Promotor-Enhancer-Reportergenanalysen, Expressionsstudien in Zellen der Patienten oder mittels CRISPR/Cas9 generierten Zellmodellen konnte aufgezeigt werden, dass ein ungefähr 3 kb großes regulatives Element innerhalb des deletierten Bereiches die Expression des circa 1 Mb entfernten *SMC2*-Gens reguliert. Das SMC2-Genprodukt ist struktureller Bestandteil des Kondensin-Komplexes. Kondensin hat strukturelle und funktionelle Ähnlichkeiten mit Cohesin, was die phänotypischen Gemeinsamkeiten dieser Patienten mit CdLS erklären könnte. Im Folgenden identifizierten wir zwei weitere Patienten, die ähnliche phänotypische Merkmale aufzeigten und eine heterozygote *Frameshift*-Variante innerhalb des *SMC2*-Gens aufweisen. Sich anschließende funktionelle Untersuchungen zeigten eine gemeinsame Fehlregulation verschiedener Kondensin-Gene sowohl in Patientenzellen als auch in mittels CRISPR/Cas9 generierten Zellmodellen. Die im Rahmen

meiner Promotion erzielten Ergebnisse zeigten Veränderungen der Expression und Funktion von Kondensin-Genen als ursächlich für ein CdLS-ähnliches Krankheitsbild. Zusätzlich konnten neue Einblicke in die Funktion des Kondensin-Komplexes bei der Regulation der Genexpression gewonnen werden. Diese überlappenden molekularen Mechanismen des Cohesin- und Kondensin-Komplexes innerhalb der Transkriptionsregulation bietet somit eine mögliche physiologische Erklärungsgrundlage der phänotypischen Ähnlichkeiten der hier beschriebenen Patienten mit dem Cornelia de Lange-Syndrom.

Im Rahmen meiner Promotionsarbeit wurden verschiedene genetische und funktionelle Untersuchungen genutzt, um die molekularen Mechanismen von CdLS und CdLS-überlappenden Phänotypen aufzuklären. Die Erweiterung des Spektrums von Mutationen in bereits bekannten, sowie neuen Genen, die Identifizierung und Charakterisierung somatischer Mosaike und deren Relevanz für die phänotypische Variabilität und für die Abschätzung des Wiederholungsrisikos bei ratsuchenden Paaren, hat zu einer Verbesserung der molekularen Diagnostik und genetischen Beratung geführt. Zusätzlich zeigen unsere Daten die Wichtigkeit funktioneller regulatorischer Elemente innerhalb nicht-codierender Bereiche des Genoms. Während sich die humangenetische Forschung der letzten Jahrzehnte primär auf den protein-codierenden Bereich des humanen Genoms konzentriert hat, ist die klinische Relevanz der verbleibenden 98 % des nicht-codierenden Bereiches größtenteils unbekannt. Die schnelle Entwicklung neuer Sequenzierungstechnologien ermöglicht uns in Zukunft, diesen großen Teil des Genoms zu verstehen, um nicht nur genetische Ursachen für humane Erkrankungen zu identifizieren, sondern auch phänotypische Variabilität zu verstehen und bessere Vorhersagen zum Phänotyp und/oder zum Krankheitsverlauf geben zu können.

1 Introduction

Cornelia de Lange syndrome (CdLS, OMIM #122470, 300590, 610759, 300882 and 614701) is a clinically and genetically heterogeneous rare developmental disorder. It is dominantly inherited and almost all disease causing genetic variants occur *de novo*. The exact incidence is unknown but estimated to be 1 in 10,000 – 30,000 live births.

In 1933, Dutch pediatrician Cornelia de Lange (1871-1950) reported two unrelated infant girls with similar clinical features such as severe intellectual disability, multiple abnormalities of the skull, face and extremities, and named the disease ´Typus Degenerativus Amstelodamensis´, after the city in which she worked. Earlier in 1916, Winfried Robert Clemens Brachmann (1888-1969) had described a child with similar features [de Lange, 1933; Brachmann, 1916]. Although some examples in the literature refer to the disorder as Brachmann-de Lange syndrome it is more widely referred as Cornelia de Lange syndrome.

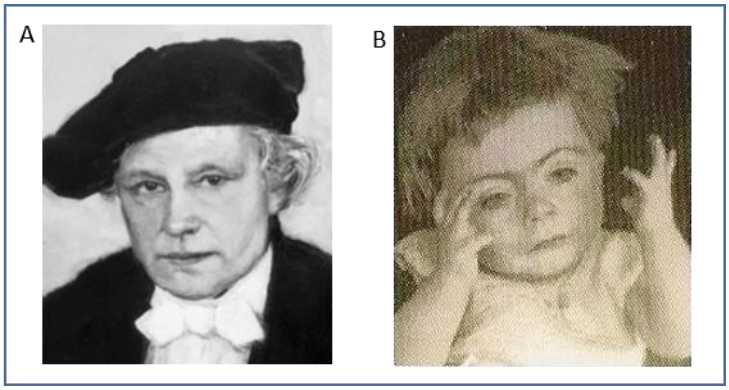


Figure 1: Dr. Cornelia de Lange (A) and her first patient (B), described in 1933

1.1 Clinical features

Although rare, CdLS is widely known and clinically well described. While almost any organ system can be affected, most commonly involved are neurodevelopmental, craniofacial, gastrointestinal and musculoskeletal systems. Diagnosis can be established clinically as well as molecularly, and the treatment is directed towards the specific symptoms that are apparent in each individual.

Affected individuals often show striking prenatal and postnatal growth retardation, typically with a microcephaly. Facial features are the clinical hallmark of the syndrome with a hirsute

forehead, synophrys, arched eyebrows, thick and long eyelashes, low-set and/or posteriorly rotated ears, broad or depressed nasal bridge and anteverted nares. The philtrum is often long and prominent, and the mouth is typically with a thin upper lip and down-turned corners.

Figure 2: Typical facial features of some CdLS patients, including a milder case on the left, and a severe patient on the right (adapted from Liu and Krantz, 2009)

Appearance of the extremities can help to establish the clinical diagnosis. Hands and feet are small in more than 90% of the cases, with disproportionate shortening of the first metacarpal, proximately placed thumb, brachydactyly and fifth finger clinodactyly. Nearly one third of the patients show malformations of the upper limbs, often asymmetrically, which range from small hands to oligodactyly, ulnar deficiency and absent forearm (figure 3). Besides the upper limbs deficiencies, orthopedic manifestations include hip dislocation or dysplasia, scoliosis and delayed maturation of the bones. Lower limbs are less involved with occasional clubbed foot deformity and syndactyly of 2nd and 3rd toes [Halal and Preus, 1979; Jackson et al., 1993; Mehta et al., 2016].

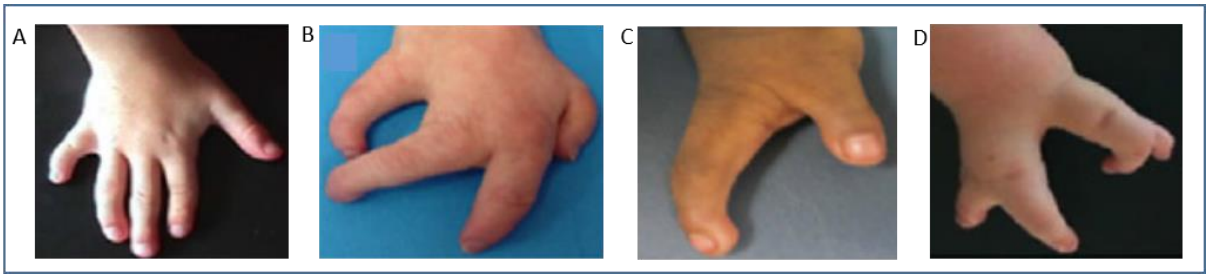


Figure 3: Malformations of the upper limbs in CdLS, including hypoplastic 5th finger (A), absent 5th finger (B), absent 3rd, 4th and 5th finger with ulnar hypoplasia (C) and oligodactyly (D) (adapted from Mehta et al., 2016)

There is general hirsutism in a large number of the patients, noticeable on the face, neck and extremities. In addition, some of the patients present with cleft palate, dental problems and hearing loss.

Multiple organ systems can be affected in CdLS. Feeding problems are typical in infancy and early childhood, with gastroesophageal reflux disease (GERD) being rather common. Congenital heart malformations, renal malformations and hypoplastic genitalia are also quite common. Many patients display emotional changes, self-injury, obsessive-compulsive behavior, anxiety and aggression. Development is mostly delayed with IQ ranging from borderline values with learning difficulties to profound mental retardation. Speech and language abilities are very often affected but acquiring of new skills occurs throughout life without regression [Kline et al., 2007].

Initial diagnosis of CdLS is primarily based on characteristic clinical features. In approximately 75% of patients, clinical diagnosis can be confirmed on genetic level. In addition to the classical phenotype that includes more severely affected patients, cases with milder phenotype have constantly been reported. Based on specific diagnostic systems proposed previously [Halal and Preus, 1979; Allanson et al., 1997; Clinical Advisory Board of the CdLS Foundation USA (CAB) and Scientific Advisory Committee of the World CdLS Federation (SAC)], Kline and colleagues have suggested minimal diagnostic criteria [Kline et al., 2007].

1.2 Genetic aspects

To date, disease-causing genetic variants in five different genes encoding either structural components (*SMC1A*, *SMC3*, *RAD21*) or regulators (*NIPBL*, *HDAC8*) of the cohesin complex have been identified in at least 75% of CdLS cases.

The first described gene was *NIPBL* [Krantz et al., 2004; Tonkin et al., 2004] and heterozygous mutations in *NIPBL* still remain the main cause of the disease with approximately 65-70% patients with a variant in this gene. The second and third most frequently mutated are the X-linked genes *SMC1A* and *HDAC8*. *SMC1A* is mutated in ~5% of the cases [Musio et al. 2006; Deardorff et al., 2007; Huisman et al., 2017] and *HDAC8* in ~4% [Deardorff et al., 2012b; Kaiser et al., 2014; Deardorff et al., 2016]. Finally, *SMC3* and *RAD21* are mutated in approximately 1% of the CdLS cases [Deardorff et al., 2007; Deardorff et al., 2012a].

1.2.1 *NIPBL*

NIPBL consists of 47 exons and encodes at least for two isoforms of delangin. Isoform A contains 2804 and B 2697 amino acids [Krantz et al., 2004; Tonkin et al., 2004]. NIPBL contains various domains including an N-terminal MAU2 interaction domain, a glutamine-rich, a predicted coiled coil, an undecapeptide repeat domain, nuclear localization signal (NLS) and five heat repeats, that represent putative platforms for interactions with other proteins [Neuwald and Hirano, 2000; Yan et al., 2006] (figure 4).

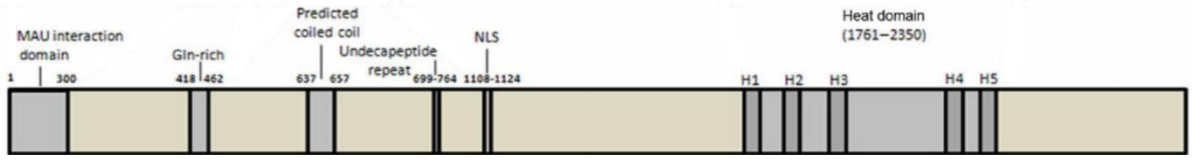


Figure 4: NIPBL protein and its domains: MAU2 interaction domain (1-300), glutamine rich domain (418-462), predicted coiled coil (637-657), undecapeptide repeat (699-764), nuclear localization signal (NLS, 1108-1124) and HEAT domain (1761-2350), consisting of five repeats (adapted from Mannini et al., 2013)

NIPBL and its orthologues, *Nipped-B* in *Drosophila*, *Scc2* in *Saccharomyces cerevisiae* and *Xscc2* in *Xenopus* encode a protein that acts as a regulator of the cohesin complex and is necessary for loading cohesin onto chromatin [Ciosk et al., 2000; Gillespie and Hirano, 2004; Rollins et al., 2004]. Although *NIPBL* depletion in *Xenopus* and HeLa cells causes defects in sister chromatid cohesion and association of cohesin with chromatin [Gillespie and Hirano, 2004; Watrin et al., 2006] analysis performed on lymphoblastoid cell lines obtained from CdLS patients revealed no obvious cohesion defects but altered binding to DNA [Castronovo et al., 2009; Revenkova et al., 2009]. Interestingly, cells of CdLS patients with heterozygous loss-of-function mutations in *NIPBL* and cells from heterozygous *Nipbl* knockout mice showed reduction of mRNA levels only to ~70%, instead of the expected 50% [Liu et al., 2009a; Kawauchi et al., 2009; Kaur et al., 2016]. In addition, defects in NIPBL might lead to transcriptional changes and altered expression of developmentally important genes, causing a CdLS phenotype [Zuin et al., 2014b].

Various types of mutations have been described in *NIPBL* affecting coding sequences, non-coding regions and gross genomic alterations. Most of the mutations are unique but some have been identified in more than one unrelated patients. The majority of these mutations

are nonsense, splice site or frameshift variants, resulting in a truncated protein and loss of NIPBL function, indicating haploinsufficiency as an underlying mechanism for CdLS [Mannini et al., 2013]. Protein truncating mutations tend to be associated with severe phenotypes and a high frequency of limb reductions, while missense mutations tend to cause a milder phenotype (figure 5).

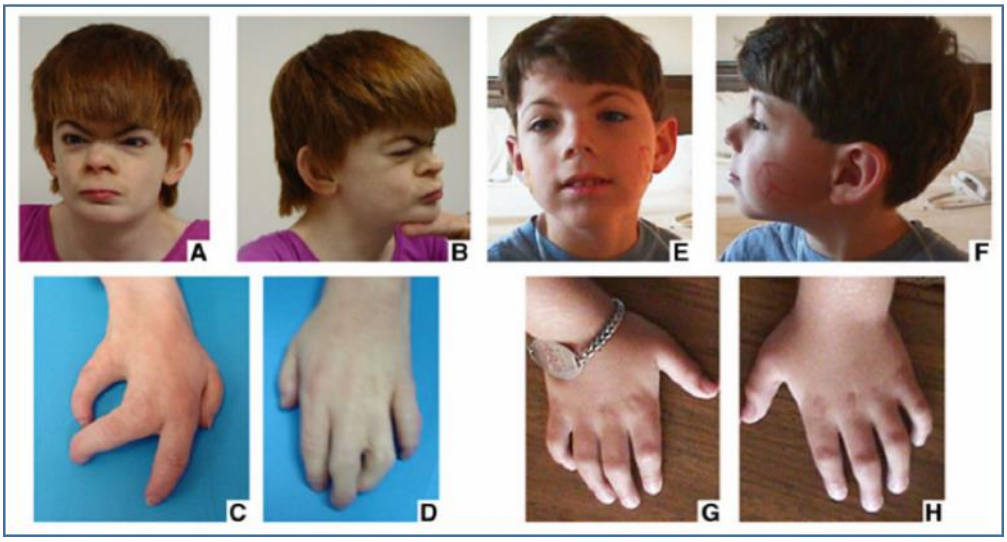


Figure 5: A-D: 28-year-old girl with a truncating mutation in *NIPBL*. E-H: 7-year old boy with a missense mutation in *NIPBL* (adapted from Mannini et al., 2013)

Although most of the *NIPBL* mutations are *de novo*, some familial cases have been reported [Gillis et al., 2004], as well as few cases of gonadal mosaicism [Niu et al., 2006; Weichert et al., 2011; Slavin et al., 2012]. In addition, Huisman and colleagues showed that somatic mosaicism in *NIPBL* is a frequent phenomenon. Namely, 23% of individuals with clinically diagnosed CdLS negative for a mutation in lymphocytes DNA by Sanger sequencing were found to carry a mutation in buccal mucosa DNA [Huisman et al., 2013].

1.2.2 *SMC1A*

SMC1A consists of 26 exons and encodes for a protein of approximately 143 kDa size. It is located on the chromosome X in a region that partially escapes X-inactivation in humans. Both hemizygous male and heterozygous female individuals have been reported with a male-to-female ratio 1:2 [Mannini et al., 2010; Gervasini et al., 2013]. It has been shown that the

SMC1A protein levels are approximately 50% higher in females than in males but there are no significant differences in the total amount of the SMC1A protein comparing healthy females and SMC1A patients [Liu et al., 2009b; Parenti et al., 2014]. Until recently, all the reported variants in this gene connected with CdLS were missense mutations or small in frame deletions supporting the hypothesis that truncating mutations in *SMC1A* are lethal [Liu et al., 2009b].

The CdLS phenotype caused by *SMC1A* variants overlaps to some extend with the phenotype of individuals with *NIPBL* variants. Individuals with *SMC1A* variants were originally reported to have less marked facial features, less effects on growth and less severe limb malformations [Musio et al., 2006; Deardorff et al., 2007]. Subsequent publications have reported a more variable phenotype [Ansari et al., 2014; Gervasini et al., 2013; Hoppman-Chaney et al., 2012].

Figure 6: Facial appearance of two individuals with CdLS, carrying a missense mutation (F and F´) or a small in-frame deletion (G and G´) in *SMC1A* (adapted from Gervasini et al., 2013)

1.2.3 *HDAC8*

HDAC8 catalyzes deacetylation of SMC3 for its removal from chromatin to enable proper dissolution of pro-cohesive elements and allow recycling of cohesin for the next cell cycle. Loss of HDAC8 activity results in an increased SMC3 acetylation and inefficient dissolution of the ´used´ cohesin from chromatin, leading to a CdLS phenotype [Deardorff et al., 2012b]. In recent years, a number of children with clinical features resembling those of CdLS have been found to have loss-of-function missense or nonsense mutations in *HDAC8*. Although facial features of these patients certainly overlap with those of the patients caused by *NIPBL* mutations, there are often some notable differences. Typical features for individuals with *HDAC8* mutations include abnormal skull formation and delayed closure of the anterior fontanelle, hypertelorism, telecanthus and/or forehead nevus flammeus. Notably, Hdac8 knockout mouse demonstrates extreme growth failure and markedly delayed closures of skull

structures [Haberland et al., 2009]. Additional facial features in individuals with *HDAC8* mutations include hooding of the eyelids and dental anomalies, suggesting that facial features might be useful in defining the phenotype more precisely (figure 7). In addition, these individuals have a happy cheerful behavior in contrast to the ´classical´ CdLS phenotype that includes aggressive and self-injurious behavior.

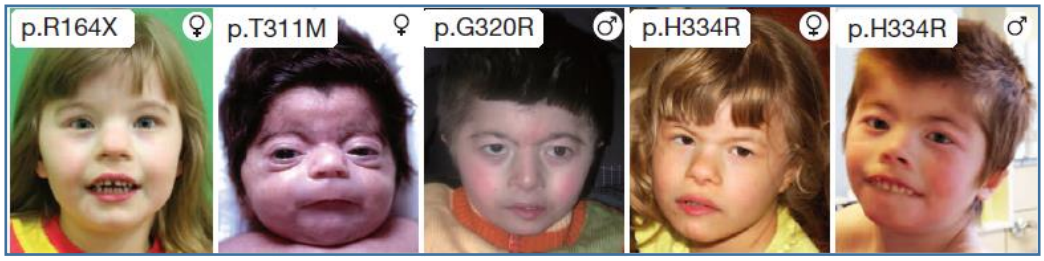


Figure 7: Facial features of individuals with *HDAC8* mutations, labeled with the corresponding mutation and sex (adapted from Deardorff et al., 2012b)

To date, nearly 60 patients with variants in *HDAC8* have been reported and nearly 2/3 of the reported patients are females. Located on the long arm of the chromosome X, *HDAC8* is among the genes that are subject to X-inactivation in females. This results in only one allele either wild type or mutant being expressed in most cells. Enzymatic assays of all reported missense mutations demonstrated reduced or absent activity, providing important insights into biological function of HDAC8 catalysis and human developmental disorders [Kaiser et al., 2014; Decroos et al., 2014; Decroos et al., 2015]. It has been reported that derivatives of N-acylthiourea act as selective activators of low-activity HDAC8, and it appears that N-(phenylcarbamothioyl)benzamide can restore catalytic activity of compromised HDAC8 mutants such as those identified in CdLS, providing a potential treatment strategy [Singh et al., 2011; Deardorff et al., 2016].

1.2.4 *SMC3*

SMC3 consists of 29 exons encoding for a protein of approximately 141 kDa that forms a heterodimer with SMC1A. Until recently, only one patient with a 3bp deletion in *SMC3* has been reported as a 'mild CdLS' phenotype [Deardorff et al., 2007]. Subsequently, a *SMC3* missense mutation was reported within a large cohort of individuals with autism spectrum

disorder, albeit without a clinical description [Sanders et al., 2012]. The low frequency of *SMC3* mutation detection might suggest that this gene may play important roles beyond cohesion [Mannini et al., 2013]. This is supported by the findings that SMC3 acetylation controls fork processing in human cells. In addition, cohesion of sister chromatids is established during S phase by SMC3 acetylation at tandem lysine residues K105 and K106, catalyzed by N-acetyltransferases ESCO1 and ESCO2 [Zhang et al., 2008; Terret et al., 2009]. On the other hand, SMC3 deacetylation by HDAC8 is involved in cohesin recycling during the cell cycle and increased SMC3 acetylation leads to inefficient dissolution of cohesin from chromatin in both prophase and anaphase [Deardorff et al., 2012b]. As *SMC3* is a central determinant of these processes, it is likely that its mutations are negatively selected and that only mutations in specific regions will result in a CdLS phenotype.

Figure 8: A 57-year old man with a 3bp deletion in *SMC3*, also as a teenager in ´Q´ (adapted from Mannini et al., 2013)

1.2.5 *RAD21*

RAD21 plays an important role in the structure and function of the cohesin complex as it physically connects SMC1A/SMC3 heterodimer with STAG and regulates association and dissociation of cohesin with chromatin [Nasmyth and Haering, 2009]. In 2012, Deardorff and colleagues have reported six individuals with mutations in *RAD21* with mild but overlapping clinical features with CdLS. They have also speculated that the mild nature of this

cohesinopathy could lead to an underestimation of the diagnosis. Cell lines expressing mutant RAD21 showed lower cell survival after irradiation and higher levels of chromosomal rearrangements compared to wild type cells [Deardorff et al., 2012a].

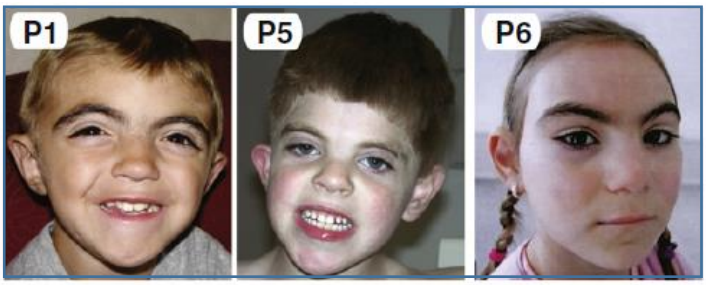


Figure 9: Facial features of the children with *RAD21* mutations (adapted from Deardorff et al., 2012a)

1.3 The cohesin complex

Cohesin is a multi-subunit protein complex, well conserved throughout evolution and crucial for cell survival. It is ring-shaped and consists of the four core subunits: SMC1A, SMC3, RAD21 and STAG. SMC (Structural Maintenance of Chromosomes) proteins are long polypeptides that fold back on themselves by antiparallel coiled-coil interactions to produce molecules with a ´hinge´ domain at one end and ATPase ´head´ on the other end. The hinge domains of SMC1A and SMC3 are tightly bound together while their heads are bridged by RAD21 (figure 10). Loading of cohesin complexes on chromosomes is facilitated by the heterodimeric Kollerin complex consisting of NIPBL and MAU2, orthologs of Scc2 and Scc4 in yeasts [Ciosk et al., 2000]. Cleavage of the RAD21 subunit at the onset of anaphase releases cohesin from chromatin and allows chromosome segregation. HDAC8 deacetylates SMC3 in vertebrates and helps dissolution of the ´used´ cohesin from chromatin [Deardorff et al., 2012b].

The cohesin complex was originally discovered as a protein complex required for sister chromatid cohesion, a function that is conserved in all eukaryotes [Michaelis et al., 1997]. Besides the canonical function in sister chromatid cohesion, other roles of cohesin include DNA repair, chromatin modification, DNA condensation, regulation of gene expression and structural organization of the genome.

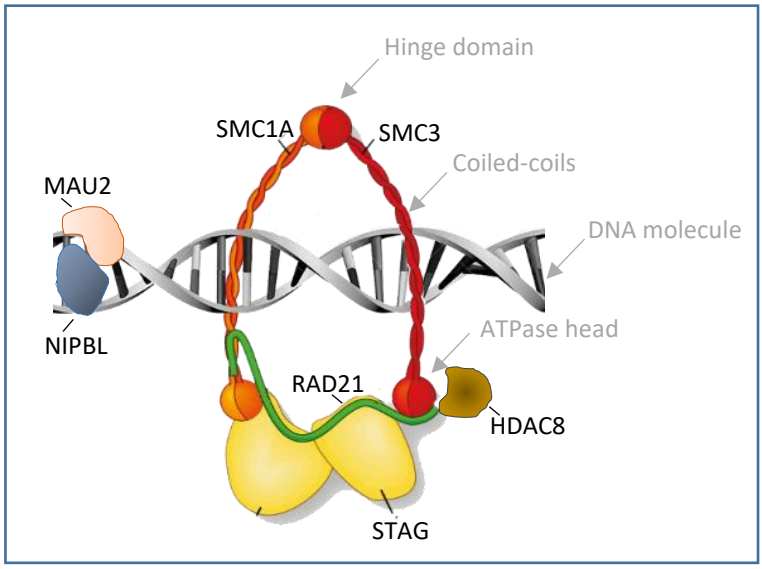


Figure 10: Subunits and regulators of the cohesin complex, significant in CdLS (adapted from Haering and Gruber 2016)

The cohesin complex was originally discovered as a protein complex required for sister chromatid cohesion, a function that is conserved in all eukaryotes [Michaelis et al., 1997]. Besides the canonical function in sister chromatid cohesion, other roles of cohesin include DNA repair, chromatin modification, DNA condensation, regulation of gene expression and structural organization of the genome.

1.3.1 Sister chromatid cohesion

DNA replication in synthesis (S) phase of the cell cycle creates two identical DNA molecules, called sister chromatids. After nuclear envelope breakdown in eukaryotic cells the chromosomes start to condense obtaining the specific mitotic shape when each of the sister chromatids goes to an opposite spindle pole of the mother cell enabling its subsequent division into two genetically identical daughter cells. This physical connection between the replicated sister chromatids that lasts from S phase of interphase until anaphase of mitosis, called sister chromatid cohesion (SCC), opposes the pulling forces that are generated by microtubules. Without cohesion, sister chromatids would not be segregated symmetrically between the forming daughter cells which would result in aneuploidies. Therefore, the

integrity of the cohesin complex is crucial for the proper segregation of the mitotic chromosomes [Gruber et al., 2003; Losada and Hirano, 2005; Nasmyth and Haering, 2005]. Different models propose the way cohesin binds sister chromatids either by topologically embracing both of them [Ivanov and Nasmyth, 2005; Haering et al., 2008] or binding a single DNA molecule, where then the two cohesin rings interact to connect the two sister chromatids [Huang et al., 2005; Zhang et al., 2008].

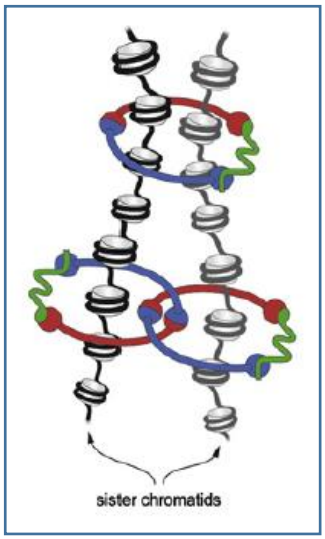


Figure 11: The two models for chromosome organization by cohesin: either the two sister chromatids are entrapped within the same ring or two cohesin rings, each encircling a single sister chromatid, interact to provide sister chromatid cohesion (adapted from Haering and Jessberger, 2012)

1.3.2 DNA repair

In addition to the canonical role of cohesin in mediating sister chromatid cohesion, cohesin is also essential during G2 phase of the cell cycle to allow repairing of double-stranded DNA breaks by homologous recombination. For homologous recombination to occur the two sister chromatids have to be brought in the close proximity where the intact one will serve as a template for repairing the damaged one. This proximity is thought to be established by cohesin [Watrin and Peters, 2006]. Although it remains unclear how cohesin is recruited to the sites of double-stranded DNA breaks there is evidence that SMC1A is phosphorylated in response to ionizing radiation, which facilitates cohesin mobilization to these sites [Kim et al., 2002; Yazdi et al., 2002; Musio et al., 2005].

1.3.3 Chromatin modification

In eukaryotes, the basic units of DNA packaging are nucleosomes consisting of DNA segments wrapped around histone octamers connected by stretches of ´linker´ DNA. In this way, large eukaryotic genomes are packed within the nucleus still ensuring appropriate access to the DNA and regulation of gene expression. Dynamic modification of chromatin structure that allows access of condensed genomic DNA to the regulatory transcription machinery proteins is called chromatin remodeling. It is carried out either by covalent histone modification enzymes or by ATP- dependent chromatin remodeling complexes. In 2002, Hakimi and colleagues reported that SNF2h, a component of the SWI/SNF chromatin remodeling complex interacts with RAD21 subunit of the cohesin complex [Hakimi et al., 2002], suggesting that SNF2h mediates association of cohesin and chromatin. It has also been hypothesized that chromatin structure plays a role in determining whether and where cohesin binds to chromosomes in eukaryotic cells [Riedel et al., 2004]. In addition, histone deacetylases 1 and 3 (HDAC1 and HDAC3) were described to interact with the cohesin loading protein NIPBL [Jahnke et al., 2008], while HDAC8 deacetylates SMC3 component of the cohesin complex [Deardorff et al., 2012b].

1.3.4 DNA condensation

Chromosome condensation is one of the key events in chromosome morphogenesis required for their faithful segregation in mitosis. Condensation of mitotic chromosomes is mainly mediated by condensin, a pentameric protein complex evolutionary similar to cohesin. The balance between SCC and condensation is an essential determinant in shaping mitotic chromosomes in eukaryotic cells [Losada and Hirano, 2001]. Various studies suggest that cohesin can influence condensin localization on the chromosomes. In support of this, it has been demonstrated in budding yeast that chromosome condensation occurs in two steps and that the first step depends on cohesin [Lavoie et al., 2004]. In addition, D´Ambrosio and colleagues have found that condensin DNA binding sites overlap with binding sites of cohesin-loading Kollerin complex [D´Ambrosio et al., 2008]. This implies that condensin and cohesin are recruited to the same sites, at least in budding yeast. However, precise molecular mechanisms of interaction between cohesin and condensin are still unclear.

1.3.5 Regulation of gene expression and 3D genome architecture

A first evidence that cohesin might play a role in transcriptional regulation was published by Rollins and colleagues, who reported that *Drosophila* mutant flies carrying a variant in *Nipped-B* (ortholog of mammalian *NIPBL*) show alterations in activation of homeobox genes [Rollins et al., 1999]. In following studies, involvement of cohesin loader and cohesin subunits in regulation of gene expression has been demonstrated in other model organisms such as zebrafish, budding yeast, fission yeast, mice etc. The mechanism by which cohesin influences transcription is still not fully understood and different mechanisms have been postulated. In *Drosophila*, cohesin co-localizes with Nipped-B throughout the entire non-repetitive genome, transcribed regions and overlapping with RNA polymerase II [Misulovin et al., 2008]. In mammals, besides co-localizing with NIPBL, cohesin also accumulates at the sites where CCCTC-binding factor (CTCF) binds [Wendt et al., 2008; Parelho et al., 2008]. Therefore, these two distinct types of cohesin sites might mediate cohesin´s roles in transcription: strong sites coincide with the binding of CTCF, whereas weaker sites map to active promoters and enhancers. At weaker sites cohesin is co-localized with its loader NIPBL, Mediator complex and tissue-specific transcription factors. However, NIPBL was not found at CTCF sites [Kagey et al., 2010; Merkenschlager and Odom, 2013; Mehta et al., 2013].

Cohesin/CTCF sites are involved in long-range interactions of the chromatin fibers, mediating contacts between promoters and distal enhancers and gene activation at number of developmentally important loci [Hadjur at al., 2009; Nativio et al., 2009]. Cohesin deficient of CTCF also mediates long-range interactions and transcriptional activation, likely depending on the co-binding of specific additional transcription factors. These long-range interactions are mediated by DNA looping bringing together loci distant on the linear scale (figure 12).

Genome-wide chromatin interaction maps provided by the Hi-C technique show an organization of chromatin into discrete domains named topologically associated domains or TADs [Dixon et al., 2012]. TADs are conserved through cell types and species, confining the interactions between enhancers and their target promoters. It is worth noting that both cohesin and CTCF are enriched at TAD boundaries, suggesting their role as architectural proteins in defining TADs. Zuin and colleagues proposed that cohesin is mainly involved in chromatin interactions within a TAD, while CTCF is important for their spatial segregation [Zuin et al., 2014a].

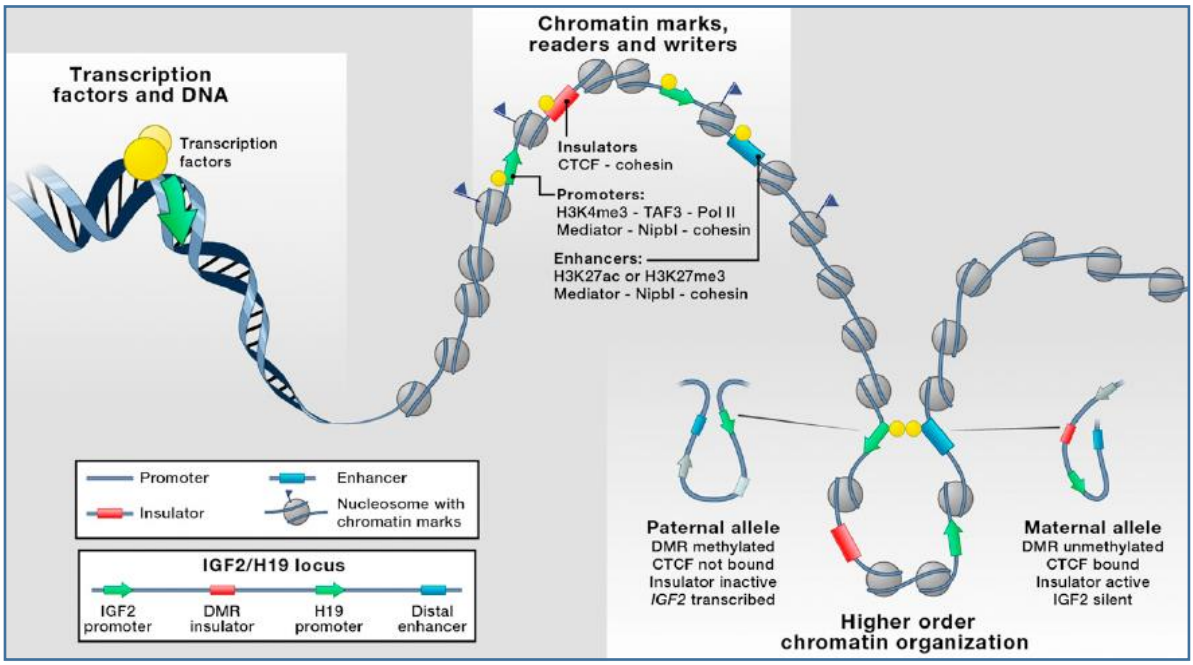


Figure 12: Chromatin modifications and DNA looping, mediated by various transcription factors and proteins including CTCF, cohesin, NIPBL, and Mediator complex. The picture in the middle represents two different cohesin-binding sites: strong sites that correspond to cohesin/CTCF interactions and their insulating roles, and weaker sites where cohesin co-localizes with NIPBL and Mediator. The figure shows regulatory elements of the imprinted *IGF2/H19* locus, as it is well studied (adapted from Merkenschlager and Odom, 2013)

Within a TAD, enhancers have access to and can interact with their receptive promoters. A recent study has shown that rather than enhancers having an inherent specificity to their cognate promoters, this communication is at least partially driven by the CTCF-mediated chromatin architecture [Hanssen et al., 2017]. According to the authors, it seems likely that intergenic variants within critical CTCF-cohesin binding site will underlie changes in gene expression associated with wide variety of complex traits and diseases. Two additional very recent studies have shown that depletion of cohesin complex from either non-dividing mouse hepatocytes or HCT-116 human colon cancer cells causes disappearance of TADs genome wide and dysregulation of certain genes and non-coding regions. On the other hand, genome compartmentalization to active (A) and inactive (B) chromatin is still preserved, indicating that chromatin has an intrinsic tendency to form compartments based on epigenetic landscape and transcriptional activity, but cohesin interferes with these subdivisions by bringing together loci with opposite states [Schwarzer et al., 2017; Rao et al., 2017].

1.4 The condensin complex

Another complex structurally similar to cohesin is condensin. This complex plays important roles in the cell, overlapping with cohesin.

Condensin is a large pentameric protein complex consisting of a SMC2/SMC4 heterodimer and three non-SMC subunits: CAP-H, CAP-D2 and CAP-G are additional components of the condensin I, while CAP-H2, CAP-D3 and CAP-G2 belong to the condensin II complex (figure 13).

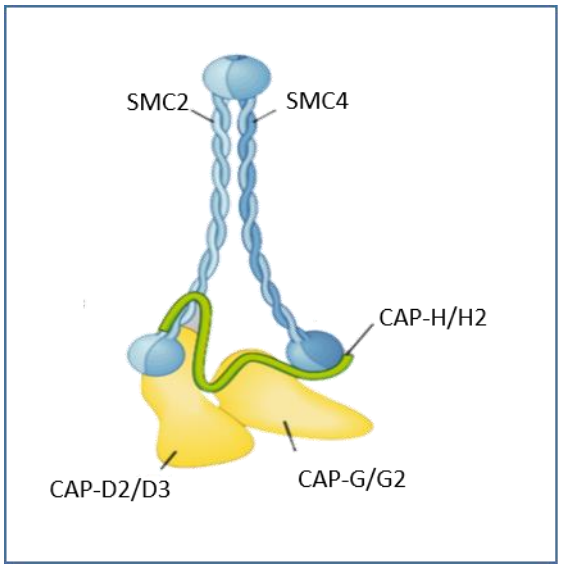


Figure 13: Subunits of the condensin complex. SMC2 and SMC4 proteins compose both condensin I and condensin II. However, these two complexes differ in their non-SMC subunits. In the picture are shown components of condensin I and condensin II, respectively (adapted from Haering and Gruber 2016)

This complex is highly conserved throughout evolution and plays important roles in many cellular processes such as proper condensation and segregation of mitotic chromosomes, 3D organization of genome architecture throughout the cell cycle, regulation of transcription and DNA repair [Hirano and Mitchison, 1994; Saitoh et al., 1994; Hudson et al., 2003; Frosi and Haering, 2015].

Despite structural similarities between condensin I and condensin II their cellular localizations and functions differ. Condensin I is located in cytosol and gains access to chromosomes after nuclear envelope breakdown when it contributes to lateral compaction of chromosomes. On

the contrary, condensin II is located in nucleus during interphase and contributes to the axial shortening of chromosomes [Hirota et al., 2004; Shintomi and Hirano 2011]. Besides being enriched on the loci important for chromosome segregation, condensin I is enriched at promoter regions of highly transcribed genes, tRNA and rRNA genes [Kim et al., 2013].

Mutations in genes encoding for components of the condensin complex have been associated with gastric, colorectal and lung cancer [Davalos et al., 2012; Je et al., 2014; Zhang et al., 2016], T-cell lymphoma [Woodward et al., 2016] and microcephaly [Martin et al., 2016].

1.5 Outline

Mutations in the five genes, *NIPBL*, *SMC1A*, *HDAC8*, *SMC3* and *RAD21* can be identified as genetic cause in at least 75% of the patients with a clinical diagnosis of CdLS. However, the remaining ~25% of the CdLS cases are still unsolved, and this percentage was even higher at the time I started my PhD studies (2013/2014).

Some of the previously unsolved cases were found to carry mosaic mutations in CdLS genes, mainly in *NIPBL*, that are not detectable by conventional Sanger sequencing using DNA of peripheral blood [Huisman et al., 2013]. Recently, variants in other chromatin regulators such as *ANKRD11*, *AFF4*, *EP300*, or *KMT2A* were described in some patients with CdLS [Ansari et al., 2014; Woods et al., 2014; Izumi et al., 2015; Yuan et al., 2015]. Notably, mutations in all these genes have already been described as genetic causes of other developmental disorders that share phenotypical features with CdLS, such as KBG syndrome, CHOPS, Rubinstein-Taybi, or Wiedemann-Steiner syndrome. Furthermore, a direct interaction between Scc2, yeast ortholog of NIPBL and SWI/SNF chromatin remodeling complex pertains to the similarities between CdLS and Coffin-Siris syndrome [Lopez-Serra et al., 2014]. Recently, biallelic variants in non-SMC subunits of the condensin complex have been reported to cause microcephaly, short stature and intellectual disability [Martin et al., 2016]. However, variants in SMC subunits of the condensin complex have still not been connected with any distinguishable phenotype.

1.6 Objectives of the present work

The first aim of the present work is to find molecular causes in patients clinically diagnosed as CdLS and to functionally characterize identified variants. The second aim is identification and molecular characterization of non-coding functional elements in the genome relevant for expression of the ´CdLS genes´. Therefore, the work presented here is divided into the following sections:

- Identification of the causative mutations in the known and new CdLS genes
- Regulation of *NIPBL*
- Identification and characterization of a non-coding regulatory element relevant for expression of condensin gene *SMC2*

2 Materials and methods

2.1 Materials

2.1.1 Reagents

Name	Manufacturer
Acetic acid	Merck
Aluminium sulphate	Sigma Aldrich
Acrylamide (30 %)	Carl Roth
Agar	Fluka Biochem
Agarose	StarLab
Ammonium sulphate	Fluka Biochem
Boric acid	Merck
Bromophenol blue	Sigma Aldrich
BSA (bovine serum albumin)	Sigma Aldrich
Buffer tablets pH 7.2	Merck
Calcium chloride	Serva
Chloroform	Merck
Cycloheximide	Sigma Aldrich
DMEM	Gibco
DMSO (dimethyl sulfoxide)	Sigma Aldrich
DTT (dithiothreitol)	Sigma Aldrich
EDTA	Merck
Ethanol	J. T. Baker
Fetal bovine serum (FBS)	Gibco
FuGENE® HD Transfection Reagent	Promega
Giemsa´s azur eosin methylene blue solution	Merck
Glucose	Merck
Glycin	Sigma Aldrich
HEPES	Sigma Aldrich
Kalium chloride	Fluka Biochem
Kalium dihydrogen phosphate	Merck
Kalium acetat	Sigma Aldrich
KaryoMAX® Colcemid solution in HBSS	Life Technologies
Lipofectamine RNAi max	Invitrogen
Lithium chloride	Sigma Aldrich
Magnesium chloride	Sigma Aldrich
Midori Green	Nippon Genetics
Milk (powder)	Merck
Methanol	J. T. Baker
β-mercaptoethanol	Sigma Aldrich
Natrium chloride	Merck
Natrium chloride, 0.9%	Fresenius Kabi
Natrium hydroxide	Merck
Natrium dodecyl sulphate	Sigma Aldrich
di-Natrium hydrogen phosphate	Merck
Natrium carbonat	Sigma Aldrich

NP-40 (Nonidet P-40)	Fluka Biochem
PEG (polyethylene glykol) 4000	Fluka Biochem
Pepton	Gibco
Phenol	AppliChem
ProLong® Gold antifade reagent with DAPI	Life Technologies
RPMI	Gibco
TEMED	Carl Roth
Tris base	MP Biomedicals
Triton-X-100	Sigma Aldrich
Tryptone	Beckton Dickinson
Tween 20	Sigma Aldrich
qPCRBIO Probe Mix Hi-ROX	PCR Biosystems
Yeast extract	Beckton Dickinson
Yeast nitrogen base	Beckton Dickinson

2.1.2 Devices

Name	Manufacturer
Agarose gel electrophoresis system:	
Electrophoresis chambers	Serva
Dark chamber FAS-Digi	Nippon Genetics
Camera MX-1	Pentax
Incubators:	
Innova 4300 Incubator Shaker	New Brunswick/Eppendorf
Galaxy 170S	New Brunswick/Eppendorf
IN75	Memmert
Microscopes:	
Axioplan 2 (HBO 100)	Carl Zeiss
Axioskop 40	Carl Zeiss
Fluorescent lamp ebq 100 isolated	Kübler
Photometers/spectrophotometers:	
Biospectrometer basic	Eppendorf
µCuvette G1.0	Eppendorf
TriStar^2 LB Multidetection Microplate Reader	Berthold
SDS-PAGE and Western Blot:	
PowerPack Basic Power Supply	Biorad
Trans-Blot® Turbo™ Transfer System	Biorad
Mini-PROTEAN Tetra Cell System	Biorad
Glass plates	Biorad
Thermocyclers:	
Thermal Cycler 2720	Applied Biosystems
Veriti 96 well Thermal Cycler	Applied Biosystems
C1000 Touch Thermal Cycler	Biorad

Centrifugation:	
Centrifuge 5804	Eppendorf
5415C	Eppendorf
5415D	Eppendorf
Microfuge 22R Centrifuge	Beckmann Coulter
Avanti J-30I Centrifuge	Beckmann Coulter
Biofuge 13	Heraeus Sepatech
CombiSpinFVL-2400 Centrifuge/Vortex	PeqLab Biotechnologie GmbH
Sequencing:	
3130 Genetic Analyzer	Applied Biosystems
3130*xl* Genetic Analyzer	Applied Biosystems
PyroMark Q24	Qiagen
Real time PCR:	
7300 Real Time PCR System	Applied Biosystems
Other:	
Autoclave- CS/VFKT	Webeco
Easypet-Pipetboy 9V DC	Eppendorf
Fume hood	Köttermann
Ice machine	Ziegra
Laminar flow hood Hera Safe	Heraeus
Laminar flow hood Heraeus Lamin Air HLB 2448GS	Heraeus
Magnetic stirrer MR3002	Heidolph
Microprocessor pH meter WTW	WTW
Reax Overhead Shaker	Heidolph
Shaker Polymax 1040	Heidolph
Scale Kern	Sartorius
Scale Kern ABJ	Sartorius
Thermomixer comfort	Eppendorf
Thermomixer compact	Eppendorf
Vacuum pump WOB-L PRES/Vac-Dry	Welch
Vortex Top Mix 11118	Fisher Bioblock Scientific
Vortex Genie	Bender & Hobein
Water bath GFL 1003	GFL

2.1.3 Solutions and buffers

If not underlined differently all the solutions were prepared in distilled water (*Aqua dest*).

2.1.3.1 Agarose gel electrophoresis

Name	Composition
Agarose solution	1-2% (w/v) agarose 1x TBE buffer Midori green DNA stain

TBE buffer	445 mM Tris-Base 445 mM boric acid 10 mM EDTA
DNA loading buffer	0.1% bromophenol blue 0.1% xylene cyanol 25% glycerin 10 mM EDTA

2.1.3.2 DNA extraction

Name	Composition
TNE buffer	10 mM Tris, pH 8.0 100 mM NaCl 1 mM EDTA, pH 8.0
TNE plus buffer	10 mM Tris, pH 8.0 100 mM NaCl 1 mM EDTA, pH 8.0 1% SDS 100 µg/ml Proteinase K
Plasmid preparation (´Mini preparation´) buffers	
Mini preparation buffer 1	50 mM Tris-HCl, pH 8.0 10 mM EDTA 100 µg/ml RNase
Mini preparation buffer 2	0.2 M natrium hydroxide (NaOH) 1% SDS
Mini preparation buffer 3	3 M kalium acetate (CH_3COOK)

2.1.3.3 SDS-PAGE and Western blot

Name	Composition
10x SDS running buffer	2 M glycine 1.25 M Tris-Base 1% (w/v) SDS
4x resolving buffer	1.5 M Tris-Base, pH 8.8 0.4 % (w/v) SDS
4x stacking buffer	0.5 M Tris-base, pH 6.8 0.4 % (w/v) SDS

4x loading buffer	62 mM Tris-Base, pH 6.8 2% (w/v) SDS 10% (v/v) glycerin 5% (w/v) DTT 5% (v/v) β-mercaptoethanol 0.025% (w/v) bromophenol blue
Transfer buffer	48 mM Tris-Base 39 mM glycine 0.05% (w/v) SDS 20% methanol
Coomassie, colloidal	10% (v/v) ethanol 2% (v/v) orthophosphoric acid 5% (w/v) aluminium sulphate 0.02% (w/v) Coomassie G250
Destainer for colloidal Coomassie	10% (v/v) ethanol 2% (v/v) orthophosphoric acid
RIPA buffer (Radioimmunoprecipitation assay buffer)	50 mM HEPES 1 mM EDTA pH 8.0 1% Nonidet-P40 (NP-40) 0.5 M lithium chloride (LiCl) pH 7.6 (adjust with NaOH) Working solution is RIPA:PIC (protease inhibitor cocktail) in ratio 250:1
TBST	1 M NaCl 1 M Tris 0.05% Tween

2.1.3.4 Mediums and additives for human cells

Name	Composition
Cycloheximide solution (CHX)	10 mg/ml cycloheximide in DMSO
DMEM 101 (full medium)	1x DMEM 10% FBS, heat inactivated 1% antibiotic/antimycotic solution
Medium for freezing the cells	80% (v/v) DMEM 101 20% (v/v) DMSO
PBS	137 mM NaCl 2.6 mM KCl 6.5 mM Na_2HPO_4 1.5 mM KH_2PO_4

RPMI (full medium)	1x RPMI 1640 medium 10% FBS, heat inactivated 1% antibiotic/antimycotic solution
Trypsin/EDTA	40 g/l trypsin 0.7 mM EDTA in PBS

2.1.3.5 Mediums and additives for bacterial cells

Name	Composition
LB medium	20 g/l tryptone peptone 10 g/l bacto yeast extract 20 g/l NaCl
Agar plates	LB medium 16 g/l agar 40 µg/ml X-gal 100 µg/ml ampicillin or 35 µg/ml kanamycin

2.1.3.6 Buffers for treating chromosomes

Name	Composition
TEEN	1 mM triethanolamine-HCl pH 8.5 0.2 mM NaEDTA 25 mM NaCl
RSB	10 mM Tris pH 7.4 10 mM NaCl 5 mM $MgCl_2$
Giemsa solution	1:5 in buffer pH 7.2 according to WEISE
Propidium iodide staining solution	50 µg/ml propidium iodide (PI) 10 µg/ml RNase A in PBS

2.1.4 Human cell lines

Name	Origin
HEK293	Human embryonic kidney, fibroblast-like
HeLa	Human cervix carcinoma, epithelial, female
1-7HB2	Human mammary epithelial cells
SH-SY5Y	Human neuroblastoma, female
Fibroblasts	Human (patients or healthy controls)

LCLs (lymphoblastoid cell lines)	Human (patients B-lymphocytes infected with Epstein-Barr virus)

2.1.5 Vectors

Name	Manufacturer
pGEM-T Easy (cloning vector)	Promega
pGL4.10 [*luc*2] (eukaryotic expression vector)	Promega
pGL4.74 [*hRluc*/TK]	Promega
pcDNA 3.1 (A, B, C) (eukaryotic expression vector)	Invitrogen
pEGFP-N3 (eukaryotic expression vector)	Takara Clontech
pFLAG-N3 (eukaryotic expression vector)	from Dr. Tarik Möröy (*Department of Medicine, Division of Experimental Medicine*, University of Montreal, Canada)
pSpCas9(BB)-2A-Puro (PX459) V2.0	from Dr. Feng Zhang via Addgene

2.1.6 Oligonucleotides

Primers for sequencing reactions were synthesized by the company Microsynth, while cloning primers were synthesized by Biomers.

TaqMan probes were ordered from Thermo Fisher Scientific as follows: SMC2- Hs00197593_m1; SMC4- Hs00374522_m1; SNAPIN- Hs00276176_m1; GAPDH- Hs99999905_m1; NADH-Hs00190020_m1.

2.1.7 Size markers

Name	Manufacturer
GeneScan™-120 LIZ™ Size Standard	Applied Biosystems
GENESCAN® 400HD ROX® Size Standard	Applied Biosystems
Hyperladder I (DNA)	Bioline
Low DNA Mass ladder (DNA)	Invitrogen
Precision Plus Protein Standard: Dual Color (protein)	Biorad

2.1.8 Enzymes

Name	Manufacturer
ExoSAP-IT®	Affimetrix

Expand Long Range dNTPack	Roche
Fast Digest®	Thermo Fisher Scientific
HpaII	New England Biolabs
RNase A	Qiagen
PrimeSTAR GXL DNA Polymerase	Takara, Clontech
T4 ligase	Roche
Taq DNA polymerase	MP Biomedicals
Taq DNA polymerase	Qiagen
Trypsin	Serva

2.1.9 Kits

Name	Manufacturer
Amaxa® Cell Line Nucleofector Kit V	Lonza
BigDye® Terminator v3.1 Sequencing Kit	Applied Biosystems
Dual-Luciferase® Reporter Assay	Promega
DNeasy Blood & Tissue Kit	Qiagen
DyeEx 2.0 Spin Kit	Qiagen
pGEM-T-Easy-Vector System	Promega
Monarch DNA gel extraction kit	New England Biolabs
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific
Plasmid Plus Midi Kit	Qiagen
Puregene® Buccal Cell Core Kit A	Qiagen
PyroMark Gold Q24 Reagents	Qiagen
ReliaPrep®RNA Cell Miniprep System	Promega
SNaPshot® Multiplex System	Applied Biosystems
Super Script III First Strand Synthesis System	Thermo Fisher Scientific
SuperSignal West Femto Chemiluminescent Substrate	Thermo Scientific
QIAquick Gel Extraction Kit	Qiagen
QuickChange Site directed Mutagenesis Kit	Agilent Technologies

2.1.10 Antibodies

Name	Manufacturer
Primary antibodies:	
α-c-Myc (9B11) (mouse)	Cell Signaling

α-FLAG M2 (monoclonal, mouse)	Sigma Aldrich
α-HDAC8 (H-145) (polyclonal, rabbit)	Santa Cruz
α-SMC1A (ab21583) (polyclonal, rabbit)	Abcam
α-SMC3	from Dr. Erwan Watrin (*Institut de Génétique et Développement de Rennes* Rennes, France)
α-SMC2 (A300-058A) (polyclonal, rabbit)	Bethyl
α-SMC4 (A300-064A) (polyclonal, rabbit)	Bethyl
α-RAD21 (D213)	Cell Signaling
α-β-tubulin (monoclonal, mouse)	Sigma Aldrich
α-acetyl-Histone H3 (polyclonal, rabbit)	Millipore
α-NCAPD2	from Dr. Erwan Watrin (*Institut de Génétique et Développement de Rennes* Rennes, France)
Secondary antibodies:	
Goat-α-rabbit IgG-HRP	Thermo Scientific
Goat-α-mouse IgG-HRP	Thermo Scientific

2.1.11 Antibiotics

Name and working concentration	Manufacturer
Ampicillin (200 µg/ml)	Roth
Antibiotic-antimycotic solution	Life Technologies
Kanamycin (70 µg/ml)	Sigma Aldrich
Puromycin (10 mg/ml)	Life Technologies

2.1.12 Programs and web sites

Name	Manufacturer
Chromas Lite	Technelysium Pty Ltd
Edit Seq	DNA Star Lasergene
ENCODE	http://genome.ucsc.edu/ENCODE/
Ensembl	http://www.ensembl.org/index.html
ExAC Browser	http://exac.broadinstitute.org/
ExPASy	http://www.expasy.org/
Face2Gene	http://suite.face2gene.com/
FlowJo	FlowJo LLC, Ashland, OR
GeneMapper™ Software	Applied Biosystems

Genome Engineering 5.0	http://www.genome-engineering.org/
gnomAD browser	http://gnomad.broadinstitute.org/
Ikaros	MetaSystems, Heidelberg, Germany
Isis	MetaSystems, Heidelberg, Germany
Mutation Discovery	http://www.mutationdiscovery.com
Mutation Taster	http://www.mutationtaster.org/
NebCutter V2.0	http://www.labtools.us/nebcutter-v2-0/
NCBI-Blast	https://blast.ncbi.nlm.nih.gov/Blast.cgi
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/
Seqman	DNA Star Lasergene
Sequencing analysis	Applied Biosystems
SIFT	http://sift.jcvi.org/
Taq-Man Software	Applied Biosystems
UCSC Genome Browser	http://genome-euro.ucsc.edu/

2.2 Methods

2.2.1 DNA extraction

Genomic DNA was extracted from blood, cells in culture (fibroblasts, HEK293, SH-SY5Y), buccal mucosa or urine cells.

DNA extraction from blood and cells in culture was performed with DNeasy Blood & Tissue Kit (Qiagen) and from buccal mucosa with Puregene® Buccal Cell Core Kit A (Qiagen) according to the manufacturer´s instructions. Starting material was 500 µl of whole blood, or ~ $5x10^6$ cells. DNA extraction from urine samples was performed using 50 ml of urine. Cells were sedimented by centrifugation at 3500 rpm for 10 minutes at 4°C. The pellet was resuspended in 1 ml TNE buffer and transferred into a 1.5 ml tube. Next centrifugation step was performed at 3000 rpm for 10 minutes at 4°C and the pellet was resuspended in 400 µl TNE plus buffer. This suspension was incubated overnight in a thermomixer at 55°C and 500 rpm. Next day, 200 µl phenol and 200 µl chloroform were added to the suspension, vortexed and incubated for 5 minutes at the room temperature. Upon the incubation, centrifugation at 10000 rpm for 10 minutes at room temperature was performed in order to separate phases. DNA in the supernatant was transferred into a new tube and precipitated with 40 µl of 3M Na-acetate

and 1 ml of isopropanol. Upon vortexing and centrifugation for 25 minutes at full speed and 4°C supernatant was discarded and DNA pellet was washed in 500 µl 70% ethanol (centrifugation at full speed for 5 minutes at 4°C). Purified DNA was finally resuspended in 30 µl of distilled water and the concentration was measured using Biospectrometer basic (Eppendorf).

2.2.2 PCR (Polymerase Chain Reaction)

Genomic regions of interest were amplified with specifically designed primers (please find enclosed a list of primers as supplementary data). Denaturation step (95°C) was followed by annealing step (temperature dependent on the primer length) and elongation (time dependent on the product length). Appropriate Taq or other polymerases were chosen specific for product length and the need for fidelity. For PCR products shorter than 1 kb, with the aim of sequencing reaction Taq polymerase from MP Biomedicals or Qiagen was used according to the protocol provided by the company. For amplification of longer products high fidelity polymerases such as PrimeSTAR GXL DNA Polymerase, Expand Long Range, or QuickChange Site directed Mutagenesis Kit were used according to the companies´ instructions.

2.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis aims to separate different DNA fragments based on their size. In an electric field negatively charged DNA migrates towards the anode (positively charged electrode) and smaller fragments move faster than the bigger ones. Agarose gels of 2% were used for separation of DNA fragments smaller than 1 kb, while 1% agarose gels were used for larger fragments. Agarose was dissolved in 0.5 x TBE buffer and Midori Green was added to the gel to stain the DNA. Simultaneously with loading the samples a DNA ladder was used in order to recognize the correct fragment size. After separation DNA was visualized with Blue Light LED illuminator (FAS-Digi Gel Documentation System, Nippon Genetics) and pictures were captured with documentation system (Camera MX-1, Pentax).

2.2.4 Gel extraction

Upon agarose gel electrophoresis PCR products or products of DNA digestion were cut out from the gel. DNA extraction from the agarose gel was performed either with QIAquick Gel

Extraction Kit (Qiagen) or Monarch DNA gel extraction kit (New England Biolabs) according to the manufacturers´ instructions.

2.2.5 Ligation

Within the first ligation step of the cloning procedure PCR products were inserted into the pGEM-T Easy vector (Promega) according to the manufacturer´s instructions, taking into account that the ratio between the vector and an insert should be 1:3. The linearized vector contains a single 3´-terminal thymidine at both ends, while the Taq DNA polymerase adds an adenine to the 3´ end of the PCR product. For PCR amplification PrimeSTAR GXL DNA Polymerase (Takara) was used. This high fidelity polymerase creates blunt ends, therefore an additional step with another Taq DNA polymerase (MP Biomedicals) was added after the PCR reaction in order to obtain 3´-terminal adenine and enable ligation. Ligation reaction was catalyzed by T4 ligase which creates a phosphodiester bond between juxtaposed 5' phosphate and 3' hydroxyl termini in duplex DNA.

Subcloning the correct insert from pGEM vector into an expression vector includes another ligation step. Here, the insert and the vector have to be digested with the same restriction enzymes (Fast Digest, Thermo Fisher Scientific). Finally, the products of the restriction enzymes digestion were purified and ligated with T4 ligase (Roche).

2.2.6 Transformation

Introducing plasmid DNA into a bacterial cell is known as transformation. In this way, bacteria are used for replication of plasmids. Besides the gene of interest, plasmids contain a bacterial origin of replication and an antibiotic resistance gene. 50 µl chemically competent cells were mixed with 5 µl of the ligated mixture and incubated for 10 minutes on ice, followed by 2 minutes incubation at 42°C. This ´heat - shock´ treatment improves membrane permeability of bacterial cells and enables introducing the plasmids. Next, 450 µl of LB medium were added to the mixture and samples were incubated for 1 hour at 37°C and 250 rpm, in order to copy the bacteria (and the plasmid) before plating them on agar plates with an specific antibiotic. The plates were incubated over night at 37°C allowing bacteria to form colonies, whereby each colony originates from a single bacterial cell.

2.2.7 Mini preparation

Mini preparation is a small - scale isolation of plasmid DNA from bacteria, based on the alkaline lysis method. Clones that had previously grown on agar plates after transformation were picked in 3 ml LB medium with an antibiotic and allowed to grow over night at 37 °C and 250 rpm.

In the first step of the purification, bacterial cells were pelleted by centrifugation at 5000 rpm for 5 minutes and then resuspended in the buffer 1. In the next step, alkaline lysis of bacteria cells was done with buffer 2 followed by neutralization reaction with buffer 3. Another centrifugation step was performed at 10000 rpm, for 15 minutes, at 4°C, with the aim to separate plasmid DNA from precipitated proteins and cell debris. Isolated plasmid DNA was then washed twice in 100% or 70% ethanol, respectively. Finally, after evaporating ethanol in a vacuum concentrator, plasmid DNA was resuspended in HPLC- purified water and analyzed by restriction enzymes and Sanger sequencing.

2.2.8 Midi preparation

Once the correct plasmid was obtained it was amplified in larger amounts by midi preparation for further functional experiments. Inoculation was performed in 50 ml LB medium with an antibiotic and bacteria were grown over night at 37°C and 250 rpm. Plasmid purification was performed with Plasmid Plus Midi Kit (Qiagen) according to the manufacturer´s instructions.

2.2.9 Plasmid digestion with restriction enzymes

Upon mini and midi preparation all the plasmids were checked for the correct insert with the corresponding restriction enzymes used for cloning. The reaction was done as follows:

10x FastDigest buffer	1.5 µl
DNA (~500 ng)	x µl
Enzyme 1	1 µl
Enzyme 2	1 µl
H_2O	x µl
	15 µl

The samples were then incubated at 37°C for 20 minutes and analyzed by agarose gel electrophoresis.

2.2.10 Sanger sequencing

This method of DNA sequencing is based on the chain-termination method. Deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs) are incorporated in a growing chain by DNA polymerase. The chain will be extended as long as dNTPs are incorporated but when a ddNTP is incorporated the reaction will terminate. DdNTPs lack a 3'-OH group required for the formation of a phosphodiester bond between two nucleotides causing DNA polymerase to stop extension of DNA when a modified ddNTP is incorporated. Each of the four ddNTPs is labeled with a different fluorescent dye which enables detection. By this, DNA fragments of a different length will arise with each of them having the specific fluorescent signal at the end of the sequence. The fragments are then separated depending on their size through capillary electrophoresis and fluorescence is detected. The outcome was illustrated by electropherogram, where each peak represents a single base within the DNA sequence.

The sequencing reaction was performed as follows:

	Plasmids	PCR products			
5x Sequencing buffer	0.5 µl	1.5 µl	96°C	1 min	
Primer (5 µM)	1 µl	1 µl	96°C	10 sec	
DNA (~500 ng)	x µl	x µl	Ta	10 sec	25x
BigDye® Terminator Mix v3.1	1.5 µl	0.5 µl	60°C	1-3 min	
H_2O	x µl	x µl	4°C		
	10 µl	10 µl			

Upon purification with DyeEx 2.0 Spin Kit (Qiagen), samples were run on 3130/3130*xl* Genetic Analyzer.

2.2.11 Pyrosequencing

This is a method of DNA sequencing based on detection of pyrophosphate release upon nucleotide incorporation. A DNA segment of interest was amplified by PCR and biotinylated. After denaturation the biotinylated single-stranded PCR amplicon was hybridized with a sequencing primer and incubated with the enzymes DNA polymerase, ATP sulfurylase, luciferase, and the substrates adenosine 5' phosphosulfate (APS) and luciferin. DNA polymerase catalyzes addition of dNTPs to the sequencing primer and each incorporation event is accompanied by the release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide. ATP sulfurylase converts PPi to ATP in the presence of

adenosine 5' phosphosulfate (APS). This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and seen as a peak in the pyrogram. The height of each peak (light signal) is proportional to the number of nucleotides incorporated. The allelic dosage was quantified on PyroMark Q24, with PyroMark Gold Q24 Reagent Kit (Qiagen).

2.2.12 Array-CGH

Comparative Genomic Hybridization (CGH) is a method for analyzing copy number variations (CNVs) of a test sample compared to a reference sample. Each DNA sample is independently labeled with a different fluorophore, denaturized and hybridized competitively onto an array with nucleic acid targets. Test sample is labeled with a red fluorophore (Cyanine 5), while the reference is labeled with a green fluorophore (Cyanine 3). As the two DNA samples are loaded in the same amount resulting ratio of the fluorescence intensities is proportional to the ratio of the copy numbers of DNA sequences in the test and reference genomes.

Array-CGH analysis was carried out using Human Genome CGH 244k Microarray Kit (Agilent Technologies, CA, USA) according to the manufacturer´s protocol. In order to precisely define the deletion found by array-CGH, we have performed a breakpoint spanning PCR.

2.2.13 SNaPshot assay

This is a primer extension-based method developed for the analysis of single nucleotide polymorphisms (SNPs) which uses fluorescently labeled ddNTPs. We have used SNaPshot® Multiplex System (Applied Biosystems) in order to confirm low levels of the disease-causing variants previously detected by next generation sequencing techniques. After PCR amplification of the region of interest and purification of the PCR product with ExoSAP-IT® (Affimetrix) SNaPshot reaction with single primers was performed. Products of the SNaPshot reaction were purified and run on 3130/3130*xl* Genetic Analyzer for fragment analysis together with GeneScan™-120 LIZ™ Size Standard (Applied Biosystems). GeneMapper™ Software (Applied Biosystems) was used for data analyses.

2.2.14 X-chromosome inactivation assay

Detection of the X-inactivation status was performed at the HUMARA (Human Androgen Receptor) locus. 500 ng of DNA was digested with methylation-sensitive restriction enzyme *Hpa*II (New England Biolabs) overnight at 37°C. In parallel with the patients´ samples a male control DNA sample was treated the same way. In order to check if the digestion is complete samples were loaded onto 1% agarose gel. Upon digestion, heat inactivation of the enzyme was performed at 80°C for 20 minutes. Digested and nondigested DNA was PCR-amplified with the primers targeting exon 1 of the HUMARA gene. Sequence of the forward primer is 5´-FAM-CGGGCCCCAGGCACCCAGA-3´, and sequence of the reverse primer is 5´-GCTGTGAAGGTTGCTGTTCCTCAT-3´. Upon purification, samples were run on 3130/3130*xl* Genetic Analyzer for fragment analysis together with GENESCAN® 400HD ROX® Size Standard (Applied Biosystems). GeneMapper™ Software (Applied Biosystems) was used for data analyses. A deviation from equal (50%-50%) inactivation of each parental allele in a larger cell population is defined as random X-inactivation, with skewed X-inactivation being defined as inactivation of the same allele in 80-100% of cells.

2.2.15 RNA extraction and cDNA synthesis

RNA was extracted from fibroblasts, LCLs, HEK293 or SH-SY5Y cells with ReliaPrep™ RNA Cell Miniprep System (Promega) according to the manufacturer´s protocol. The Super Script III First Strand Synthesis System (Thermo Fisher Scientific) was used to reversely transcribe 1-5 µg of RNA with random hexamers adjusting the same starting amounts of the RNAs between the different samples.

2.2.16 Real time PCR

This method monitors amplification of a DNA molecule during the PCR (or in a real time), unlike the conventional PCR which detects only the final product of the reaction. For measuring gene expression levels TaqMan assay was used, a technology that relies on 5´−3´ exonuclease activity of Taq polymerase.

TaqMan probe consists of an oligonucleotide sequence specific for the gene of interest with a covalently attached fluorophore on its 5´ end and a quencher on its 3´ end. As long as the fluorophore and the quencher are in proximity quenching inhibits any fluorescent signals. As the Taq polymerase synthesizes the new DNA strand its 5´−3´ activity degrades the probe that

had previously annealed to the template. Degradation of the probe releases the fluorophore from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of the fluorophore. Hence, fluorescence detected in the quantitative PCR is directly proportional to the fluorophore released and the amount of DNA template present in the PCR.

TaqMan gene expression assays were purchased from the Thermo Fisher Scientific and expression levels of transcripts were measured on 7300 Real Time PCR System (Applied Biosystems). Relative gene expression was determined using the ΔΔCt method, as previously described [Livak and Schmittgen, 2001].

2.2.17 Cell culture

Handling the cells was performed under sterile conditions using a laminar flow hood with sterile solutions and materials. All the mediums and adequate solutions were previously warmed up in a water bath at 37°C. The cells were grown at 37°C in an atmosphere with 5% CO_2.

Skin biopsies were performed and used to establish fibroblast cell lines which were grown in DMEM, supplemented with 10% FBS and 1% antibiotic/antimycotic solution (DMEM 101, full medium). HEK293, HeLa, HB2 and SH-SY5Y cells were grown under the same conditions in DMEM 101 full medium. LCLs were grown in RPMI medium, supplemented with 10% FBS and 1% antibiotic/antimycotic solution (RPMI full medium).

2.2.18 Harvesting adherent cells

Medium was removed and the cells carefully washed 2 times with PBS. Upon washing, incubating the cells in trypsin/EDTA solution for 3 minutes was performed in order to detach the adherent cells. The activity of the trypsin/EDTA was stopped by adding medium.

Alternatively, cells were harvested with a cell scraper, upon two washing steps with PBS.

Finally, living cells were centrifuged for 5 minutes at the room temperature, at 1500 rpm, and the supernatant was removed. Sedimented cells were stored at -80°C for further experiments.

2.2.19 Cryoconservation of cells

Cells were harvested, well resuspended and counted with an automatic cell counter. Amount of the cells per cryotube was $2x10^6$, mixed with the medium for freezing the cells (80% (v/v)

DMEM 101 and 20% (v/v) DMSO). Cryotubes with cells were kept for a while in an isopropanol-filled container at -80°C, and then transferred into the liquid nitrogen, for a longer period of storage.

2.2.20 Thawing the cells

Thawing the cryostocks of the cells was performed at 37°C in a water bath. When thawed, cells were centrifuged for 5 minutes at 1500 rpm in order to remove the rest of the medium and DMSO. Upon centrifugation, cells were resuspended in fresh DMEM 101 medium and transferred into a sterile culture flask. After overnight incubation medium was changed again.

2.2.21 Transient transfection of the human cells with plasmid DNA

One day before the transfection cells were seeded and incubated to 70-80% confluency for transfection. The amount of cells, medium, plasmid DNA and FuGENE® HD transfection reagent are given in the following table:

Dish	Amount of cells	Amount of medium	Amount of plasmid DNA	Amount of FuGENE
24-well plate	$0.03\text{-}0.05 \times 10^6$	0.5 ml	0.2-2 µg	2 µl
6-well plate	0.3×10^6	2 ml	1-5 µg	4 µl
10 cm Petri dish	2.2×10^6	10 ml	2-8 µg	10 µl

On the day of transfection, fresh medium (DMEM 101) was given to the cells. Appropriate amount of plasmid DNA and FuGENE® was mixed with DMEM pure (100 µl for a 24-well plate, 200 µl for a 6-well plate, 500 µl for a 10 cm Petri dish) and incubated for 20 minutes. Upon incubation, the mixture was added dropwise to the cells and incubated at 37°C for 24 or 48 hours.

While HEK293 and HeLa cells were transfected with FuGENE®, SH-SY5Y were transfected with Amaxa® Cell Line Nucleofector Kit V, according to the manufacturer´s instructions.

2.2.22 Transient transfection of the human cells with siRNA

HeLa or SH-SY5Y cells were transfected with 150-200 nM of each siRNA using Lipofectamine RNAi max and Opti-MEM medium (Life Technologies) and incubated for 48 hours.

2.2.23 CRISPR/Cas9

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system was originally discovered as a prokaryotic immune system that protects bacteria from invading DNA molecules such as viruses and plasmids. The enzyme Cas9 (CRISPR-associated gene 9) is a DNA endonuclease that creates site-specific double-stranded DNA breaks guided by an RNA molecule (gRNA). This system can function in a variety of cells and organisms, allowing permanent modification of the target DNA. Upon a double-stranded DNA break, the cell repairs it either by non-homologous end joining or homologous recombination.

In order to mimic the deletion on chromosome 9 found in one of our patients, we have created HEK293 and SH-SY5Y cell lines deficient for the 3 kb regulatory element. The appropriate gRNAs were inserted in the plasmid (pSpCas9(BB)-2A-Puro (PX459) V2.0, Addgene) as described by Ran et al., 2013. HEK293 cells were co-transfected with 1 µg of each gRNA, marking the region for the double cut, with FuGENE-HD (Promega, Madison, USA). SH-SY5Y cells were transfected with the same gRNAs, albeit with the Amaxa® Cell Line Nucleofector Kit V (Lonza), according to the manufacturer´s instructions. 24 hours after the transfection, puromycin was added as a selection marker to the final concentration of 2 mg/ml. Single clones were picked, expanded, PCR analyzed and Sanger-sequenced for the presence of the desired deletion. Upon the selection process, four independent homozygous clones were obtained for the HEK293 cell line, whereas only one homozygous clone was obtained for the SH-SY5Y cell line.

2.2.24 Luciferase reporter gene assay

In order to investigate transcriptional activity and interaction between a gene promoter and a putative enhancer element, we have used the Dual-Luciferase reporter assay system (Promega, Madison, USA). Luciferases are a class of oxidative enzymes found in several species that enable the organisms to produce bioluminescence or emit light. In this assay, light is produced by converting chemical energy of luciferin oxidation through an electron transition forming the product molecule oxyluciferin. This reaction is catalyzed by Firefly (*Photinus pyralis*) luciferase in the presence of ATP, oxygen and Mg^{2+} (figure 14). The fluorescent product of the reaction can then be quantified by measuring the released light. As an internal control for the transfection experiment, Renilla (*Renilla reniformis*) luciferase was used. The

luminescent reaction catalyzed by Renilla luciferase utilizes oxygen and coelenterate luciferin (coelenterazine).

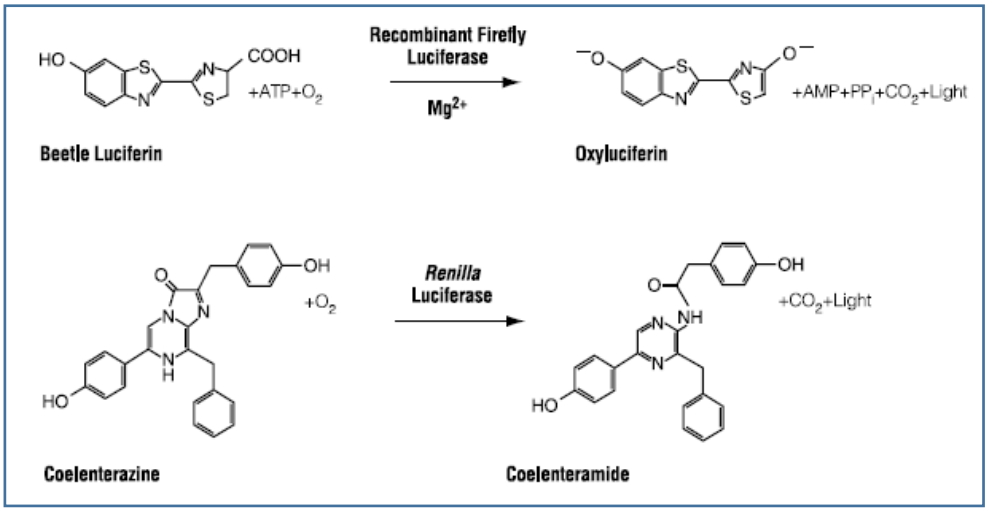


Figure 14: Bioluminescent reactions catalyzed by firefly and Renilla luciferase (adapted from Promega manual)

One day before the experiment 3-5x10^4 cells per well were seeded in a 24-well plate. The transfection mix per well consists of 400 ng pGL4.10 (firefly) construct, 1.2 µl FuGENE-HD (Promega, Madison, USA), and 25 ng pGL4.74 (Renilla) in 100 µl of pure DMEM medium (without serum or antibiotics).

Cells were transfected with different constructs containing firefly luciferase gene under the control of the promoter of interest. With that aim, we have used pGL4.10 vector and the cloning strategy is as described on the figure 15:

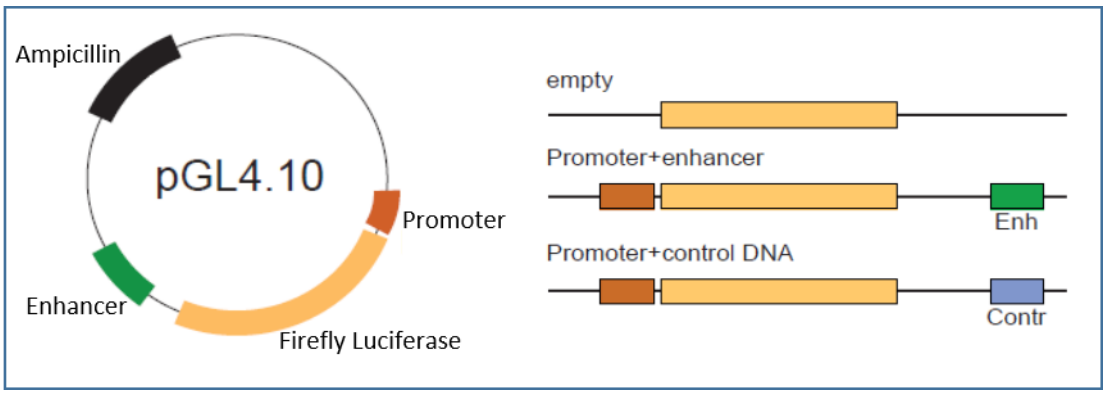


Figure 15: Cloning strategy and different constructs for the luciferase assay. Firefly luciferase is under the control of the promoter, cloned in pGL4.10 vector together with either the putative enhancer, or the size-matched control

Activity of the firefly and Renilla luciferase was measured after 24 hours of incubation with the Dual Luciferase Reporter Assay (Promega) in TriStar^2 LB Multidetection Microplate Reader (Berthold, Bad Wildbad, Germany). All the measurements were verified in a minimum of three independent experiments and as triplicates in each experiment.

2.2.25 Cycloheximide treatment of the cells

In order to block nonsense-mediated decay of the mutant mRNA transcript, we have treated patients´ cell lines with cycloheximide (CHX). Therefore, patients´ fibroblasts or LCLs were seeded in a 6-well plate (seeding density is $0.3x10^6$) one day before treatment. On the day of the treatment medium was removed from the cells and a fresh medium containing cycloheximide to the final concentration of 50µg/ml was added per well. The cells were incubated with CHX overnight and pelleted the next day for further experiments.

2.2.26 Metaphase spreads

To obtain chromosomes at the mitotic stage, colcemid (Life Technologies) was added to the cells at a final concentration of 0.1 µg/ml. Cells were incubated for 3 hours before harvesting in order to depolymerize microtubules and block the cells in mitosis. This was followed by hypotonic treatment and lysis of the cells with 0.56% KCl for 5 minutes. Following the hypotonic treatment, the samples were fixed in a solution containing methanol and acetic acid in ratio 3:1. Samples were then applied dropwise on the microscope slides, dried and stained with Giemsa solution, and chromosomes were observed with immersion oil under the microscope (Axioskop 40, Zeiss). For karyotyping we have used the Ikaros software (MetaSystems, Heidelberg, Germany). At least 15 mitoses were investigated.

2.2.27 ICS assay

Intrinsic Chromosome Structure (ICS) assay is a qualitative *in vitro* assay, described by Hudson et al., 2003 and adapted by Martin et al., 2016. It assesses whether chromosomes are able to recover their native metaphase structure when arrested in mitosis and treated with two different buffers. The first one is a low-salt TEEN buffer, which increases DNA-DNA repulsion forces through cation depletion and consequently leads to a decompaction of mitotic chromosomes. In the next step, the chromosomes are treated with RSB buffer which aims to recover their native structure. After two rounds of unfolding/refolding, mitotic chromosomes

from all the patient cell lines were compared to the healthy control fibroblasts. Principle of the assay is given in the figure 16:

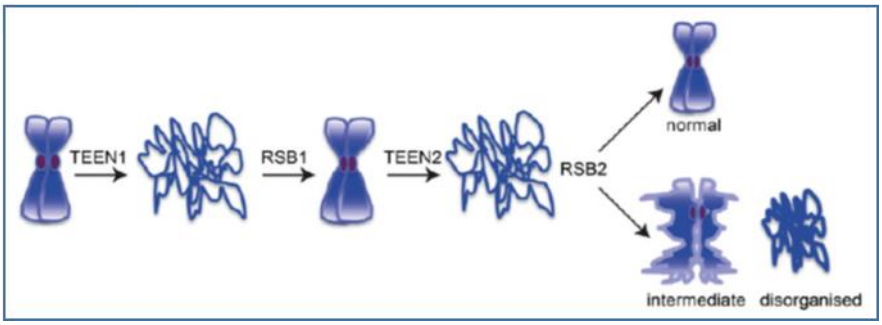


Figure 16: ICS assay is based on structural ´memory´ of chromosomes, in a way that they can be swollen and detangled by TEEN buffer, and can restore their original morphology by treatment with RSB buffer. Sequential decompaction with TEEN and refolding with RSB buffer was used to establish the structural integrity of mitotic chromosomes (adapted from Martin et al., 2016)

Primary fibroblasts were grown on coverslips and blocked in mitosis with 0.1 µg/ml colcemid overnight. Cells were then lysed with 0.56% KCl for 5 minutes followed by TEEN buffer for 30 minutes and RSB for 10 minutes. The TEEN-RSB cycle was repeated, cells were fixed in methanol:acetic acid (3:1) for 30 minutes, washed with PBS twice and mounted with DAPI (ProLong® Gold antifade reagent with DAPI, Life Technologies). Several microscopic slides were prepared for each sample in three independent experiments and at least 30 pictures were taken per sample per experiment (approximately 100 per sample in total) using an Axioplan 2 microscope. As fluorescence imaging platform we have used Isis software (MetaSystems, Heidelberg, Germany). The pictures of the chromosomes were then scored into three groups: normal, intermediate and disorganized, and results are shown as % of mitotic cells.

2.2.28 Cell cycle analysis

The cell cycle is a series of events leading to cell division and duplication of its DNA to produce two daughter cells. It consists of four distinct phases: G1, S, G2 and M. The first phase of the interphase is G1 (*gap 1*), also called growth phase, because here the cell grows in size, along with increased protein synthesis and production of organelles (mitochondria and ribosomes). The next phase is S (*synthesis*) phase, when DNA synthesis starts and the DNA amount

doubles. Production of histones occurs during this phase, while the rates of RNA transcription and protein synthesis are very low. G2 phase occurs after DNA replication and corresponds to a period of increased protein synthesis and rapid cell growth, to prepare the cell for mitosis. Finally, mitotic phase consists of nuclear division (karyokinesis) and it is a relatively short period of the cell cycle.

Fibroblasts and HEK293 cells were grown on a 10 cm Petri dish and fixed in cold 70% ethanol at 70-80% of confluency. After washing cells with PBS and RNase A treatment, DNA was stained with propidium iodide (PI). PI is a fluorescent intercalating agent that binds DNA in a stoichiometric ratio proportionally to the DNA amount present in the cell. By this, cells in S phase will have more DNA than cells in G1 phase. The fixed and stained cells were then analyzed with BD LSR II flow cytometer (BD Biosciences, San Jose, CA, USA) and results were processed in FlowJo software (FlowJo LLC, Ashland, OR) using Watson´s mathematical model.

2.2.29 Protein extraction

Sedimented cells were resuspended in RIPA lysis buffer that contains protease inhibitor cocktail (PIC) in a ratio 250:1. Upon 5 minutes incubation at room temperature, the lysed cells were centrifuged for 5 minutes at 10000 rpm and 4°C. RIPA buffer disrupts also the nuclear membrane; therefore it can be used also for isolation of nuclear proteins. After centrifugation cell debris were sedimented while the supernatant contains all soluble proteins. Supernatant was transferred to a new tube and mixed with the 4x SDS-PAGE loading buffer for subsequent analysis.

2.2.30 Protein quantification

Upon extracting cellular proteins, concentration was determined with Pierce BCA Protein Assay Kit according to the manufacturer´s instructions. This assay is based on the well-known reduction of Cu^{+2} to Cu^{+1} in an alkaline medium by protein (the biuret reaction). Protein concentrations were determined with reference to standards such as bovine serum albumin (BSA). A series of dilutions of BSA solutions with known concentration were prepared and absorbance was measured along with the unknown samples at 562 nm. The standard curve, which represents a correlation between the absorbance and the concentration of the standards, was used to calculate the concentration of the investigated/unknown samples.

2.2.31 SDS-PAGE

A very common method for separating proteins by electrophoresis uses polyacrylamide gel as a support medium and sodium dodecyl sulphate (SDS) to denature the proteins. Therefore, this method is called sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). SDS is a negatively charged anionic detergent that destroys complex structure of proteins. At the same time, polyacrylamide gel restrains large molecules to move as fast as the smaller ones. That means that the proteins will move in an electric field based on their size. The polyacrylamide gel consists of the resolving and stacking gel:

Resolving gel	8 %	10 %	12 %
4x resolving buffer	2.5 ml	2.5 ml	2.5 ml
Acrylamide	2.8 ml	3.33 ml	4.2 ml
H_2O	4.6 ml	4 ml	3.2 ml
10% APS	150 µl	150 µl	150 µl
TEMED	15 µl	15 µl	15 µl

Stacking gel	4%
4x stacking buffer	1.25 ml
Acrylamide	0.65 ml
H_2O	3.1 ml
10% APS	75 µl
TEMED	7.5 µl

The samples were mixed with the 4x loading buffer that contains β-mercaptoethanol in a ratio 20:1, denatured for 5 minutes at 95°C and loaded onto the gel, along with the size marker (Precision Plus Protein Standard: Dual Color, Biorad).

2.2.32 Western blot

After the separation by SDS-PAGE, proteins were transferred to a PVDF (Polyvinylidene difluoride) membrane by semi-dry method. The PVDF membrane was first activated with 100% ethanol and then equilibrated in the transfer buffer for 30 minutes. Whatman paper was previously also soaked in the transfer buffer and the transfer was done as shown in the figure 17. Transfer of the proteins was performed in the Trans-Blot® Turbo™ Transfer System (Biorad) for 30 minutes and with 1 A (25 V). After transfer, the membrane was briefly washed in distilled water and unspecific bindings were blocked with 4% milk in PBS for one hour. As next, the blocking solution was removed and a primary antibody, also diluted in 4 % milk in PBS, was added to the membrane and incubated overnight at 4°C.

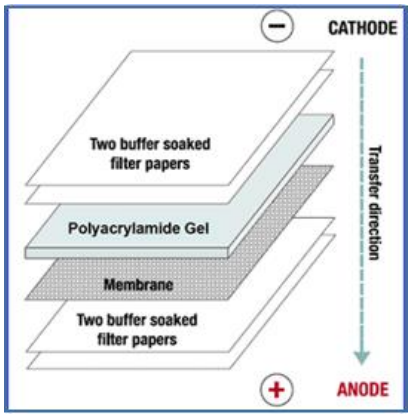


Figure 17: Components for the western blot transfer

The next day primary antibody was removed and the membrane was washed 3 times for 10 minutes in TBST. Finally, the membrane was incubated for 1 hour at the room temperature in the secondary antibody (diluted in 4% milk in PBS) before washing 3 times in TBST and proceeding to the detection of the specific proteins. The secondary antibody is conjugated with the horseradish peroxidase (HRP) which enables the detection by the chemiluminescent reaction. SuperSignal West Femto Substrate enables detection of antigen by oxidizing luminol in the presence of HRP and peroxide. This reaction produces chemiluminescence that can be visualized on X-ray film or an imaging system.

3 Results

3.1 Identification of the causative mutations in the known CdLS genes

Sequencing analysis was conducted on a cohort of patients clinically diagnosed as CdLS or CdLS-overlapping phenotypes. The cohort consists of more than 200 patients and originates from an international collaboration. We have used different sequencing approaches to identify and confirm causative mutations, including Sanger sequencing, targeted gene panel sequencing, SNaPshot, exome- and pyrosequencing.

3.1.1 *NIPBL*

Recent findings indicated a high percentage of CdLS-causing mosaic mutations in *NIPBL*, previously not detected by conventional Sanger sequencing on DNA isolated from blood samples [Castronovo et al., 2010; Huisman et al., 2013; Baquero-Montoya et al., 2014]. Huisman and colleagues detected a *NIPBL* mutation in buccal mucosa cells by Sanger sequencing in 10 of 13 tested patients previously negative for a mutation in lymphocytes [Huisman et al., 2013].

3.1.1.1 Identification of patients with somatic mosaicism in *NIPBL*

We have established a high-coverage targeted CdLS gene panel as described in Braunholz et al., 2015. When possible the analysis was performed on tissues other than blood (preferably fibroblasts or buccal mucosa) to enable the detection of mosaic mutations. By this technique we have originally found three mosaic *NIPBL* mutations in the patients who were previously negative for a mutation detection by Sanger sequencing in blood leukocytes DNA: a nonsense (c.4094 T>A, p.(L1365X)), a missense (c.4751 T>G, p.(L1584R)) and a splice-site mutation (c.5328+1G>C).

All mutations were verified by Sanger sequencing and SNaPshot analysis on DNA from three different tissues: blood lymphocytes, buccal mucosa and fibroblasts. Interestingly, although buccal mucosa DNA was previously suggested as a preferred tissue for detection of the mosaic *NIPBL* variants [Huisman et al., 2013] we were not able to confirm the mutations in this tissue by Sanger sequencing. However, all three mutations were detected on DNA extracted from

fibroblasts. Besides conventional Sanger sequencing we have used SNaPshot assays to verify the mutations in different tissues. These results are presented in figure 18.

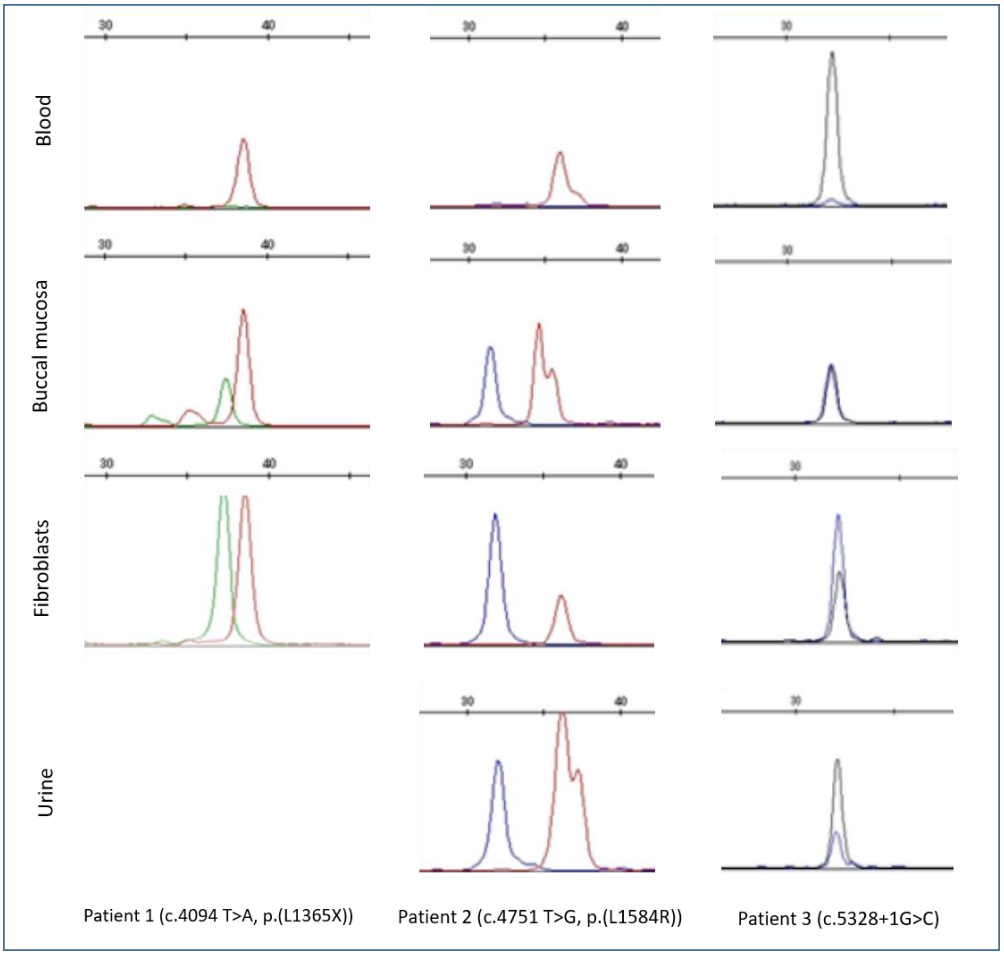


Figure 18: Verification of the identified mutations by SNaPshot analysis. The mutant allele is clearly visible in buccal mucosa, fibroblast and urine DNA. Blood samples of patients 1 and 2 show only wild type allele, while the blood sample of patient 3 shows a faint signal of the mutant allele (from Braunholz et al., 2015)

3.1.1.2 Expanding the clinical and molecular spectrum of *NIPBL* mosaicism

We have reported two additional patients with a clinical diagnosis of CdLS found to be negative for pathogenic variants in the five CdLS genes by conventional Sanger sequencing approaches on blood DNA. Disease-causing variants were subsequently identified by NGS and confirmed by different sequencing approaches.

Patient 4 is a 32-years old male, presenting with facial features typical for CdLS, namely synophrys, long eyelashes, depressed nasal bridge, upturned nose, thin upper lip and low set ears. He also presents with split hand deformity on the left hand and ulnar oligodactyly with

missing fifth finger on the right hand. He is able to write and read with certain limitations and to take care of himself, even working in a carpenter´s shop. The unusually mild intellectual disability indicates the presence of mosaicism and this is, to our current knowledge, a unique combination of features among CdLS patients. He was found to carry a nonsense mutation in 30% of the sequencing reads (c.4399 A>T, p.(K1467X)). The variant was confirmed by Sanger sequencing on buccal DNA, as well as on DNA obtained from epithelial urine cells, whereas it was not detectable on blood DNA. Patient 5 is a female with facial features typical for CdLS, including synophrys and long eyelashes. She displayed early feeding difficulties and foramen ovale with right/left shunt. The patient was selected to undergo a trio-based exome sequencing analysis, which was performed on the HiSeq2500 system (Ilumina, San Diego, CA, USA). This analysis, performed on DNA extracted from fibroblasts, detected a *de novo* missense mutation c.5483 G>A, p.(R1828Q) in 15% of the sequencing reads. Interestingly, we were not able to confirm this mutation by Sanger sequencing neither on blood nor on fibroblast DNA but on the buccal mucosa DNA. Subsequent pyrosequencing analysis revealed the highest amount of the mutant allele in the buccal mucosa. Ratio between mutant and wild type allele was found to be 2:98 in blood, 14:86 in fibroblast, and 17:83 in buccal mucosa DNA, respectively (Figure 19).

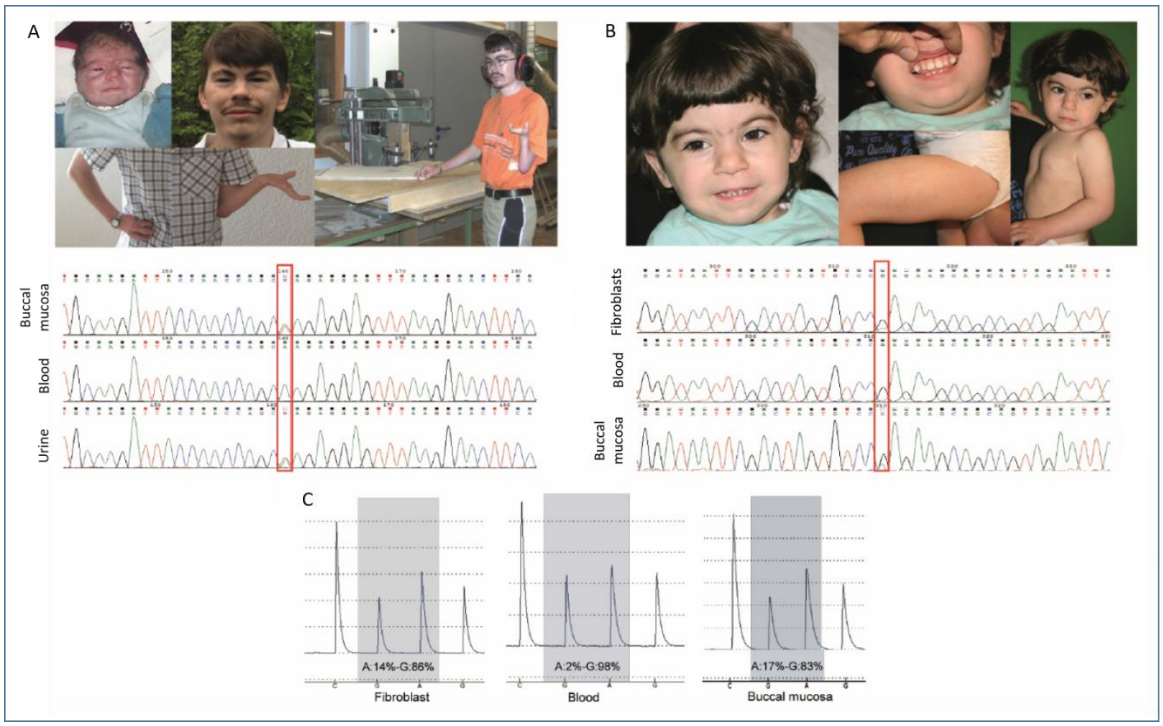


Figure 19: Photographs and sequencing analysis results of the patients 4 (A) and 5 (B), and pyrosequencing results of three different tissues for the patient 5 (C) (from Pozojevic et al., 2017)

3.1.2 *HDAC8*

We have reported clinical and molecular characterization of 11 patients (8 females and 3 males) with 10 distinct variants in *HDAC8*. In order to find and verify the causative mutation, different sequencing approaches such as Sanger sequencing, targeted CdLS panel, whole exome sequencing, SNaPshot and pyrosequencing were applied.

3.1.2.1 Somatic mosaicism in *HDAC8*

Patients 6 and 10 are siblings and share the mutation c.839 C>T, p.(T280I) which was inherited from their apparently healthy mother. To further investigate the mother, we have performed SNaPshot analysis and pyrosequencing on her DNA from blood, buccal mucosa and urine. Along with these three samples, we have used DNA from her affected daughter (patient 6) as the positive control for the mutation. The ratio between wild type and mutant allele was found to be 78:22, 88:12 and 64:36 in blood, buccal mucosa and urine sample, respectively (Figure 20). As an additional method for allele quantification pyrosequencing was performed by Dr. Diana Braunholz and Dr. Ilaria Parenti. Pyrosequencing result showed 19% of the mutant allele in the urine, 2% in the blood and a complete absence of the mutant allele in the buccal DNA of the mother, while it showed 50-50% ratio of the two alleles in the patient´s DNA sample.

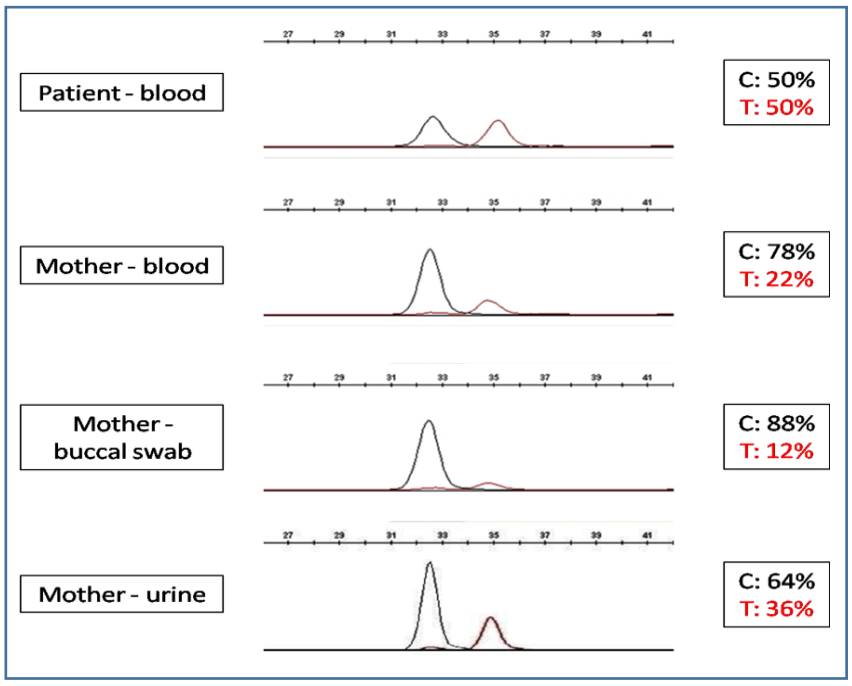


Figure 20: SNaPshot analysis of mosaicism detected in the apparently healthy mother of the patients 6 and 10

3.1.2.2 Functional analysis of a variant affecting a canonical splice donor site

Investigation of the patient 8 mRNA showed that splice site mutation (c.910+1G>A) results in an aberrantly spliced transcript causing frameshift that results in a premature stop codon (p.(V247Ffs*37)). PCR amplification of the patient´s cDNA obtained from fibroblasts and Sanger sequencing showed two bands: a wild type band of 712 bp and an aberrant band of 444 bps. Sanger sequencing of the aberrant band showed skipping of exons 8 and 9 (Figure 21).

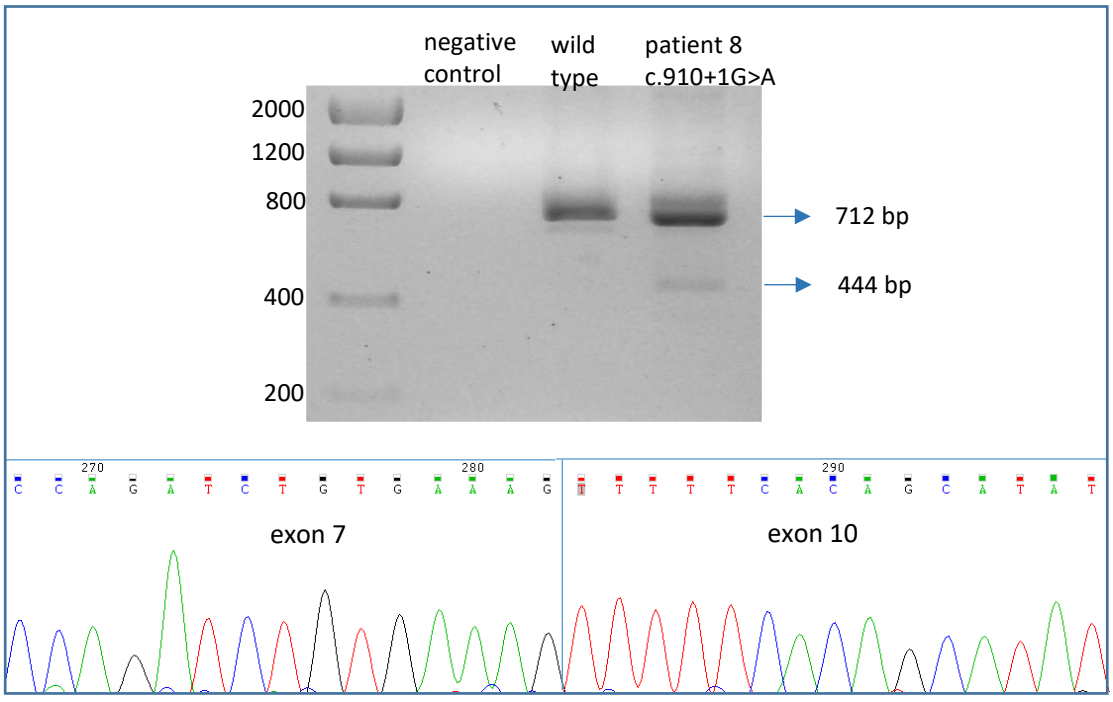
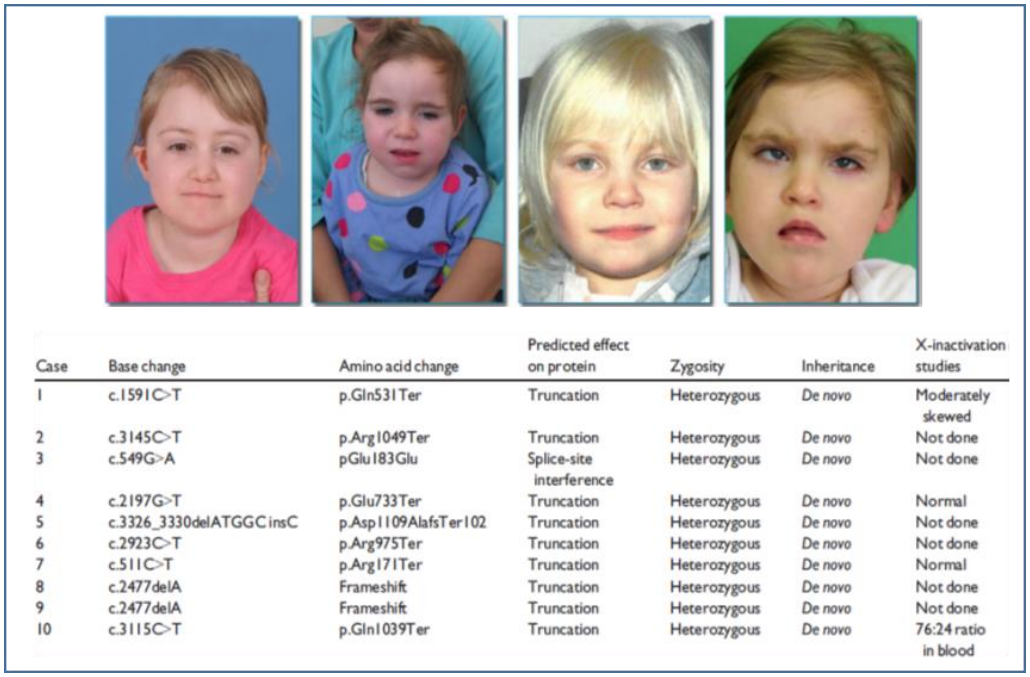


Figure 21: PCR amplification and Sanger sequencing of the patient 8 mRNA shows skipping of exons 8 and 9

3.1.3 *SMC1A*

Missense mutations and small in frame insertions/deletions in *SMC1A* cause CdLS. Within an international collaboration, we have reported a cohort of patients with truncating mutations in *SMC1A*. Mutations that result in premature stop of the protein encoding reading frame were identified in 10 female patients presenting with drug-resistant epilepsy. None of these cases were diagnosed as CdLS. All patients show developmental impairment, abnormalities in electroencephalogram (EEG) and no expressive language. Short stature and progressive

microcephaly were seen in the majority of cases. Facial appearances and the summary of genetic findings for all 10 cases are presented in the figure 22.



Case	Base change	Amino acid change	Predicted effect on protein	Zygosity	Inheritance	X-inactivation studies
1	c.1591C>T	p.Gln531Ter	Truncation	Heterozygous	*De novo*	Moderately skewed
2	c.3145C>T	p.Arg1049Ter	Truncation	Heterozygous	*De novo*	Not done
3	c.549G>A	pGlu183Glu	Splice-site interference	Heterozygous	*De novo*	Not done
4	c.2197G>T	p.Glu733Ter	Truncation	Heterozygous	*De novo*	Normal
5	c.3326_3330delATGGC insC	p.Asp1109AlafsTer102	Truncation	Heterozygous	*De novo*	Not done
6	c.2923C>T	p.Arg975Ter	Truncation	Heterozygous	*De novo*	Not done
7	c.511C>T	p.Arg171Ter	Truncation	Heterozygous	*De novo*	Normal
8	c.2477delA	Frameshift	Truncation	Heterozygous	*De novo*	Not done
9	c.2477delA	Frameshift	Truncation	Heterozygous	*De novo*	Not done
10	c.3115C>T	p.Gln1039Ter	Truncation	Heterozygous	*De novo*	76:24 ratio in blood

Figure 22: Facial appearances, from left to right, of the cases 6, 4, 8 and 10, along with summary of genetic findings for all 10 cases (from Symonds et al., 2017)

Besides these data published as Symonds et al., 2017, additional functional investigations were performed for case 10. Analyses of metaphase spreads revealed a normal karyotype (figure 23).

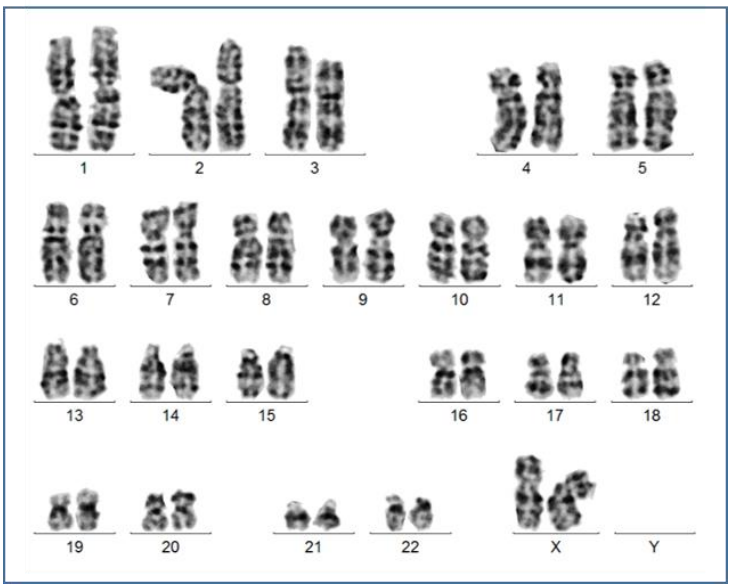


Figure 23: Metaphase spreads analyses of the case 10 reveals a normal karyotype 46, XX

Since *SMC1A* is located on X-chromosome we have analyzed X-chromosome inactivation at the HUMARA locus. Therefore, DNA samples from blood, fibroblasts and LCLs were used. Blood shows borderline skewing with the ratio 76:24 between the two alleles, fibroblasts show extreme skewing (90:10), while the ratio in LCLs was 65:35. Thus, our analyses of all three cell types show expression of mutant as well as wild type *SMC1A* alleles.

We have also investigated nonsense-mediated decay (NMD) of the aberrant *SMC1A* transcript. To chemically block NMD, we treated patient´s fibroblasts and LCLs with cycloheximide (CHX). While the CHX-treated cells showed similar amounts of both transcripts, mutant transcript was almost completely degraded in non-treated cells (figure 24).

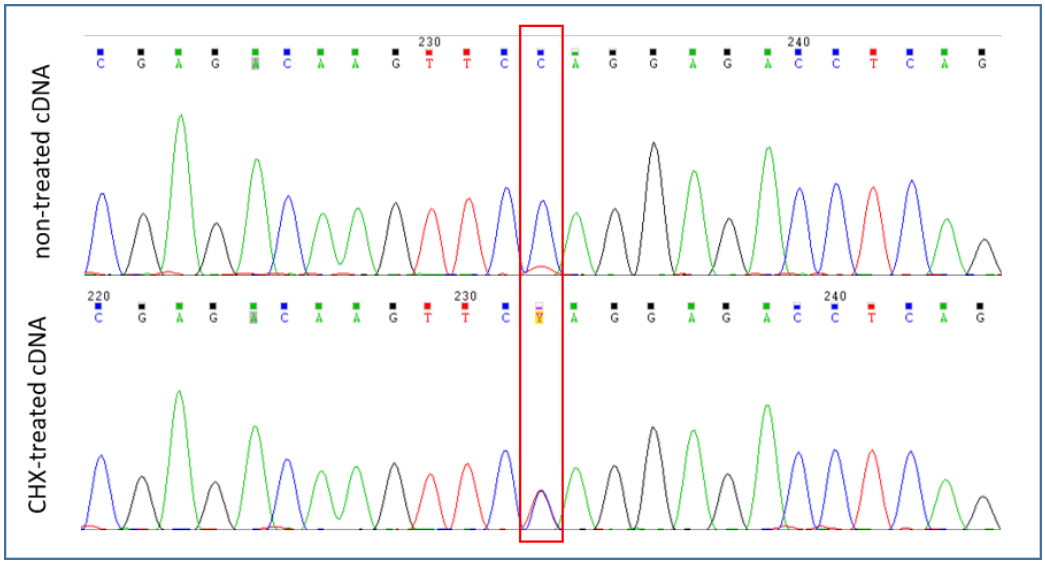


Figure 24: NMD efficiency in the fibroblasts of the case 10. Electropherograms show cDNA of the non-treated cells, compared to the cells treated with cycloheximide

3.2 Identification of genetic variants in new CdLS genes

During the period of my PhD studies, our group has identified variants in genes other than the five known ´CdLS genes´. Within an international collaboration and internationally assembled cohort of patients with CdLS, disease-causing variants in *ANKRD11*, *KMT2A*, *SETD5* and different subunits of the SWI/SNF chromatin-remodeling complex were identified.

3.2.1 ANKRD11

By exome sequencing we have identified two heterozygous loss-of-function mutations in *ANKRD11* in two patients with the initial clinical diagnosis of CdLS. Nonsense mutation c.5483G>T, p.(S1828X) in the patient A was found in 31% of the sequencing reads, thus indicating for a somatic mosaicism. Sanger sequencing on blood and fibroblast DNA showed indeed a higher amount of the mutant allele in fibroblasts (figure 25). Subsequent pyrosequencing (performed by Dr. Parenti) identified 68:32 ratio of the two alleles in blood and 54:36 in fibroblast DNA.

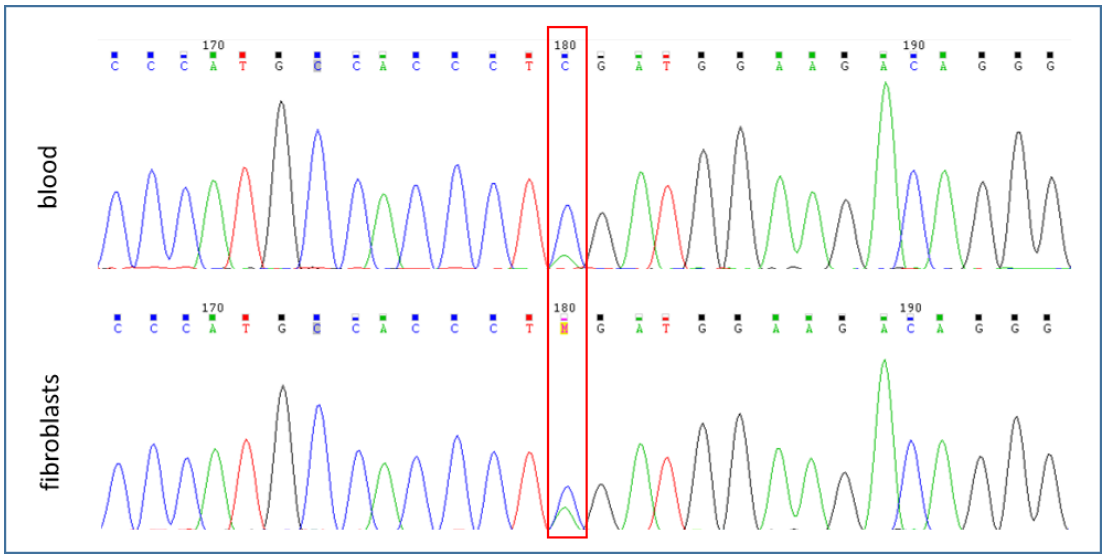


Figure 25: Sanger sequencing results of the patient A on blood and fibroblast DNA, indicating presence of the mosaicism

3.3 Deciphering roles of non-coding regulatory elements in CdLS

Thanks to the international collaboration with the groups of Dr. Kerstin Wendt (Rotterdam, The Netherlands) and Dr. Erwan Watrin (Rennes, France), we have performed functional investigations of various regulatory elements relevant for CdLS and chromatin-associated processes. One project includes investigation of *NIPBL* regulation by a long non-coding RNA (lncRNA) and a distal enhancer, while the other project includes identification and functional characterization of a non-coding regulatory element on chromosome 9.

3.3.1 Regulation of *NIPBL* by a long non-coding RNA and a distal enhancer

In contrast to an expected 50% reduction of *NIPBL* expression, cells of CdLS patients with a heterozygous loss of one *NIPBL* allele as well as heterozygous knockout cells from animal models are able to maintain ~70% of *NIPBL* transcript levels [Liu et al., 2009a; Kawauchi et al., 2009]. Further investigations revealed that in severely affected patients ~65% *NIPBL* expression is observed, while ~75% *NIPBL* expression is observed in less severely affected patients [Kaur et al., 2016]. This indicates an existence of unknown mechanisms which upregulate transcription of the remaining wild type allele suggesting that precise *NIPBL* expression levels are critical for development. In order to investigate these mechanisms and regulatory elements of the *NIPBL* gene we have applied different molecular approaches. We investigated the role of the previously reported long non-coding RNA (lncRNA) upstream of *NIPBL* [Velmashev et al., 2013] and demonstrated that a distal enhancer controls expression of the *NIPBL* as well as the lncRNA.

3.3.1.1 Long non-coding RNA upstream and antisense of *NIPBL*: *NIPBL-AS1*

During the course of our project, Velmeshev and colleagues have identified a *NIPBL* promoter-associated antisense transcript as a part of a study aiming to analyze various autism-related non-coding RNAs. They named it *NIPBL-AS1* and reported that it is located upstream of *NIPBL*, encoding a lncRNA of 5.3 kb size transcribed antisense to *NIPBL* [Velmeshev et al., 2013]. *NIPBL* and *NIPBL-AS1* are transcribed from the same bidirectional promoter, while their annotated transcription start sites (TSS) are separated by 77 bps. The promoter is characterized by a

DNase hypersensitive region flanked by H3K4me3 (trimethylation of lysine 4 on histone H3), indicating active transcription in both directions (data not shown).

Because antisense transcription can be an important regulator of gene expression, we hypothesized that *NIPBL-AS1* transcript might be involved in *NIPBL* expression control. Upon depletion of *NIPBL-AS1* by Antisense Oligonucleotides (ASO), transcription of *NIPBL* was not dramatically affected neither in HEK293 nor in HB2 cells. However, when the actual transcription of either *NIPBL-AS1* or *NIPBL* was blocked by inactive Cas9 enzyme (CRISPRi), it affected expression of the other transcript (data not shown). Therefore, the results show that the transcription of *NIPBL* and *NIPBL-AS1* are interconnected and that active transcription of *NIPBL-AS1* is important to maintain transcription of *NIPBL* and *vice versa*. These experiments were performed in laboratory of Dr. Kerstin Wendt by Dr. Jessica Zuin and Dr. Valentina Casa.

3.3.1.2 Identification of a distal enhancer that controls *NIPBL* expression

We next aimed to identify distal regulatory elements of *NIPBL* and *NIPBL-AS1* using chromosome conformation capturing (3C) technique in the HB2 and HEK293T cells. When a viewpoint was located within the *NIPBL* promoter (VP1), contacts with two intergenic regions R1 and R2 were identified. These two regions are located either 130 kb or 160 kb upstream of the *NIPBL* promoter (figure 26). These experiments were also performed in collaboration with the laboratory of Dr. Kerstin Wendt in Rotterdam.

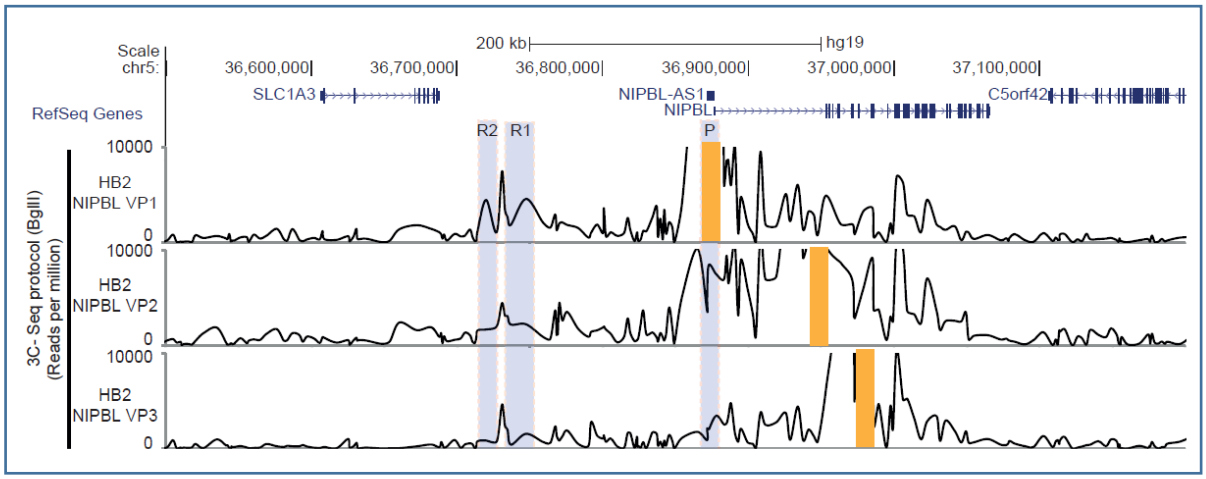


Figure 26: 3C-sequencing results with three different viewpoints (VP1-3) showing contacts with the intergenic regions R1 and R2, upstream of the *NIPBL* promoter. The result is shown for HB2 cell line (from Zuin et al., 2017)

In order to validate these results, higher resolution 3C experiments were performed in HEK293T and HB2 cells, confirming interaction with the R1 region but not with R2. In addition to these data, chromatin immunoprecipitation (ChIP) sequencing data for different histone marks and DNase I hypersensitivity tracks for six cell lines (GM1287, K562, HeLa-S3, HMEC, HSMM, HUVEC) available from the ENCODE (ENCyclopedia Of DNA Elements) project revealed that the R1 region correlated with open chromatin and enhancer marks in different cell lines. Therefore, we hypothesized that this region might be a distal enhancer of *NIPBL*. The R2 region correlated with enhancer marks only in one cell line (GM1287), and therefore might represent a tissue-specific alternative enhancer. Consistent with our observations, publicly available Hi-C data confirm the interaction between the R1 region and *NIPBL* promoter in seven different human cell lines [Rao et al., 2014]. This indicates that this long-range interaction is conserved through different cell types and possibly also species, since Hi-C data from mouse CH12 cells also showed long-range contacts of the *NIPBL* promoter with a potential distal enhancer next to the *Slc1a3* gene (data not shown).

Within the R1 region of 11 kb size, we have identified two smaller regions hypothesized to act as putative enhancers: R1-1 of 2 kb size and R1-2 of 3 kb size (genomic coordinates of these two regions are given in the figure 27). Each of these two regions shows enriched histone marks typical for enhancers, actively transcribed elements and contains CTCF sites (CTCF#1 in R1-1 and CTCF#2 in R1-2).

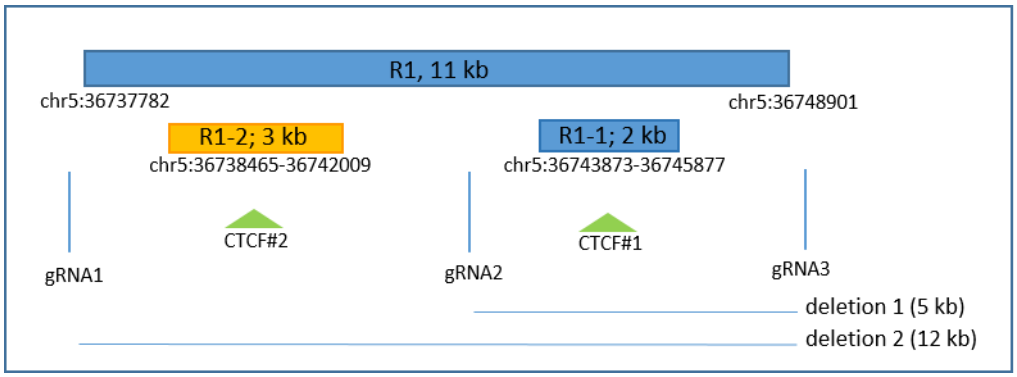


Figure 27: The R1 region of 11 kb size and two smaller regions within it: R1-1 and R1-2, where each of them contains a CTCF site (CTCF#1 and CTCF#2, respectively). The picture shows the two fragments investigated in the luciferase assay, as well as the two deletions created by CRISPR/Cas9 in HEK293T cell line and guide RNAs (gRNAs) targeting them. Genomic coordinates of the regions are according to the hg19 assembly, along with their sizes in kilobases (kb)

3.3.1.2.1 Investigating interaction of the two putative enhancers with the *NIPBL* promoter

To test if these two regions within R1 (R1-1 and R1-2) indeed interact with the *NIPBL* promoter we have inserted each of them into the pGL4.10 plasmid that already contained the *NIPBL* promoter. This reporter gene plasmid has been generated by Dr. Diana Braunholz and was available within our laboratory. It contains the *NIPBL* promoter controlling the *LUC2* luciferase gene. As a control, DNA fragments of similar size to the candidate enhancers R1-1 or R1-2 were inserted into the pGL4.10 plasmid containing the *NIPBL* promoter. Please find the list of primers used in the supplementary table. In comparison to a control DNA of the same size, region R1-1 leads to a clear 2.5 fold increase of the luciferase activity, while R1-2 does not seem to have an effect (figure 28). This result suggests a functional interaction of the R1-1 region with the *NIPBL* promoter and its role of an enhancer.

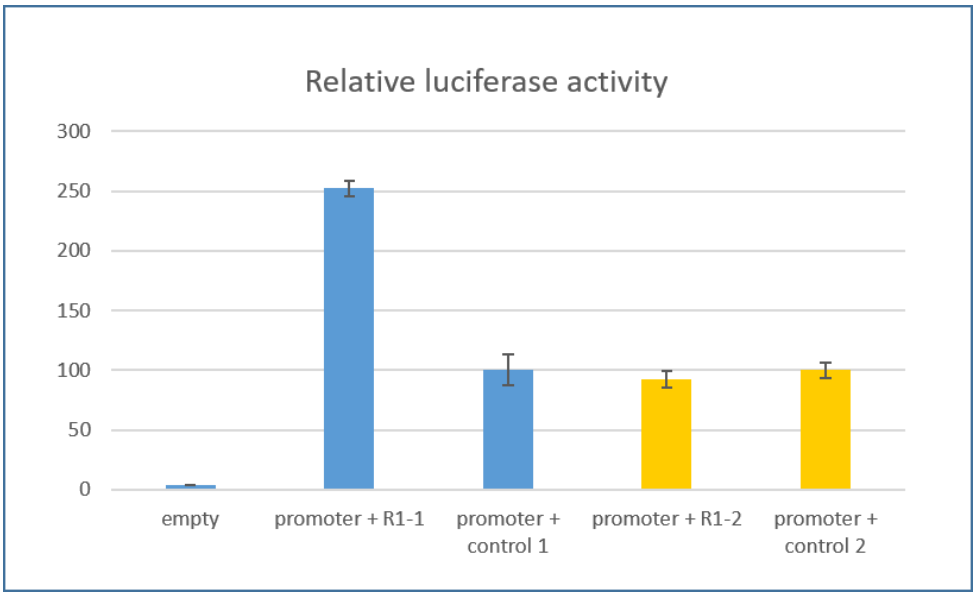


Figure 28: Relative luciferase activity obtained for the two investigated regions R1-1 and R1-2 within the R1 region. Each of the two regions is compared to a size-matched control DNA (2 kb or 3 kb). Blue bars represent results for R1-1 and its size-matched control, while yellow bars represent R1-2 and its respective control

3.3.1.2.2 Investigating effects of the enhancer deletion on *NIPBL* and *NIPBL-AS1* expression

We further wanted to test whether R1-1 and R1-2 regions within R1 display enhancer activity in regulating *NIPBL* and *NIPBL-AS1* expression. Therefore, the group of Dr. Kerstin Wendt used CRISPR/Cas9 system to delete the R1 region in HEK293T cell line. This experiment was started before we have verified the functional relevance of R1 with the previously described luciferase

reporter gene assay. Three gRNAs were designed to specifically recognize regions at edges (gRNA1 and gRNA3) or inside (gRNA2) the R1 region. By this, two deletions were generated as shown in the figure 27. While a 5 kb deletion was generated using gRNA2 and gRNA3 ('deletion 1'), a 12 kb deletion was generated using gRNA1 and gRNA3 ('deletion 2').

Gene expression analysis in genetically modified cells showed reduced *NIPBL* and *NIPBL-AS1* transcript levels (figure 29). These results are in accordance with the luciferase reporter assay data.

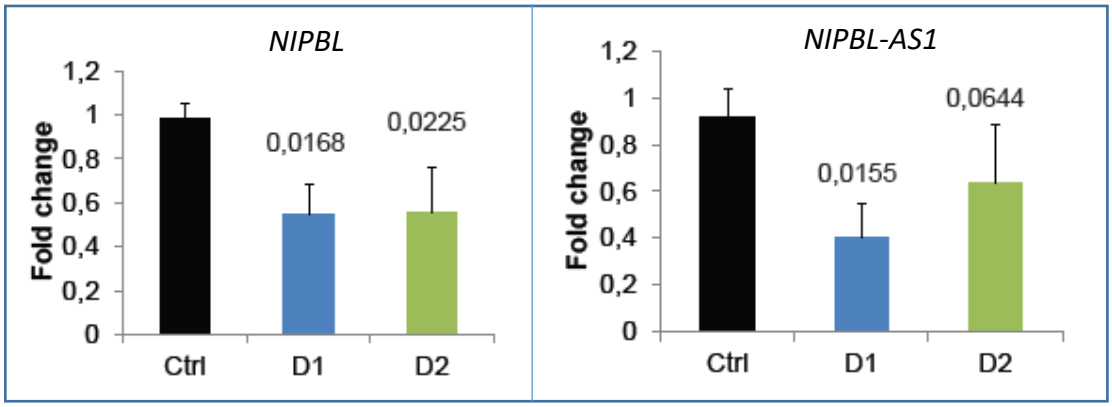


Figure 29: Expression analyses results for *NIPBL* and *NIPBL-AS1* in HEK293T cells, after deleting 5 kb (D1) or 12 kb (D2) region. Ctrl stands for a control and represents wild type HEK293T cells. The blue and green bars represent average transcript levels of the investigated clones (from Zuin et al., 2017)

3.3.1.3 Relevance for Cornelia de Lange Syndrome

After verifying R1-1 region as an enhancer regulating *NIPBL* expression, we screened a cohort of patients with CdLS for a putative disease-causing variants within this region. In addition to this experiment, we have checked *NIPBL* expression and quantified the two alleles in LCLs of patients with *NIPBL* truncating mutations.

3.3.1.3.1 Sanger sequencing of R1-1 overlapping region in patients with CdLS

We have screened a cohort of 30 CdLS patients by Sanger sequencing for a possible variant in the *NIPBL* enhancer region R1-1. Genomic coordinates of the analyzed region are chr5:36744148-36745929 (hg19), and it was covered by four different PCR reactions (please

see supplementary table for the primer sequences). However, sequencing analysis did not reveal any variant that was *de novo* or without a Reference SNP cluster ID.

3.3.1.3.2 *NIPBL* expression and allele quantification

NIPBL transcript levels were found to be reduced in CdLS patients with *NIPBL* mutations, but the cause of the downregulation remains unclear [Liu et al., 2009a; Zuin et al., 2014b]. We aimed to investigate *NIPBL* and *NIPBL-AS1* transcript levels in cells from CdLS patients with heterozygous truncating mutations in *NIPBL*. Investigation of *NIPBL* transcript levels in three patient LCLs (described in Liu et al., 2009a) showed a reduction to 60-70%, consistent with previous reports [Liu et al., 2009a; Kaur et al., 2016] (data not shown; experiments performed in the laboratory in Rotterdam). Since our data show that transcription of *NIPBL* and *NIPBL-AS1* are coupled and *NIPBL-AS1* levels were not found to be significantly affected (data not shown), we have investigated nonsense-mediated mRNA decay in these LCLs. Therefore, we quantified the amount of wild type and mutant allele by pyrosequencing. Here we observed percentages of 38% (patient 1) or 24% (patient 3) of mutant *NIPBL* transcripts, which increase up to 46-48% after blocking NMD with cycloheximide (table):

		DMSO		cycloheximide	
	Mutation	wild type	mutant	wild type	mutant
patient 1	c.4604C>T (nonsense)	62%	38%	52%	48%
patient 3	c.2449_2450insG (frameshift)	76%	24%	54%	46%

This result indicates at least a partial degradation of mutant *NIPBL* transcripts in patient cells, while both alleles remain to be actively transcribed.

3.3.2 A non-coding regulatory element on chromosome 9 and relevance of the condensin complex for CdLS

3.3.2.1 Clinical description of the patient

This project initially started with a patient diagnosed as ´CdLS-like´ by Prof. Dr. Gillessen-Kaesbach in Lübeck. The patient (´patient 1´) is the oldest son of non-consanguineous parents, has a healthy sister and a younger brother that also shows the phenotype. Besides a microcephaly reported at birth (25.5 cm, -3.8 SD) and Apgar score 1/4/6, a gastric tube was embedded. The patient shows mild intellectual disability, aggressive behavior and hearing loss. Clinical evaluation at the age of 9 years showed a more friendly behavior, albeit microcephaly and short stature. He presented with hirsutism on his back and arms. Facial features include synophrys, short nose with anteverted nostrils, long philtrum and thin upper lip. Thumbs of the hands are hypoplastic. Based on these clinical features, Prof. Dr. Gillessen-Kaesbach established the clinical diagnosis of a mild Cornelia de Lange syndrome. In addition, based on the facial features, *in silico* diagnosis by Face2Gene also prioritized CdLS as clinical diagnosis. Further clinical investigations of the patient´s mother also showed very similar clinical features (figure 30).

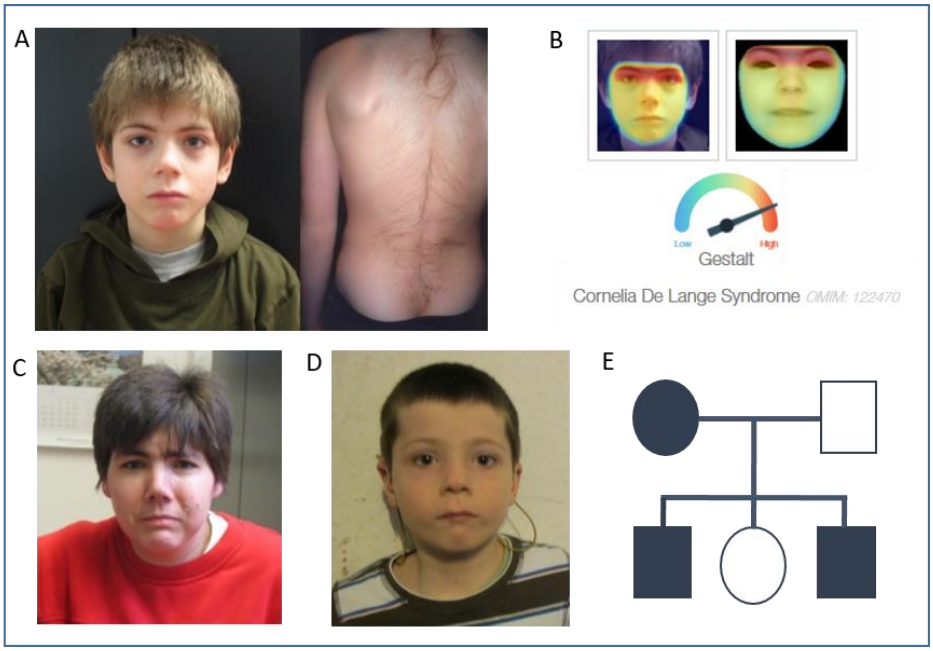


Figure 30: Clinical features of the family diagnosed as CdLS-like. Facial features of the patient 1 and hirsutism at his back at the age of 9 years (A). *In silico* clinical diagnosis with Face2Gene shows a high probability for a CdLS phenotype, based on facial features (B). Patient´s mother (C) and the younger brother (D) showing similar facial features. Family pedigree showing the affected individuals (E)

3.3.2.2 Genetic analyses

Initial Sanger sequencing in the five CdLS genes did not identify any disease-causing variant. As next, gene panel analysis was performed and a variant in *SMARCB1*, encoding for a component of the SWI/SNF chromatin-remodeling complex was detected. However, the same variant was identified in the patient´s healthy sister and therefore excluded to be disease causing. The patient was then selected to undergo exome sequencing, but no putative disease-causing variant was identified.

As next, array CGH revealed a deletion on chromosome 9 inherited from the mother and shared by the younger brother. The exact breakpoints were mapped by Sanger sequencing to chr9:105,763,201-105,823,296 (hg19), confirming the array CGH analysis and defining a heterozygous deletion of ~60 kb (figure 31).

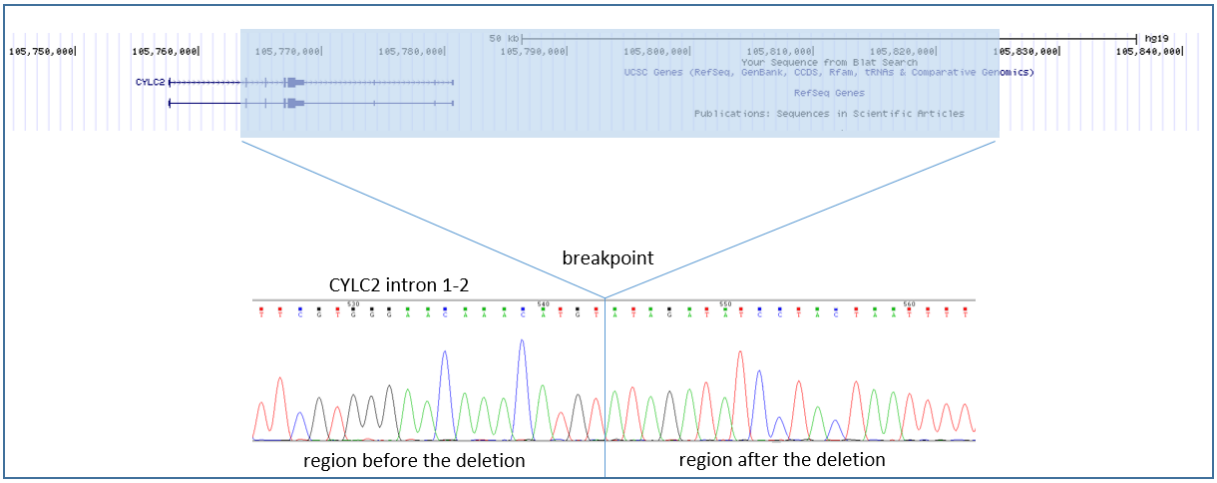


Figure 31: The deleted region on chromosome 9 is marked in blue; the exact coordinates of the deletion are defined by the breakpoint spanning PCR

The deletion includes the exons 2 – 8 of eight exons encoding Cylicin 2 gene (*CYLC2*). *CYLC2* is exclusively expressed in testis and part of the cytoskeletal calyx of mammalian sperm heads. Dr. Diana Braunholz has sequenced a cohort of CdLS patients for a putative disease-causing variant in *CYLC2* but no variant was found.

Because of the testis-specific expression and no obvious connection to the CdLS phenotype, *CYLC2* was excluded as the genetic cause of the disease. Additional *in silico* analysis to identify

putative regulatory elements within the non-coding part of the deleted region revealed an enhancer element of 3 kb. This prediction was based on ENCODE data such as histone modifications, transcription factor binding sites, DNase I hypersensitivity regions and chromatin state segmentation in human embryonic stem cells (hESC) and human testis carcinoma cells (NT2-D1) (figure 32).

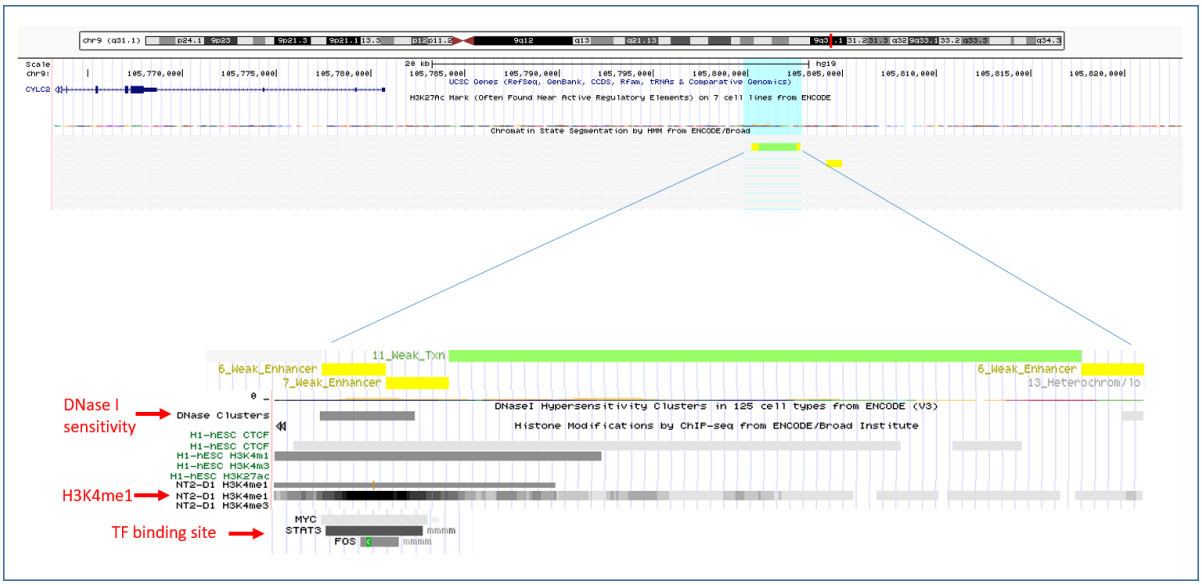


Figure 32: *In silico* analysis of the 3 kb enhancer element within the deleted region on chromosome 9 found in the patient. The figure represents a screenshot from UCSC browser with the ENCODE data information relevant for our investigation (H3K4me1- monomethylation of lysine 4 on histone H3; TF- transcription factor)

We then started searching for candidate genes that might be regulated by this enhancer. The enhancer is approximately 1 mega base (Mb) distant from the next protein-coding gene *SMC2*, so we aimed to investigate if it interacts with the *SMC2* promotor and regulates *SMC2* expression.

3.3.2.3 *In vitro* analyses

In order to investigate interaction between the *SMC2* promoter and the putative enhancer on chromosome 9, we have performed luciferase assays. Therefore, we inserted either the entire 3 kb region or one of its fragments (F1, F2 or F3) into the pGL4.10 vector that drives *LUC2* reporter gene expression under control of the *SMC2* promoter. This experiment was performed under my supervision by our Bachelor student Barbara Geidner and is described in

detail within her Bachelor thesis (´Funktionelle Charakterisierung eines nicht-kodierenden Bereiches auf Chromosom 9 und dessen Relevanz bei der Entstehung eines CdLS-ähnlichen Phänotypen´). Please find the list of primers used for cloning in the supplements. Strategy for generation of reporter plasmids is illustrated in figure 33. Briefly, *SMC2* promoter (chr9:106,855,012-106,857,544; hg19) was inserted into pGL4.10 vector, controlling expression of the *LUC2* gene. Next, full-length (3 kb) enhancer (chr9:105,799,735-105,802,835; hg19) or one of its three fragments were inserted.

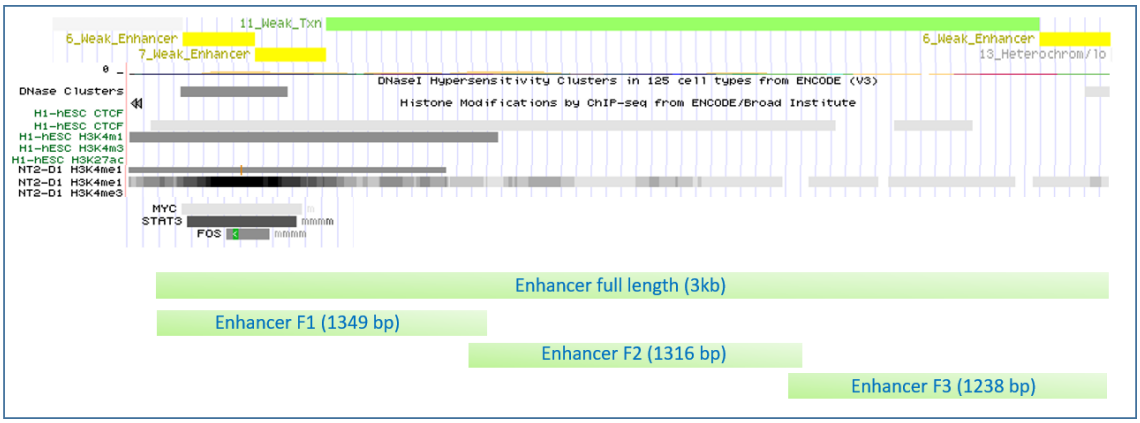


Figure 33: Predicted enhancer (3 kb) and the cloning strategy for the luciferase assay

Relative luciferase activity shows almost total inactivation of the *SMC2* promoter when containing the full-length enhancer or enhancer fragment 1, compared to a vector containing a size-matched control (figure34).

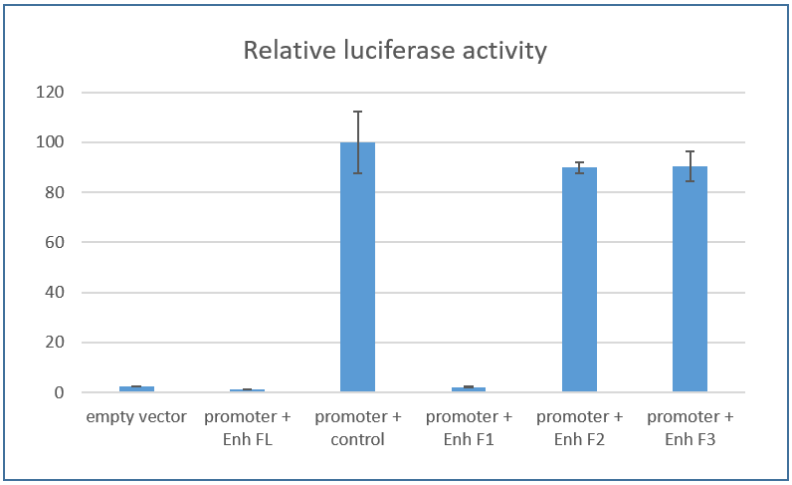


Figure 34: Relative luciferase activity, obtained for the 3 kb enhancer/regulatory element and its three fragments. Enh- enhancer; FL- full-length; F1/F2/F3- three fragments of the enhancer

3.3.2.4 Expression analyses in ´patient 1´ cell line

After functional validation of the regulatory element by *in vitro* studies, we investigated SMC2 protein levels in patient´s cells by western blot analyses. Therefore, we obtained fibroblasts from skin biopsies of the index patient and his healthy sister who does not carry the deletion. As a control, we investigated SMC4, another condensin protein known to form heterodimers with SMC2, as well as β-actin levels. Western blot analysis shows decreased SMC2 and SMC4 levels in patient cells compared to unchanged β-actin levels (figure 35).

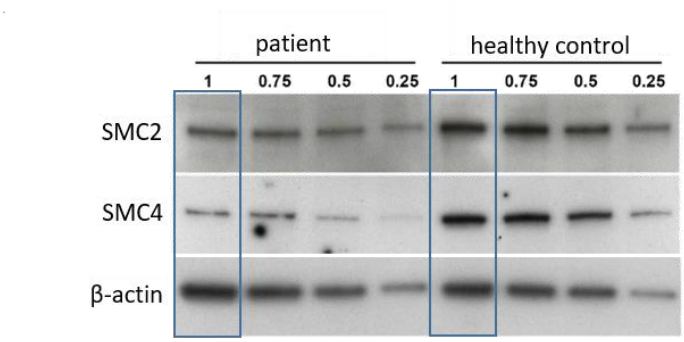


Figure 35: SMC2 and SMC4 protein amounts in the patient fibroblast cell line, compared to healthy control fibroblasts. Different lanes (1/0.75/0.5/0.25) represent different concentrations for loading the samples on the gel. β-actin serves as the loading control

Next, we investigated alterations on transcript levels by real time PCR. RNA was extracted from fibroblast cell lines of the patient and his healthy sister and compared to five healthy Caucasian controls. Analysis of cDNA could show 40% decrease of *SMC2* expression and 70% decrease of *SMC4* expression in the patient cells compared to controls (figure 36). Expression levels were normalized to *SNAPIN*, *GAPDH* and *NADH* expression.

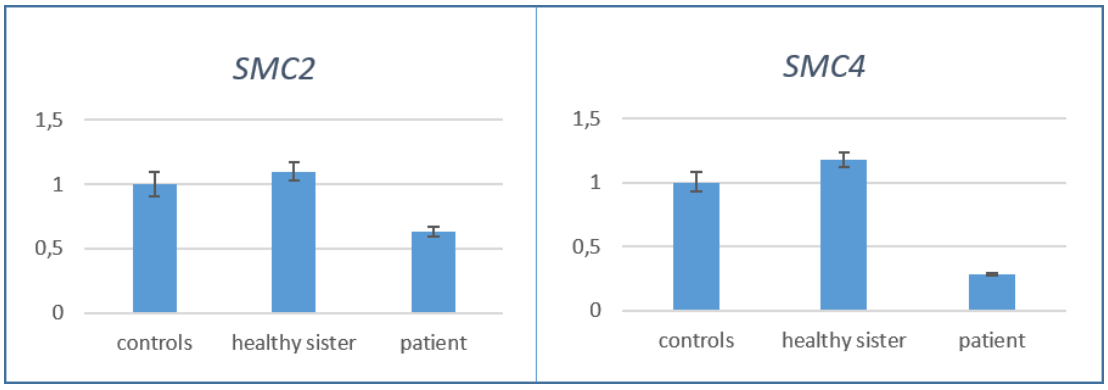


Figure 36: *SMC2* and *SMC4* transcript levels in the patient fibroblast cell line compared to the fibroblasts of his healthy sister and five healthy Caucasian controls

3.3.2.5 Generation of cells deficient for the 3 kb enhancer by CRISPR/Cas9 genome editing

To analyze the functional relevance of the enhancer/regulatory element on *SMC2* expression, CRISPR/Cas9 was used to delete the minimal 3 kb region characterized by luciferase reporter gene assays (please see 3.3.2.3). Therefore, SH-SY5Y and HEK293 cells were transfected with two gRNAs (gRNA1 and gRNA5) targeting the region for the deletion (figure 37). After puromycin selection, genomic DNA was isolated and the clones were genotyped. The strategy of verifying clones for the deletion is illustrated in figure 37. Briefly, primers for PCR verification were created approximately 100 base pairs left and right from the deletion (F and R primer). In case of a deletion PCR product was 224 bp in size, but 3195 bp in case of missing deletion. Clones with a single band of 224 bp were tested by Sanger sequencing, to check the deletion. For illustration, data of the SH-SY5Y clone is presented. Thus, we obtained a single SH-SY5Y clone and four different HEK293 clones homozygous for the deletion.

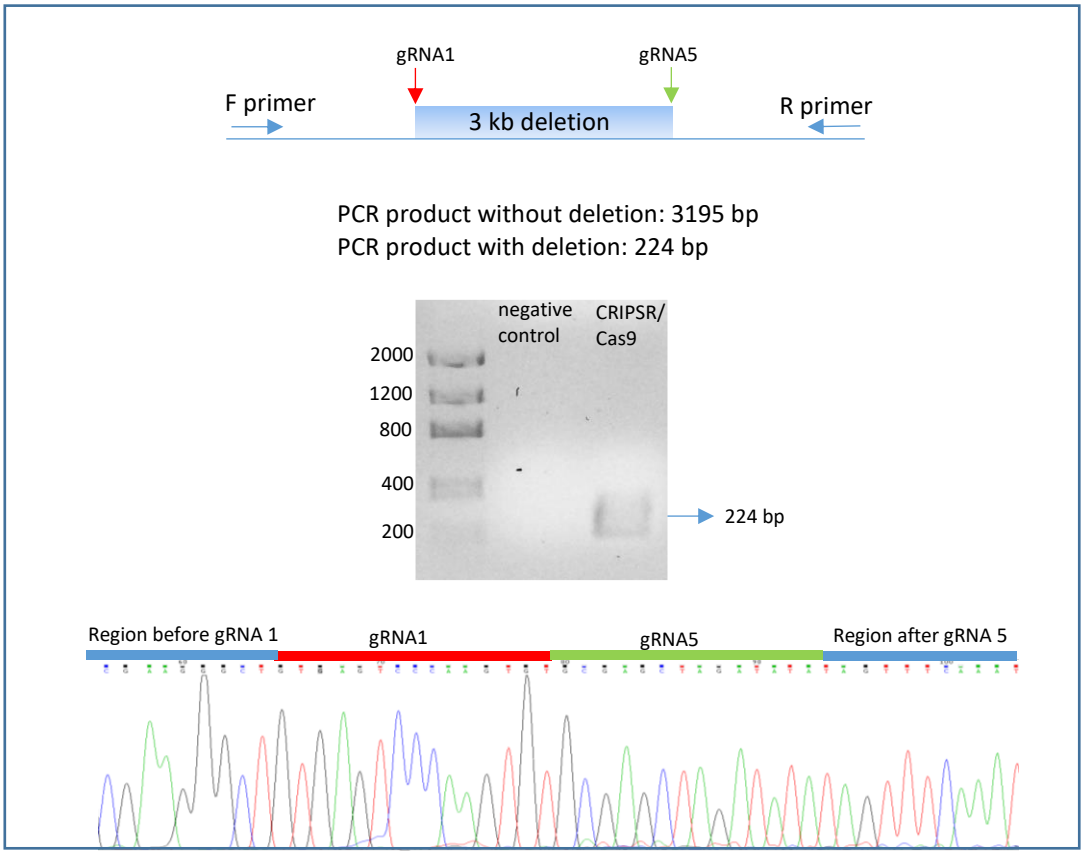


Figure 37: Creation and verification of the clones, deficient for the 3 kb regulatory element. Guide RNAs (gRNA1 and gRNA5) are targeting the region for the deletion, while the verification primers F and R (forward and reverse) bind approximately 100 base pairs (bp) before and after the deleted region. Only clones with a single 224 bp band were selected for Sanger sequencing

Gene expression analysis in HEK293$^{del/del}$ cells revealed ~50% decreased *SMC2* and *SMC4* transcripts, in accordance to the results obtained in patient's fibroblasts. However, SH-SY5Y$^{del/del}$ cells show increased expression levels of *SMC2* as well as *SMC4* compared to SH-SY5Y$^{wt/wt}$ cells (figure 38).

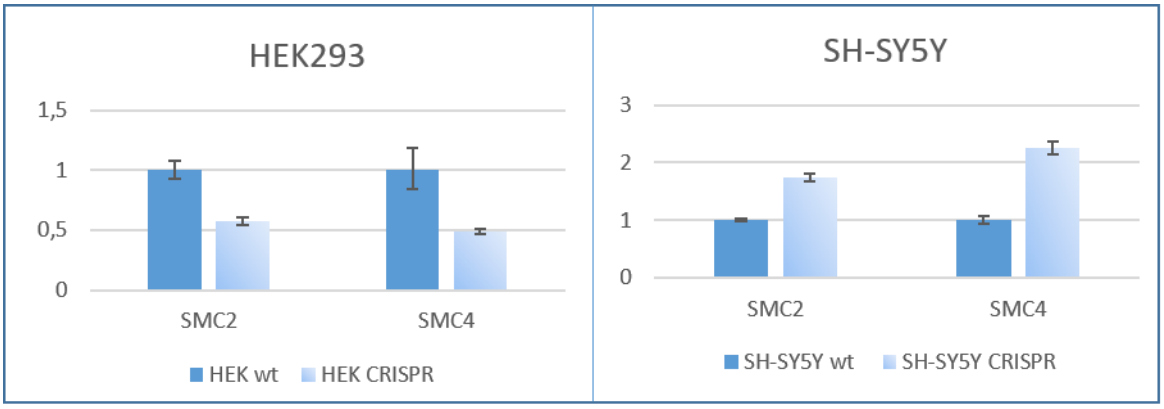


Figure 38: *SMC2* and *SMC4* transcript levels in CRISPR/Cas9 edited HEK293 and SH-SY5Y cell lines (3 kb deletion). Darker blue bars represent wild type cell lines, while all edited cell lines are marked in light blue. The result in edited HEK293 cells is in accordance with the results obtained in the patient fibroblasts cell line.

3.3.2.6 Identification of coding variants in *SMC2*

In collaboration with PD. Dr. Tim Strom from Munich an inherited frameshift mutation in *SMC2* (c.1662_1663delAG, p.(Asp555Hisfs*6)) was identified by trio based whole exome sequencing in an additional patient (´patient 2´) and his mother. The patient was identified by Dr. Kirsten Cremer from Bonn. This patient is the first child of non-consanguineous parents and has a healthy sister. He was born by caesarian section 4 weeks before term. Breastfeeding was impossible due to hypotonia. Besides the hypotonia and spine irregularities, the patient shows mild to moderate intellectual impairment, speech developmental delay as well as aggressive, self-aggressive and hyperactive behavior. The latter improved slightly after melatonin treatment. At the last clinical examination, the patient was 10 years old and his height was 131 cm (P3), weight 33.3 kg (P50-75), while occipitofrontal circumference (OFC) was not measured. The facial features included slight bilateral ptosis, defined eyebrows, high nasal bridge, thin upper lip, prominent upper incisors and prominent ears. Refraction anomaly, wide-spaced nipples and phimosis were further clinical features.

The mother shows symptoms such as mild intellectual impairment and hearing problems. Facial features include defined eyebrows with synophrys and thin upper lip. At the last clinical examination, her height was 151 cm (<P3), weight 55.3 kg and OFC 52.5 cm (P4).

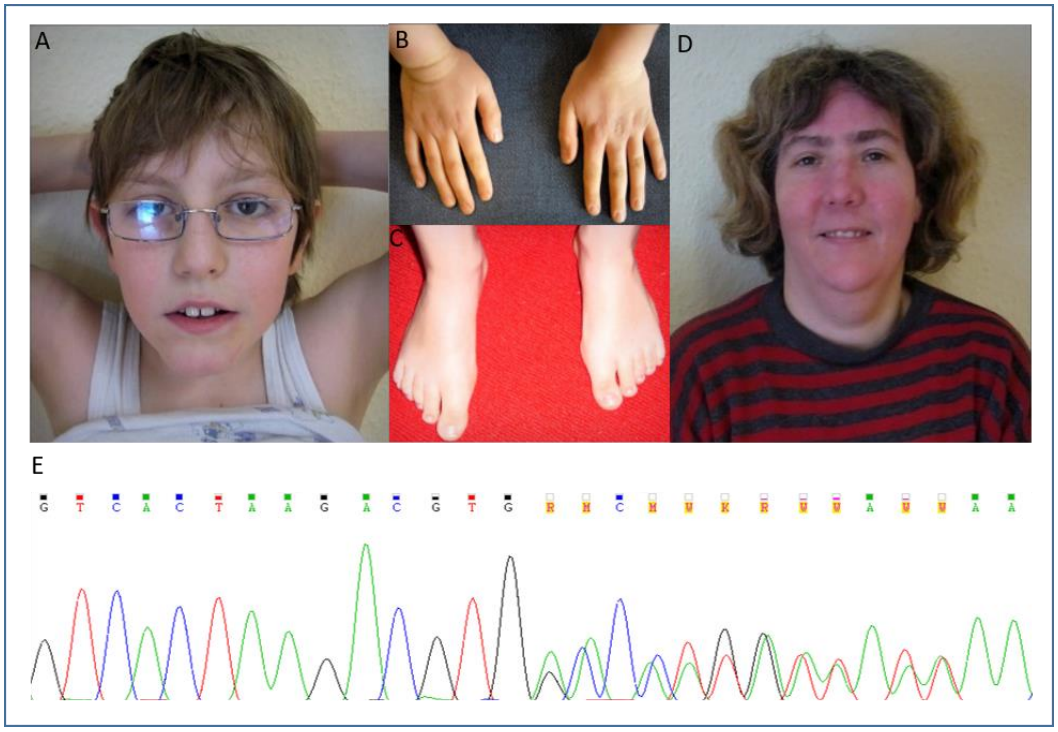


Figure 39: Facial features of the ´patient 2´ (A), along with the pictures of his hands (B) and feet (C). Facial features of the patient´s mother (D), who also has the mutation in *SMC2* (E)

3.3.2.7 Expression analyses in ´patient 2´ cell line

Dermal fibroblasts from skin biopsies of the patient 2 and his mother were used to investigate *SMC2* and *SMC4* expression and protein levels. While quantitative analyses of the transcripts show decreased expression of *SMC2* and *SMC4*, protein levels in the patient cell line were slightly increased. However, no significant alterations with a tendency to decreased protein levels were observed in cells obtained from the mother (figure 40).

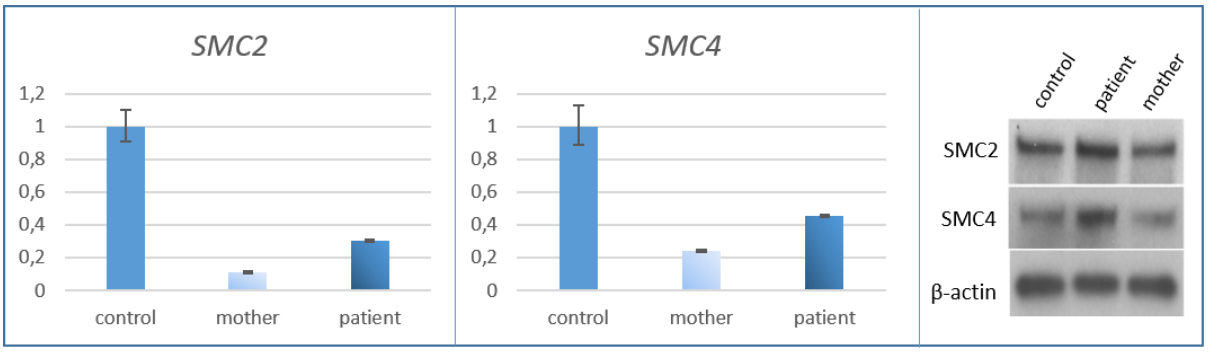


Figure 40: *SMC2*/SMC2 and *SMC4*/SMC4 levels in the fibroblast cell line of the patient and his mother, compared to healthy control fibroblasts

3.3.2.8 siRNA-mediated silencing of *SMC2* affects SMC4 protein levels

To verify that decreased *SMC2* expression affects *SMC4*/SMC4 levels, we specifically silenced endogenous *SMC2* by siRNA. Therefore, HeLa and SH-SY5Y were transfected with 8µl siRNA (SMC2 5´-UGCUAUCACUGGCUUAAAUTT-3´; control 5´-UUCUCCGAACGUGUCACGUTT-3´) and cells were incubated for 48 hours.

TaqMan analyses as well as western blotting were used to analyze transcript and protein levels of *SMC2*/SMC2 and *SMC4*/SMC4 as described (3.3.2.4). Thus, highly efficient silencing of *SMC2* was proven on RNA (95% efficiency) as well as protein level, resulting in a decrease of *SMC4* expression (33% decreased) that results in reduced SMC4 protein levels (figure 41).

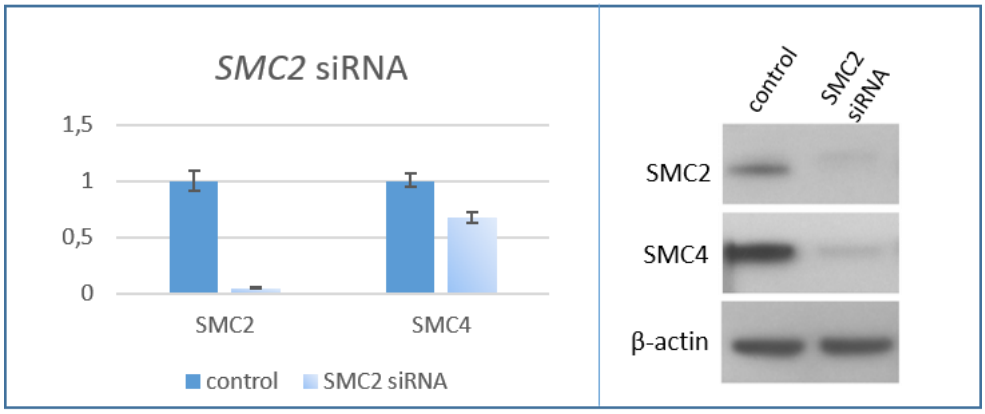


Figure 41: *SMC2*/SMC2 and *SMC4*/SMC4 levels upon transfecting the cells with *SMC2*-specific siRNA. The result is shown for HeLa cells, but we have obtained the same result in SH-SY5Y cells

3.3.2.9 Analyses of alterations in condensin functions

The second part of the project aimed to investigate consequences of these mutations on cellular level. To investigate that, we have performed various experiments such as cell cycle analysis, metaphase spreads and intrinsic chromosome stability (ICS) assay.

3.3.2.9.1 Cell cycle analysis

Fixed cells were stained with propidium iodide in order to quantify DNA amount within cells and determine the amount of cells in each phase of the cell cycle.

Cell cycle analyses were performed for all the cell lines with the deletion on chromosome 9 (patient 1 fibroblasts and CRISPR/Cas9-generated cells) and compared to the respective healthy/wild type cell line. Enrichment of cells in G1 phase and decreased number of cells in G2/M were identified for all the cells compared to respective wild type controls. Strongest G1 enrichment was observed for heterozygous ´patient 1´ fibroblast cells (91.6% cells in G1 phase, compared to 65.5% in control fibroblasts), although homozygous HEK293 and SH-SY5Y cells show significant G1 enrichment as well (34.2% $HEK^{del/del}$ in G1 phase compared to 30.2% $HEK^{wt/wt}$; 69.1% $SH\text{-}SY5Y^{del/del}$ in G1 phase compared to 56.8% $SH\text{-}SY5Y^{wt/wt}$) (figure 42).

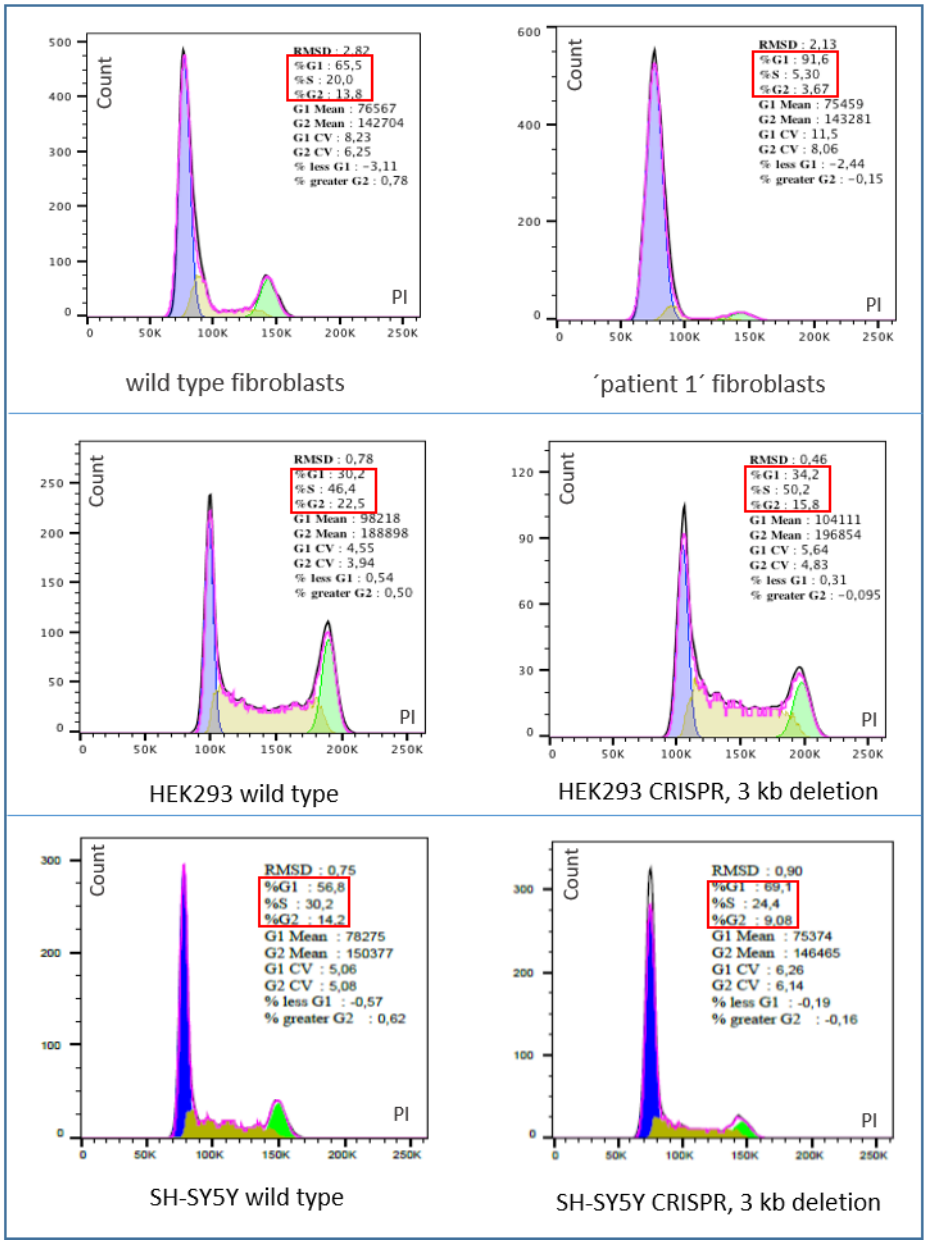


Figure 42: Cell cycle analysis the cell lines with chromosome 9 deletion. First line represents fibroblasts of the ´patient 1´ compared to the wild type fibroblasts. Second and third line represent CRISPR/Cas9 modified HEK293 and SH-SY5Y cell lines with 3 kb deletion, compared to the respective wild type cells.

3.3.2.9.2 Metaphase spreads

Since the condensin complex plays an important role in chromosome condensation and separation of sister chromatids, we investigated mitotic chromosomes in fibroblasts of patient 2 with the heterozygous frameshift variant in *SMC2*.

Metaphase spreads revealed a tendency towards normal appearance of the chromosomes. The result shown in the figure 43 represents one of ten photographs taken from fibroblast sample of patient 2, demonstrating that all sister chromatids are connected in metaphase and excluding precocious sister chromatid separation (PSCS).

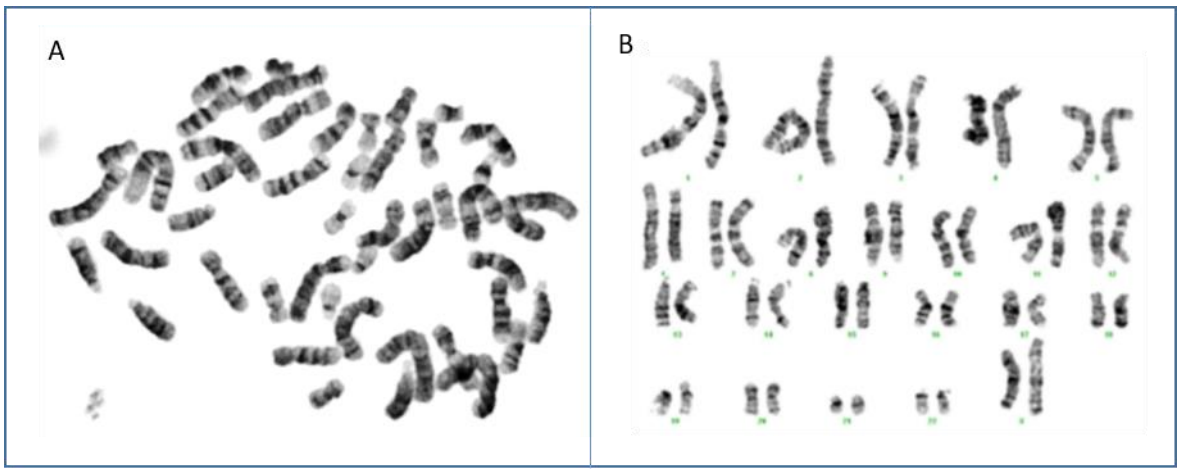


Figure 43: Metaphase spreads of the patient 2 fibroblast sample (A), along with a karyotype of a healthy control fibroblast sample (B).

3.3.2.9.3 Intrinsic chromosome stability

It has been shown previously that condensin complex is required for structural integrity of mitotic chromosomes [Hudson et al., 2003; Martin et al., 2016]. We have adapted the protocols described in both papers (see 2.2.27) to investigate intrinsic chromosome structure (ICS) in fibroblasts of patients compared to two healthy controls. Cells were grown on coverslips, arrested in mitosis by colcemid and lysed. Chromosomes were then treated with two consecutive rounds of TEEN and RSB buffer (please see figure 16), fixed and stained with DAPI. Three independent experiments were performed and approximately 30 photographs were taken per sample, per experiment (approximately 100 photographs in total for each cell line). Ability of the chromosomes to recover their structure was scored according to the criteria illustrated in figure 44 (normal, intermediate or disorganized chromosomal structure).

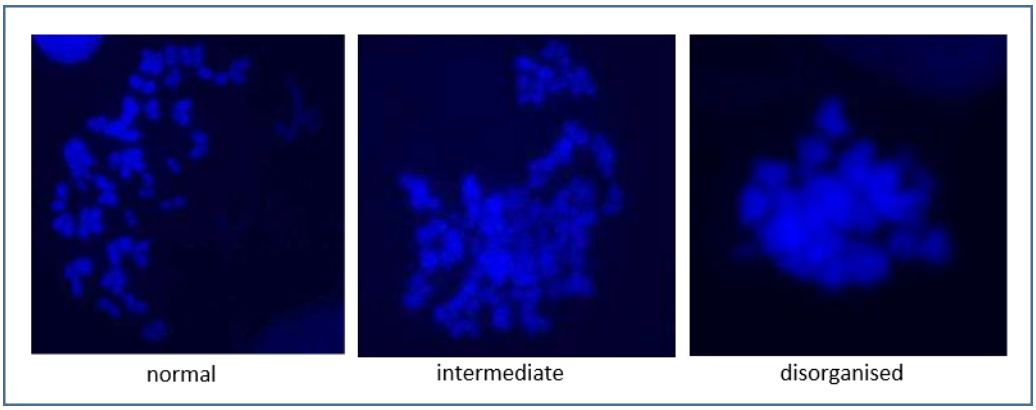


Figure 44: Criteria for scoring the pictures of the chromosomes in the ICS assay

Chromosomes of all mutated cell lines show decreased ability to recover their structure compared to two controls (Normal chromosomes: 66.7% in control 1, 57.5% in control 2, 30.6% in patient 2, 37.8% in patient 2 and 48.3% in mother of patient 2. Disorganized chromosomes: 4.9% in control 1, 10.9% in control 2, 29.6% in patient 1, 21.6% in patient 2 and 20.2% in mother of patient 2) (figure 45).

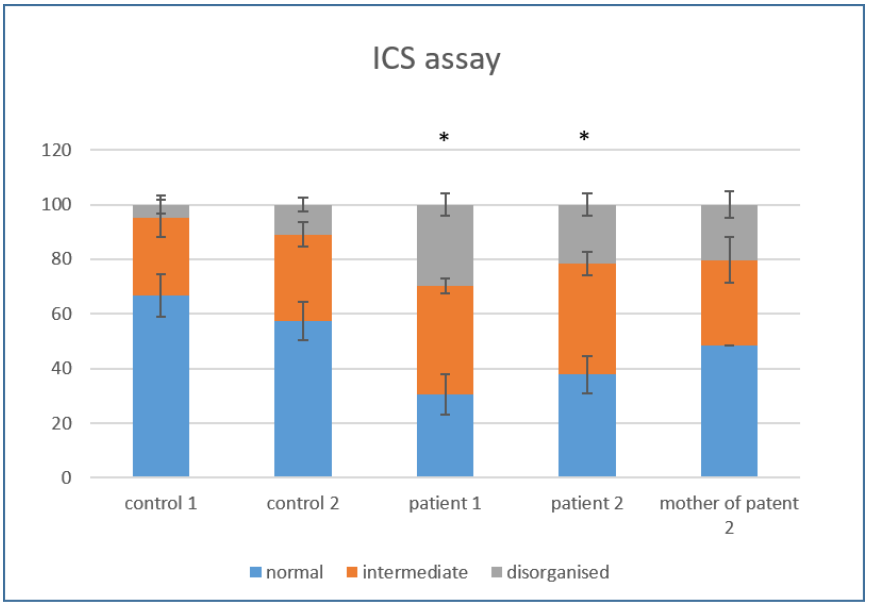


Figure 45: Quantification of abnormal chromosome refolding in ICS assay, for the fibroblasts of two healthy controls, patients and mother of the patient 2. The results are shown as % of mitotic cells (experiments=3; n≥30 mitoses per sample per experiment). Two-tailed *t*-test, (*) $p<0.05$, proportion of disorganized and normal chromosomes compared to control 1

4 Discussion

CdLS is clinically and genetically heterogeneous. The clinical spectrum is very broad, ranging from mildly to severely affected cases. Most patients display typical facial features which are clinical hallmark of the syndrome and often very helpful for establishing the diagnosis. Other typical features include growth retardation, upper limb malformations, hirsutism, problems with inner organs and intellectual impairment. Severely affected patients with characteristic facial features, growth retardation and malformations of the upper limbs are often delineated as ´classical CdLS patients´. However, clinical diagnosis is often challenging for patients with milder or atypical phenotypes.

4.1 Identification of the causative mutations in known CdLS genes

Mutations in the five genes encoding for regulators (*NIPBL*, *HDAC8*) or structural subunits (*SMC1A*, *SMC3*, *RAD21*) of the cohesin complex have been described as genetic causes of CdLS [Krantz et al., 2004; Tonkin et al., 2004; Musio et al. 2006; Deardorff et al., 2007; Deardorff et al., 2012a; Deardorff et al., 2012b]. At the beginning of my PhD studies, it was considered that ~60% of CdLS cases are due to *NIPBL* mutations, 5% caused by mutations in *SMC1A*, *HDAC8* and *RAD21*, while only a single CdLS patient was found to carry a mutation in *SMC3* [Mannini et al., 2013]. At that time, Sanger sequencing was the established gold standard in molecular diagnostics and DNA obtained from blood lymphocytes used for mutation analysis. The majority of remaining genetically unsolved cases (~35%) were patients with milder or atypical phenotype. In 2013, Huisman and colleagues showed that somatic mosaicism in *NIPBL* is frequent (23%) in reliably diagnosed CdLS patients. Of 13 individuals with CdLS tested negative for a mutation in DNA from lymphocytes, in 10 of them variants in *NIPBL* were identified in buccal mucosa DNA by Sanger sequencing [Huisman et al., 2013]. It has been described previously that Sanger sequencing has limitations and cannot detect mosaic alleles bellow 15-20% [Rohlin et al., 2009] explaining a certain portion of mutations that ´escaped´ routine diagnostics. Therefore, Huisman and colleagues suggested buccal mucosa DNA as a reliable way to investigate whether a patient may have somatic mosaicism, if lymphocyte analysis fails to show a mutation.

Based on these findings we have collected buccal mucosa samples from our patients with clinical diagnosis of CdLS but negative for a known disease-causing variant by conventional Sanger sequencing on blood DNA. In an initial pilot project using high-coverage gene panel sequencing, we have identified three mosaic *NIPBL* mutations that were undetected by classical Sanger sequencing analysis on blood as well as buccal mucosa DNA. For further validation experiments, we have collected DNA from additional sources of these patients such as urine samples and fibroblasts obtained from skin biopsies. While Sanger sequencing failed to confirm the mutation on blood and buccal mucosa DNA, all three mutations were verified on fibroblast DNA. Furthermore, NGS result was confirmed by sensitive SNaPshot analyses in different tissues: the mutant allele was clearly visible in buccal mucosa, fibroblast and urine DNA for all investigated samples. In blood samples of patients 1 and 2 only the wild type allele was visible, while a faint signal of the mutant allele was visible in blood sample of the patient 3 (please see figure 18). Therefore, we recommended fibroblasts as the most suitable source for DNA extraction to allow detection of somatic mosaicism in *NIPBL* in routine diagnostics for those patients with clear clinical diagnosis but negative for a variant analyzing DNA from blood samples [Braunholz et al., 2015].

However, we have recently described the first patient with a higher portion of somatic mosaicism in buccal mucosa DNA compared to fibroblast or blood DNA [Pozojevic et al., 2017]. To further verify the mosaicism in this patient (´patient 5´) and quantify the ratio between wild type and mutant allele we have used pyrosequencing analyses. While Sanger sequencing detected the mutation only in buccal mucosa DNA, highly sensitive pyrosequencing analysis revealed a ratio between wild type and mutant allele 98:2 in blood, 86:14 in fibroblast and 83:17 in buccal mucosa DNA (please see figure 19 B and C). This result is in accordance with the previously reported threshold sensitivity of Sanger sequencing (15-20%) [Rohlin et al., 2009], explaining why the mutation could not be detected in fibroblast DNA by conventional Sanger sequencing.

Based on several publications and case reports of CdLS patients, there are no clear clinical features that indicate for somatic mosaicism [Castronovo et al., 2010; Huisman et al., 2013; Baquero-Montoya et al., 2014; Braunholz et al., 2015]. Thus, no correlation between the mutation load and severity of phenotype has been described. As a part of my thesis, we have described the very first patient with a mosaic nonsense variant in *NIPBL* with an unusually mild

cognitive impairment (´patient 4´). Although dramatic reduction defects of the upper limbs found in this patient are typical for severely affected patients with heterozygous *NIPBL* null alleles, cognitive impairment that is only mild has not been previously reported in patients with this type of genetic variant. The mutation was confirmed on buccal mucosa and urine DNA by Sanger sequencing while it was not detectable in blood (figure 19 A). Based on these findings, it is tempting to speculate that some other tissues such as specific neuronal tissues might show only a small amount of the mutant allele, which might explain the unusual mild cognitive impairment. Interestingly, a patient with similar clinical features such as a rather mild cognitive impairment, characteristic facial features and severe limb reduction defects was reported by colleagues from Amsterdam [Bhuiyan et al., 2006]. Unfortunately, this patient was analyzed for a mutation only on DNA from blood and no variant was identified. Additional DNA samples extracted from buccal mucosa or other tissues were not available. Comparing the clinical features and based on our findings, it is tempting to speculate that this patient might be mosaic for a disease-causing variant in *NIPBL* as well.

With identifying very high portions of somatic mosaicism in *NIPBL*, variants in this gene can be found in at least 65-70% of patients with the clinical diagnosis of CdLS according to majority of the current literature. Based on our own findings, this number is even higher if only those patients with ´classical CdLS phenotypes´ would be included. The remaining 30-35% of patients negative for *NIPBL* variants include a significant number of patients with CdLS-like or overlapping phenotypes. Based on our findings, we recommend use of sensitive next generation sequencing technologies with a large number of sequencing reads for a suitable molecular diagnostics. In case no mutation can be detected on blood DNA, we recommend use of DNA from fibroblasts or other easily accessible tissues such as buccal mucosa or urine samples.

Identification of low-level mosaicism is of utmost importance not only for an accurate clinical diagnosis but also for genetic counselling and risk prediction. Another intriguing issue is how much mosaicism/mutant allele can be tolerated in healthy individuals and how much is enough to develop the phenotype in patients. In this regard, Campbell and colleagues have shown that somatic mosaicism in parents of children with simplex genetic disease is more common than previously appreciated, influencing risk prediction [Campbell et al., 2014]. This issue also is illustrated by the example of the two affected siblings from two independent

pregnancies, both with the same disease-causing variant in *HDAC8* inherited from their apparently healthy mother (patients 6 and 10 from Parenti et al., 2016b). In order to investigate mosaicism in the mother, we have collected DNA from different tissues and performed SNaPshot and pyrosequencing analyses. First we performed the SNaPshot assay with the aim to determine the ratio between two alleles, and we have obtained 78:22 ratio in blood, 64:36 in urine and 88:12 in buccal mucosa DNA, compared to the sample of her affected daughter (figure 20). As an additional method for allele quantification, pyrosequencing showed 2% mutant allele in blood, 19% in urine sample and a complete absence of the mutant allele in buccal mucosa DNA. Although these two experiments show discrepancies, trend was the same, showing that the highest number of mutated cells is in the epithelial urine cells and the lowest number is in buccal mucosa DNA. Although SNaPshot is a very sensitive technique for detection of low rates of mosaicism, a robust quantitative analysis is often impossible because of technical limitations. Namely, calculations for the SNaPshot experiment are performed relative to the reference sample based on the surface of the area under the curve. Pyrosequencing is on the other hand reproducible and gives a direct ratio between the two alleles. Limitations of conventional sequencing approaches have already been emphasized and shown on multiple examples. Even today, Sanger sequencing represents the gold standard to exclude disease-causing variants in parents of affected individuals. We have proven mosaicism in the mother only by investigating multiple tissues and by using sensitive target-sequencing technologies. It is tempting to speculate that in the future new sensitive technologies will completely replace Sanger sequencing even for validation of NGS findings, to allow a more accurate molecular diagnostic.

Our work presented here provides the first description of a mosaic variant in *HDAC8* expanding the spectrum of mosaic mutations in CdLS and underlying its relevance in genetic counseling. *HDAC8* is an X-linked gene and consequently hemizygous male patients tend to present with a more severe phenotype compared to heterozygous females. In agreement with this, the p.(T280I) missense variant described here causes a rather moderate phenotype in the female patient (patient 6) while her brother was severely affected and died at the age of 2 years due to pneumonia. Interestingly, patients 6 and 8 were initially not diagnosed as CdLS. While the clinical re-evaluation of the patient 6 (c.839 C>T, p.(T280I)) revealed features that support the molecular diagnosis, clinical diagnosis of the patient 8 (c.910+1G>A) remained unclear. Of note, Harakalova and colleagues previously reported a splice-site variant in *HDAC8*

in a large Dutch family with seven males affected by intellectual disability, truncal obesity, gynecomastia, hypogonadism and unusual facial features [Harakalova et al., 2012]. This family was found to have a phenotype that overlaps with Borjesson-Forssman-Lehmann syndrome and Wilson-Turner syndrome while no differential clinical diagnosis of CdLS is listed in this publication. Thus, specific splice-site mutations resulting in aberrant transcripts encoding aberrant but stable HDAC8 proteins may result in a different phenotype that does not necessarily overlap with CdLS.

Within the last years, our group has also actively contributed to the expanding spectrum of *SMC3* mutations: in addition to the single patient reported previously by Deardorff and colleagues [Deardorff et al., 2007], 15 new cases were reported [Gil- Rodriguez et al., 2015]. Interestingly, some of these causative variants in *SMC3* were found by exome sequencing in patients not initially diagnosed as CdLS but presenting some CdLS-overlapping features.

In addition, we were involved in an international study reporting a cohort of female patients with heterozygous loss-of-function mutations in *SMC1A* [Symonds et al., 2017]. Mutations in *SMC1A* associated with CdLS are missense mutations or small deletions that preserve the protein-reading frame. Although all these patients present phenotypic similarities to CdLS including developmental delay, short stature, microcephaly and cardiac anomalies, none of them was clinically diagnosed as CdLS. Within the last years and with the development of novel sequencing technologies, several other patients with *SMC1A* truncating mutations (mainly nonsense and frameshift) have been reported, all without a CdLS phenotype [Gilissen et al., 2014; Goldstein et al., 2015; Lebrun et al., 2015; Jansen et al., 2016; Huisman et al., 2017]. All these patients present with a drug-resistant epilepsy and developmental delay with some resemblance to individuals with Rett syndrome [Lebrun et al., 2015; Huisman et al., 2017]. Whether truncating variants in this gene should be considered independently from CdLS is a matter of debate. Currently, no male patients have been reported with loss-of-function mutations in *SMC1A*, indicating that loss of SMC1A is lethal. This is in accordance with the function of SMC1A as a core component of the cohesin complex that is essential for cell survival. The disease-causing mechanism in female patients with heterozygous loss-of-function mutations is currently not understood. Lebrun and colleagues demonstrated that there is a decreased *SMC1A* expression in the fibroblasts of the female patient with a loss-of-function mutation [Lebrun et al., 2015]. Interestingly, this gene is located on X-chromosome

in a region that partially escapes X-inactivation. Healthy males were found to have 50% reduced *SMC1A* transcript levels and 44% reduced SMC1A protein levels compared to healthy females [Parenti et al., 2014], similar to the *SMC1A* levels in the affected female patients with loss-of-function alleles. These observations indicate a gender-specific dosage sensitivity that is currently not understood on functional level and is in the focus of several research projects (personal communication, Prof. Dr. Kaiser). The level of *SMC1A* ´overexpression´ in females and X-inactivation data are sometimes discrepant comparing different studies and seems to be tissue-dependent and sensitive for different X-inactivation technique used. Our analyses of *SMC1A* transcript levels in the ´case 10´ show efficient degradation of the mutant transcript by nonsense-mediated mRNA decay, which could be chemically inhibited by CHX (figure 24). Our X-inactivation studies show expression of both wild type and mutant allele with a ratio 65:35 in LCLs, borderline skewing in blood (76:24) and extreme skewing in fibroblasts (90:10). It is still not completely understood how exactly cohesin plays a role in X-inactivation process but it has been reported that it represses *Xi* transcription, synergistically with other factors such as chromatin remodelers and modifiers. [Minajigi et al., 2015].

4.2 Identification of new genes relevant for CdLS

Besides searching for the disease-causing variant in DNA extracted from ´wrong tissues´ or with insufficient sequencing technologies (i.e. somatic mosaicism), a specific variant can cause an unusual phenotype leading to a different/misleading clinical diagnosis. In addition to expanding the numbers and spectrum of variants in the five known CdLS genes, we have identified CdLS-causing variants in genes that were previously connected with other clinically overlapping syndromes. Thus, we have reported two patients with variants in *ANKRD11*, one patient with a variant in *KMT2A*, one with *SETD5*, and three patients with pathogenic variants affecting *ARID1B*.

Mutations in *ANKRD11* were previously described to cause KBG syndrome characterized by intellectual disability, skeletal malformations and macrodontia [Sirmaci et al., 2011]. Variants in *KMT2A* formerly named *MLL* cause Wiedemann-Steiner syndrome, recognizable by excessive hair growth around elbows, short stature, intellectual disability and distinctive facial appearance [Jones et al., 2012]. *SETD5* mutations have previously been connected with 3p25

microdeletion syndrome, characterized by intellectual disability, short stature, microcephaly, hypotonia and heart defects [Grozeva et al., 2014; Kuechler et al., 2015]. Finally, mutations affecting SWI/SNF chromatin remodeling complex gene *ARID1B* cause Coffin-Siris syndrome with developmental delay, speech impairment and hypoplastic or absent fifth fingernail or toenail [Santen et al., 2012; Tsurusaki et al., 2012]. Although all these syndromes are distinct entities they share overlapping clinical features with CdLS and are often listed as a differential clinical diagnosis for CdLS. For instance, Wieczorek and colleagues described a patient with a clinical diagnosis of Coffin-Siris syndrome (CSS) who was later found to carry a *NIPBL* mutation [Wieczorek et al., 2013; Parenti et al., 2017].

Although CdLS and KBG syndrome are distinct entities, they share overlapping features such as cognitive impairment, growth retardation, brachycephaly, small hands and feet, etc. Our patient A [Parenti et al., 2016a] shows some clinical signs common to both CdLS and KBG but also additional features/signs typical only for CdLS including small head circumference, synophrys, long eyelashes and depressed nasal bridge (figure 46 F). Interestingly, she was found to carry *ANKRD11* mutation in a mosaic state, which may explain her unusual phenotype. Consistent with our observations that mutations in *ANKRD11* could lead to a clinical diagnosis that overlaps with CdLS, Ansari and colleagues had previously reported three cases with *de novo* mutations in *ANKRD11* and clinical diagnosis of CdLS [Ansari et al., 2014]. Although mutations in *KMT2A* are associated with Wiedemann-Steiner syndrome [Jones et al., 2012], there are patients reported to carry mutations in this gene with a clinical diagnosis of overlapping syndromes such as Coffin-Siris or Cornelia de Lange syndrome [Bramswig et al., 2015; Yuan et al., 2015]. In addition, we have also reported a patient with a *NIPBL* mutation previously diagnosed as a CSS patient [Wieczorek et al., 2013; Parenti et al., 2017]. Clinical diagnosis is often established based on the facial features, which can be overlapping between different syndromes as illustrated in figure 46.

Interestingly, all these genes encode for different components involved in chromatin organization and processes relevant for transcriptional regulation. While for some of these proteins first molecular data indicate a functional interaction with the cohesin complex, relevance of some proteins is currently unknown. It has been shown that in yeast the SWI/SNF-related RSC chromatin remodeling complex recruits Scc2-Scc4 (yeast ortholog of NIPBL-MAU2 cohesin loader) to broad nucleosome-free regions. Consequently, inactivation of the cohesin loader or the RSC complex had effects on nucleosome positioning, gene expression and sister

chromatid cohesion [Lopez-Serra et al., 2014]. Our work establishes for the first time the clinical link between CSS and CdLS, supporting the molecular link described by Lopez-Serra and colleagues.

Figure 46: Comparison of different but overlapping clinical phenotypes: a ´classical CdLS´ patient with a truncating variant in *NIPBL* and a severe phenotype (A) and a milder CdLS case with a missense variant in *NIPBL* (from Mannini et al., 2013) (B). A classical CSS patient with a mutation in *ARID1B* (from Santen et al., 2012) (C); a patient with a clinical diagnosis of CSS but with a variant in *NIPBL* (from Parenti et al., 2017) (D). A classical KBG patient with a variant in *ANKRD11* (from Sirmaci et al., 2011) (E) and our patient with a nonsense mosaic variant in *ANKRD11* (from Parenti et al., 2016a) (F)

SETD5 and KMT2A act as histone modifiers and mediate chromatin activity/structure associated with epigenetic transcriptional activity, e.g. promoter and enhancer activity and accessibility. For instance, KMT2A catalyzes methylation of H3K4 and it is well known that H3K4me3 is associated with active promoters, while H3K4me1 is a mark of enhancers. While methyltransferase activity of KMT2A is well characterized, function of SETD5 has yet to be determined. Based on the sequence similarity to other SET domain proteins it is predicted to

act as a methyltransferase. It was shown that SETD5 regulates histone acetylation during gene transcription and it was found to co-immunoprecipitate with histone deacetylases [Rincon-Arano et al., 2012; Osipovich et al., 2016]. Finally, ANKRD11 has also been reported to recruit histone deacetylases (HDACs) in order to inhibit ligand-dependent transcriptional activation [Zhang et al., 2004]. However, it is still unknown if there is a direct functional interaction with subunits or other regulators of the cohesin complex.

Shared molecular mechanisms and pathways might explain clinical features found in CdLS and overlapping syndromes. All these genes encode for proteins involved in chromatin organization and transcriptional regulation, further supporting the hypothesis that chromatin dysregulation and transcriptional disturbances are crucial for understanding the pathogenesis of all these phenotypical overlaps. Therefore, conventional sequencing analysis of the five ´CdLS genes´ is not sufficient and analyses of these additional genes should be considered for those patients with a suspected clinical diagnosis of CdLS, especially when negative for a variant in any of the five known CdLS genes. With the dropping costs of next generation sequencing technologies, we believe that technologies such as exome or whole genome sequencing will become the preferred technologies even in routine molecular diagnostic in the future.

4.3 Regulation of *NIPBL*

Although variants in a growing number of genes have been described as genetic causes of CdLS, *NIPBL* remains ´the main CdLS gene´ accounting for at least 65-70% of cases. Heterozygous mutations resulting in null alleles are very often associated with a more severe phenotype and haploinsufficiency has been postulated as the pathomechanism. Interestingly, cells of CdLS patients with heterozygous loss-of-function mutations in *NIPBL* and cells from heterozygous *Nipbl* knockout mice showed reduction of mRNA levels only to ~70%, instead of the expected 50% [Liu et al., 2009a; Kawauchi et al., 2009; Kaur et al., 2016]. Therefore, we were interested in the identification and functional characterization of genomic elements involved in regulation of *NIPBL* expression. Besides understanding how exactly *NIPBL* is regulated and expressed, this knowledge might be relevant for future therapeutic approaches. Single nucleotide variations (SNVs) as well as copy number variations (CNVs) outside of the coding genome can interfere with normal gene regulation, which was shown for various

diseases [Bhatia et al., 2013; Weedon et al., 2014; Lupianez et al., 2015; Spielmann and Mundlos 2017].

Within my thesis, I was involved in an international collaboration project focusing on the function of the *NIPBL* promoter. We could give functional evidence for the bidirectional activity of the *NIPBL* promoter driving expression of two divergent transcripts: *NIPBL* and the long non-coding RNA *NIPBL-AS1* transcribed antisense to *NIPBL.* Similar observations were described for a large number of promoters in the human genome [Preker et al., 2008; Seila et al., 2008; Bagchi and Iyer, 2016] but precise regulatory mechanisms are still unknown. We have shown for the very first time that blocking expression of one of each transcripts by inactive Cas9 (CRISPRi) results in increased expression levels of the other transcript. This result indicates that *NIPBL* and *NIPBL-AS1* transcription is coupled.

By chromatin-conformation capture sequencing we have identified a putative enhancer region (R1, chr5:36737782-36748901) 130 kb upstream of the *NIPBL* promoter. Analyses of the ENCODE data for this region showed that it correlates with open chromatin and enhancers marks in different cell lines. Subsequent functional studies such as reporter gene assays and expression analyses on CRISPR/Cas9 modified cells narrowed down the actual enhancer within this 11 kb region (R1) to approximately 2 kb (R1-1) (please see figure 27 for an explanation). The interaction between the 2 kb enhancer (R1-1) and the *NIPBL* promoter was demonstrated in the luciferase assay: in comparison to a size-matched control, R1-1 region leads to a clear 2.5 fold increase of the luciferase activity (figure 28). Of note, another region within the R1 was also investigated for the enhancer activity (R1-2) but the result was negative. In addition, two deletions were created in HEK293 cell line by CRISPR/Cas9: a smaller 5 kb deletion encompassing the R1-1 region, and a bigger 12 kb deletion encompassing the entire R1 region (which includes both R1-1 and R1-2). Gene expression analysis in genetically modified cells showed reduced *NIPBL* and *NIPBL-AS1* transcript levels (figure 29). Since there was no significant difference between the effects of the smaller and larger deletion on *NIPBL* and *NIPBL-AS1* expression, we conclude that the actual enhancer element is located within the smaller deletion corresponding to the R1-1 region.

Due to its close proximity to the two transcription start sites, the enhancer might stimulate transcription of both *NIPBL* and *NIPBL-AS1* suggesting a competition between the two genes for enhancer activity. This hypothesis was supported by findings that blocking the

transcription of *NIPBL* using CRISPRi results in increased transcription of *NIPBL-AS1* and *vice versa* [Zuin et al. 2017]. Based on the presented data, we suggest that the actual transcription of the lncRNA and the competition for the distal enhancer might be mechanisms relevant for the regulation of *NIPBL* expression. These insights reveal a possibility to increase *NIPBL* transcription by interfering with *NIPBL-AS1* lncRNA and eventually manipulate gene expression in patient cells.

We have also investigated *NIPBL* and *NIPBL-AS1* levels in LCLs of patients with heterozygous truncating mutations in *NIPBL.* While *NIPBL* showed a reduction to 60-70% consistent with previous reports [Liu et al., 2009a; Kaur et al., 2016], *NIPBL-AS1* levels were not significantly affected. Explanations for decreased *NIPBL* levels could be either a downregulation of *NIPBL* gene or a degradation of *NIPBL* mRNA by the nonsense-mediated mRNA decay. In the first case, we would expect that *NIPBL-AS1* levels would change, since our data show that transcriptions of these two genes are interconnected. In the second case, we would not observe a change in *NIPBL-AS1* levels. Since *NIPBL-AS1* levels were indeed not changed, we investigated NMD in patient cell lines. The contribution of wild type and mutated allele in the total RNA was determined by pyrosequencing of transcripts extracted from LCLs of two different patients. We observed percentages of 38% (patient 1) or 24% (patient 3) of mutant *NIPBL* transcripts, which increase up to 46-48% after blocking NMD with cycloheximide. This result indicates at least a partial degradation of mutant *NIPBL* transcripts in patient cells, while both alleles remain actively transcribed. This result explains why *NIPBL-AS1* transcription is not changed and supports our finding that the *NIPBL-AS1* transcript *per se* is not involved in *NIPBL* transcription.

After characterization of regulatory elements such as the enhancer relevant for *NIPBL* expression, we sequenced this enhancer region in a cohort of patients with clinical diagnosis of CdLS. Unfortunately no disease-relevant genetic variant could be identified. This might be explained by the fact that the majority of identified non-coding variants are larger copy number variations (CNVs) that would not be identified by Sanger sequencing. Another reason is that functionally relevant genetic variants in the non-coding regions are so rare that much more than 30 patients have to be included in such a sequencing study. Finally, we have identified and screened only a single enhancer that regulates *NIPBL*. Regarding current knowledge, most genes are regulated by several *cis*-regulatory elements. For instance, the

average number of distant elements interacting with a single transcription start site (TSS) is 3.9 and the average number of TSSs interacting with a distal element is 2.5, indicating a complex regulatory network [The ENCODE project consortium, 2012]. Currently we do not have enough information to interpret SNVs within regulatory elements and we do not know if they would necessarily result in the same phenotype (if any visible). Understanding the non-coding part of the genome is a new field in genetics that is growing, especially with the establishment of use of genome sequencing instead of exome sequencing. Based on current knowledge, it is still unclear how relevant single nucleotide variants within non-coding regulatory elements of the genome are.

4.4 A non-coding regulatory element on chromosome 9 and relevance of the condensin complex for CdLS

As described above, we and others have shown that besides the five known ´CdLS genes´ variants in other chromatin regulators and transcriptional factors can cause a CdLS phenotype. Within the last years, it has become clear that besides the coding part which represents only ~2% of the entire genome, the remaining ~98% of the genome not coding for a protein plays essential roles in gene regulation and disease manifestations [Mattick 2001].

Within our cohort of unsolved cases, we have identified a patient with clinical features that indicated a clinical diagnosis of CdLS, established by Prof. Dr. Gabriele Gillessen-Kaesbach. Moreover, based on the facial features only, this patient was also diagnosed *in silico* as CdLS by Face2Gene suite of phenotyping applications. After we failed to identify a candidate variant in the coding region by different sequencing approaches, we have identified a small deletion on chromosome 9 by array CGH. Subsequent analyses of other family members have identified the same deletion in the mother as well as the younger brother. Further clinical investigations revealed similar clinical features shared by all three individuals. Although the deletion disrupts the *CYLC2* gene, it was excluded to be relevant for the phenotype. *CYLC2* is exclusively expressed in testis and encodes a structural component of the cytoskeletal calyx of mammalian sperm heads. Nevertheless, a cohort of CdLS patients was sequenced for a putative disease-causing variant but no variant was identified.

Additional *in silico* analyses based on a large number of epigenetic and functional data from the ENCODE project were used to analyze the 60 kb deletion for putative regulatory elements. Based on specific modifications of the N-terminal histone tails such as H3K4me1 in different cell types, binding site of various transcription factors (MYC, STAT3, FOS) as well as DNase I sensitivity, a putative enhancer element of 3 kb was predicted within the 60 kb deletion. Further analyses of this genomic region showed the Structural Maintenance of Chromosomes gene − *SMC2* − as the first gene downstream from the predicted enhancer. Moreover, according to publicly available Hi-C data the predicted enhancer and *SMC2* are located within the same topologically associated domain (TAD). It is well known that TADs are conserved among species, cell types and tissues, preventing promiscuous enhancer activity. Co-localization of the enhancer and *SMC2* gene within the same TAD further supported a putative functional interaction of the two genomic elements.

To investigate the interaction between the *SMC2* promoter and the putative enhancer on functional level we have performed reporter gene assays. By this, we could show a very strong activity of the 3 kb enhancer element on the *SMC2* promoter compared to a similar-sized control. Although epigenetic analyses of this 3 kb region strongly indicated an enhancer activity, we could see a total inactivation of the *SMC2* promoter mediated by the 3 kb element in luciferase reporter gene assays. This observation might be explained by technical limitations of the assay such as the distance between the promoter and the enhancer in a circular plasmid in comparison to Mb-scale distance on genomic level. However, this experiment clearly shows a functional interaction between the two investigated regions further supporting relevance of the investigated enhancer. By inserting different small overlapping fragments representing the 3 kb enhancer into the reporter gene plasmid, we narrowed down the critical region that mediates interaction with the *SMC2* promoter to a fragment of ~1 kb maintaining the same activity like the full length construct (figure 34). Interestingly, our result was in accordance with the *in silico* predictions based on ENCODE data that show binding site for several transcription factors, DNase I hypersensitivity as well as H3K4me1, already predicting a weak enhancer activity (figure 33).

After validating interaction between the regulatory element and the *SMC2* promoter, we have obtained fibroblast cell lines of the index patient and his healthy sister for further functional investigations. Western blot analyses showed decreased SMC2 levels in the patient cell line

compared to his sister and additional controls, further supporting our results from *in vitro* analyses and strongly indicating that this regulatory element affects *SMC2* expression. Besides using β-actin as a loading control, we analyzed SMC4 protein levels. SMC4 forms a heterodimer with SMC2, building the structural core of the condensin complex (condensin I and condensin II). Interestingly, SMC4 levels were also decreased in the patient cell line compared to his unaffected sister or controls (figure 35). It has been reported previously that increased SMC4 levels occur in colorectal cancer cells as a consequence of increased *SMC2*/SMC2 levels and that decreased SMC4 levels occur in $SMC2^{OFF}$ cells [Davalos et al., 2012; Ohta et al., 2016], which is in accordance with our western blot result showing the regulation on the protein level. We have further investigated transcript levels and could see that *SMC2* mRNA amounts are decreased in the patient cell line, further supporting our previous findings that the deleted enhancer regulates *SMC2*. Surprisingly, *SMC4* mRNA levels were also dramatically decreased in the patient cell line compared to fibroblasts of his healthy sister and five healthy controls (figure 36). It has been assumed that this regulation occurs exclusively on protein level, but the regulatory mechanisms by which reduced *SMC2*/SMC2 levels affect SMC4 remained unknown. However, our results show decreased *SMC4* transcript levels due to a reduction of *SMC2*/SMC2, clearly indicating novel regulatory mechanisms.

To further confirm the relevance of the 3 kb enhancer region within the 60 kb deletion, we have generated two different cell lines deficient for this 3 kb region by CRISPR/Cas9 genome editing. While four independent homozygous clones were obtained for HEK293 cells, only a single clone was obtained for SH-SY5Y neuroblastoma cells. Expression analyses in HEK293 cells were in line with the results obtained in patient´s fibroblasts with ~50% decrease of *SMC2* and *SMC4* transcript levels. However, the result obtained in SH-SY5Y cells showed ~50% increase of *SMC2* and *SMC4* expression, which is contrary to the reduced expression levels in all other investigated cell types (figure 38). Although these divergent results are still not fully understood on molecular level, there are several putative explanations. Our data generated for the SH-SY5Y cell line were obtained only from a single clone because additional biological replicates were not available. Within this cell line, we cannot exclude potential off-target effects introduced by CRISPR/Cas9 that might affect the transcriptional regulation mechanisms. Although the targeting specificity of Cas9 is believed to be tightly controlled by the 20 bp-gRNA sequence and the presence of PAM (protospacer adjacent motif), potential off-target cleavage activity can still occur [Fu et al., 2013; Hsu et al., 2013; Mali et al., 2013].

Therefore, genome sequencing of the mutant cell lines should be performed in order to exclude functionally relevant off-target effects. Another assumption for obtaining only a single clone in SH-SY5Y cells is that the deletion significantly alters cell metabolisms and viability. The single clone obtained might carry an additional ´protective´ variant. This is further supported by own observations that show decreased growth efficiency of clones with homozygous deletion compared to clones with heterozygous deletion or wild type cells. Another explanation for the discrepancy in results between the different cell lines is that regulatory elements might have different activity in different cell types. Their activity is restricted to a particular tissue or cell type, a time point of life or to specific physiological, pathological or environmental conditions. It has even been reported that a certain regulatory element can either enhance or silence transcription depending on the cellular context [Bessis et al., 1997; Ernst and Kellis, 2010; Soufi et al., 2012; ENCODE project consortium 2012; Heinz et al., 2015]. Therefore, their tissue-specificity makes their identification and interpretation of the results more challenging. Our results show that the investigated 3 kb regulatory region acts as an enhancer in HEK293 and fibroblast cell lines, while it appears to be a silencer in SH-SY5Y neuroblastoma cells. If this discrepancy between the two CRISPR/Cas9 modified cell lines is indeed due to the tissue-specificity or due to additional unknown genetic or epigenetic variants is still unclear and needs to be investigated further.

Since we could show that impaired expression of *SMC2*/SMC2 results in a phenotype that overlaps with CdLS, we aimed to find additional patients with variant within the protein-coding region. After contacting PD. Dr. Tim Strom (Munich) and Dr. Kirsten Cremer (Bonn), an additional patient with a frameshift mutation in *SMC2* (c.1662_1663delAG, p.(Asp555Hisfs*6)) was identified by exome sequencing and the mutation was inherited from his mildly affected mother. Gene expression analyses in fibroblast cells of the ´patient 2´ and his mother showed decreased *SMC2* levels similar to the patient with deletion of the enhancer element. In agreement with our analyses on mutant HEK293 cells generated by CRISPR/Cas9 genome editing as well as fibroblasts of the patient 1, we could also detect reduced *SMC4* transcript levels. These results were in contrast to the western blot analysis that indicated slightly increased protein levels in the patient compared to healthy controls. Interestingly, the same analysis shows no significant alterations with a tendency to decreased protein concentrations in the fibroblasts of the patient´s mildly affected mother (figure 40). Whether these contrary observations of decreased transcript levels but unaffected or even increased protein

concentrations are based on limitations of the technologies used or represent the physiological status can only be speculated. Interpretation of quantitative western blots is often challenging and only obvious differences in protein amounts can be detected. In addition, many steps in the procedure such as loading the samples, their transfer to membrane and detection of the signal can be a prone to error. On the other hand, quantitative measurement of transcript levels by real time PCR is a more robust method than western blotting and we used several housekeeping genes for normalization. Finally, alterations of transcript levels do not always correlate with alterations of protein levels and discrepancies between mRNA and protein levels can occur. Transcription and translation are processes regulated by different mechanisms that can be either connected or independent from each other. For instance, it has been reported that transcription factor THAP1 regulates its own expression by a feedback-loop mechanism, where decreased amounts of protein lead to increased transcript levels [Erogullari et al., 2014]. Reasons for the absence of correlation between mRNA and protein levels include various post-transcriptional and post-translational mechanisms or differences between proteins in their half-lives *in vivo*, although many of these mechanisms are still not well understood [reviewed in Greenbaum et al., 2003].

Patient 2 presents with mild to moderate intellectual impairment, speech delay, aggressive behavior and distinctive facial features such as defined eyebrows, high nasal bridge, thin upper lip and prominent ears. His mother shows significant overlapping features such as mild intellectual impairment, hearing problems and facial features including defined eyebrows with synophrys and thin upper lip, but the index patient is more severely affected. It is tempting to speculate that these differences in the phenotype might be explained by different condensin protein levels. While both individuals carry the same genetic variant in *SMC2*, additional genetic variants, polymorphisms or even unknown epigenetic alterations could further explain the observed differences.

To further validate whether decreased *SMC2* expression directly affects *SMC4* transcript levels as well as SMC2/SMC4 protein concentration, we specifically silenced endogenous *SMC2* by siRNA in HeLa and SH-SY5Y cells. By this, we have shown that reduced SMC2 levels result in decrease of *SMC4*/SMC4, which is in agreement with our observations in patients´ cell lines (figure 41). Since these two proteins create a heterodimer within condensin complexes, their mutual regulation is expected. However, a common regulation of these condensin

components on transcriptional level was not described before. The amounts and activity of protein complexes can be controlled at the level of mRNA transcripts (through transcription and degradation) or protein components (through translation and protein degradation) while control of activity can also be achieved by post-translational modifications such as phosphorylation, controlling the assembly of complexes ´just-in-time´ [Lichtenberg et al., 2005; Jensen et al., 2006]. Several *in silico* studies have reported that ancient protein complexes present in every species are more likely to exhibit co-expression of their components and that transcriptional regulation may be important for these complexes associated with fundamental biomolecular processes [Simonis et al., 2006; Tan et al., 2007; Webb and Westhead 2009]. In line with this, the condensin complex is both highly conserved and ancient, predating histones [Kalitsis et al., 2017]. Therefore, our results showing that condensin complex assembly is already regulated on transcript level is in accordance with these observations. How exactly alterations of *SMC2*/SMC2 levels affect *SMC4*/SMC4 expression is unknown.

In addition, an increasing body of evidence suggests that besides their canonical roles in chromatin condensation, condensin complexes play fundamental roles in long-range genome interactions and transcription regulation [Kim et al., 2013; Huang et al., 2013; Murakami-Tonami et al., 2014; Iwasaki et al., 2015; Kim et al., 2016]. These findings are further supported by unpublished data that indicate misregulation of a group of chromatin-associated genes caused by mutations in different condensin subunits (own results and personal communication with collaborators).

By subsequent functional investigations using patient cell lines as well as CRISPR/Cas9-modified cells, we could identify alterations in canonical condensin functions. Cell cycle analyses of all the investigated cell lines show an increased amount of cells in G1 phase and decreased amount of cells in G2/M phase of the cell cycle. Since G1 arrest occurs in response to DNA damage and SMC2 transcriptionally regulates DNA damage response genes in cooperation with MYCN [Murakami-Tonami et al., 2014], our result of disturbed cell cycle progression is expected. Although the investigated metaphase spreads appear to be normal, we have shown that the cell lines of the patients lose their ability to recover chromosomes after stressing them with law-salt TEEN buffer in accordance with the previous reports [Hudson et al., 2003; Martin et al., 2016].

In summary, this project started with the functional characterization of a predicted enhancer element within a ~60 kb deletion identified in three related patients, all clinically diagnosed as CdLS-like. Subsequent functional investigations could verify the interaction of this element with the 1 Mb distant *SMC2* promoter and its regulatory activity on *SMC2* expression. In addition, we present experimental evidence that misregulation of *SMC2* transcription affects expression of *SMC4* by an unknown mechanism. SMC2 is the core subunit of both condensin I and condensin II complexes, well conserved throughout evolution and essential for the cell survival. Although numerous studies have been performed to characterize condensin function, variants in genes encoding condensin components are still rare. Most of mutations have been associated with various cancers [Davalos et al., 2012; Je et al., 2014; Woodward et al., 2016; Zhang et al., 2016], while biallelic mutations in non-SMC subunits have recently been associated with microcephaly, short stature and intellectual disability [Martin et al., 2016]. Of note, a single patient with a 12 Mb deletion on chromosome 9, that includes also *SMC2*, has been reported to share common features with Cornelia de Lange syndrome [Cao et al., 2015]. However, this deletion includes several genes some of which also associated with other phenotypes/syndromes. According to the data of a large exome study including 60,706 exomes from the Broad institute (ExAC study) *SMC2* is intolerant to loss of function mutations (probability of LoF intolerance, pLI=1) which is consistent with its essential role and a very low number of reported cases. Therefore, our patient with the frameshift variant in *SMC2* represents the first case with a distinctive phenotype (other than cancer) with a mutation in this gene. He displays certain CdLS-overlapping features such as intellectual disability, speech delay, aggressive behavior, high nasal bridge, defined eyebrows, thin upper lip and feeding difficulties at birth. The three patients with the deletion of the *SMC2* enhancer (´patient 1´, his brother and mother) were diagnosed as CdLS-like because of the facial features, mild intellectual disability, aggressive behavior, hearing loss, hirsutism etc. Although there are certainly some clinical overlaps between these patients, in has to be taken into account that variants in regulatory elements might lead not only to milder but also to different consequences than variants the protein coding parts of the respective genes. This might be partially explained by tissue-specific activity of regulatory elements that restricts regulation of the target genes to specific tissues or developmental processes as well as by data showing that a regulatory element usually controls expression of more genes, just like a single gene has more than one regulatory element to control specific expression pattern [The ENCODE

project consortium, 2012]. Identification of regulatory elements is crucial for understanding spatiotemporal transcriptional regulation during development as well as the functional impact of disease-related noncoding genetic variants. Therefore, regulatory elements have to be identified and functionally characterized to understand complex regulation of gene expression that is of great relevance for haploinsufficiency syndromes such as CdLS.

With the development and dropping costs of high-throughput sequencing technologies such as whole genome sequencing, ChIP-seq and chromosome conformation capture, importance of non-coding DNA in gene regulation and 3D chromatin folding got into the focus of numerous research projects. Loss of function and gain of function mutations in regulatory elements, disruption of TADs, and impact of GWAS-identified SNPs have been described in context of various monogenic and complex traits and diseases [Klopocki et al., 2008; Weedon et al., 2014; Claussnitzer et al., 2015; Lupianez et at., 2015]. Of note, most of the genome-wide association study (GWAS) signals map to non-coding regions and point to non-coding variants, although their functional interpretation is challenging. Accumulating evidence has shown that GWAS variants are enriched in the functional non-coding regions such as enhancers, DNase hypersensitivity regions and chromatin marks [Degner et al., 2012; Trynka et al., 2013]. In addition to SNVs/SNPs, large-scale structural variations (including CNVs) often encompass more than one gene. Besides intergenic regions, non-coding regions within a gene and non-coding RNAs could also be affected and therefore need to be taken into consideration. Finally, a compound model of inheritance of a regulatory SNP and CNVs has also been described [Albers et al., 2012; Wieczorek et al., 2014; Wu et al., 2015].

Variants in the coding regions have been considered the leading causes of Mendelian and various complex disorders. Accordingly, Sanger-, gene panel- and exome sequencing are still widely used in molecular diagnostic. However, recent progress in genetic and genomic studies has delineated the importance of functional genetic variants in non-coding regions of genome. Therefore, we expect that in the following years whole genome sequencing (WGS) will be the preferred technology in molecular diagnostics together with technical and methodological innovations. Understanding the non-coding and regulatory genome will give us insights into mechanisms that are still not well understood, reduced penetrance, phenotypical variation, genotype-phenotype correlation etc. Finally, new insights will increase the number of patients

who receive proper genetic/molecular diagnosis with a hope to treat the cause of the disease rather than palliate the consequences.

References

Albers, C.A., Paul, D.S., Schulze, H., Freson, K., Stephens, J.C., Smethurst, P.A., Jolley, J.D., Cvejic, A., Kostadima, M., Bertone, P., Breuning, M.H., Debili, N., Deloukas, P., Favier, R., Fiedler, J., Hobbs, C.M., Huang, N., Hurles, M.E., Kiddle, G., Krapels, I., Nurden, P., Ruivenkamp, C.A.L., Sambrook, J.G., Smith, K., Stemple, D.L., Strauss, G., Thys, C., Van Geet, C., Newbury-Ecob, R., Ouwehand, W.H., Ghevaert, C., 2012. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat. Genet. 44, 435–439. doi:10.1038/ng.1083

Allanson, J.E., Hennekam, R.C., Ireland, M., 1997. De Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. J. Med. Genet. 34, 645–50.

Ansari, M., Poke, G., Ferry, Q., Williamson, K., Aldridge, R., Meynert, A.M., Bengani, H., Chan, C.Y., Kayserili, H., Avci, S., Hennekam, R.C.M., Lampe, A.K., Redeker, E., Homfray, T., Ross, A., Falkenberg Smeland, M., Mansour, S., Parker, M.J., Cook, J.A., Splitt, M., Fisher, R.B., Fryer, A., Magee, A.C., Wilkie, A., Barnicoat, A., Brady, A.F., Cooper, N.S., Mercer, C., Deshpande, C., Bennett, C.P., Pilz, D.T., Ruddy, D., Cilliers, D., Johnson, D.S., Josifova, D., Rosser, E., Thompson, E.M., Wakeling, E., Kinning, E., Stewart, F., Flinter, F., Girisha, K.M., Cox, H., Firth, H. V, Kingston, H., Wee, J.S., Hurst, J.A., Clayton-Smith, J., Tolmie, J., Vogt, J., Tatton-Brown, K., Chandler, K., Prescott, K., Wilson, L., Behnam, M., McEntagart, M., Davidson, R., Lynch, S.-A., Sisodiya, S., Mehta, S.G., McKee, S.A., Mohammed, S., Holden, S., Park, S.-M., Holder, S.E., Harrison, V., McConnell, V., Lam, W.K., Green, A.J., Donnai, D., Bitner-Glindzicz, M., Donnelly, D.E., Nellåker, C., Taylor, M.S., FitzPatrick, D.R., 2014. Genetic heterogeneity in Cornelia de Lange syndrome (CdLS) and CdLS-like phenotypes with observed and predicted levels of mosaicism. J. Med. Genet. 51, 659–68. doi:10.1136/jmedgenet-2014-102573

Bagchi, D.N., Iyer, V.R., 2016. The Determinants of Directionality in Transcriptional Initiation. Trends Genet. doi:10.1016/j.tig.2016.03.005

Bessis, A., Champtiaux, N., Chantelin, L., Changeux, J.P., 1997. The neuron-restrictive silencer element: a dual enhancer/silencer crucial for patterned expression of a nicotinic receptor gene in the brain. Proc Natl Acad Sci USA. 1997 May 27;94(11):5906-11

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M.T., De Marco, R., Damante, G., Grainger, R., Van Heyningen, V., Kleinjan, D.A., 2013. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. Am. J. Hum. Genet. 93, 1126–1134. doi:10.1016/j.ajhg.2013.10.028

Bhuiyan, Z.A., Klein M., Hammond P., van Haeringen A., Mannens M.M., Van Berckelaer-Onnes I., Hennekam R.C., 2006. Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. J. Med. Genet. 43, 568–575. doi:10.1136/jmg.2005.038240

Brachmann, W. Ein Fall von symmetrischer Monodaktylie durch Ulnadefekt, mit symmetrischer Flughautbildung in den Ellenbeugen, sowie anderen Abnormitaeten (Zwerghaftigkeit, Halsrippen, Behaarung) (A case of symmetrical monodactyly, representing ulnar deficiency, with symmetrical antecubital webbing and other abnormalities, (dwarfism, cervical ribs, hirsutism)). *Jahrbuch fuer Kinderheilkunde und physische Erziehung* 84: 225-235, 1916.

Cao, R., Pu, T., Fang, S., Long, F., Xie, J., Xu, Y., Chen, S., Sun, K., Xu, R., 2015. Patients Carrying 9q31.1-q32 Deletion Share Common Features with Cornelia de Lange Syndrome. Cell. Physiol. Biochem. 35, 270–280. doi:10.1159/000369694

Castronovo, P., Gervasini, C., Cereda, A., Masciadri, M., Milani, D., Russo, S., Selicorni, A., Larizza, L., 2009. Premature chromatid separation is not a useful diagnostic marker for Cornelia de Lange syndrome. Chromosom. Res. 17, 763–771. doi:10.1007/s10577-009-9066-6

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., Nasmyth, K., 2000. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol. Cell 5, 243–54. doi:10.1016/S1097-2765(00)80420-7

Claussnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., Abdennur, N.A., Liu, J., Svensson, P.-A., Hsu, Y.-H., Drucker, D.J., Mellgren, G., Hui, C.-C., Hauner, H., Kellis, M., 2015. *FTO* Obesity Variant Circuitry and Adipocyte Browning in Humans. N. Engl. J. Med. 373, 895–907. doi:10.1056/NEJMoa1502214

D'Ambrosio, C., Schmidt, C.K., Katou, Y., Kelly, G., Itoh, T., Shirahige, K., Uhlmann, F., 2008. Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. Genes Dev. 22, 2215–2227. doi:10.1101/gad.1675708

Dávalos, V., Súarez-López, L., Castaño, J., Messent, A., Abasolo, I., Fernandez, Y., Guerra-Moreno, A., Espín, E., Armengol, M., Musulen, E., Ariza, A., Sayós, J., Arango, D., Schwartz, S., 2012. Human SMC2 protein, a core subunit of human condensin complex, is a novel transcriptional target of the WNT signaling pathway and a new therapeutic target. J. Biol. Chem. 287, 43472–43481. doi:10.1074/jbc.M112.428466

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

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

Deardorff, M.A., Wilde, J.J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Mönnich, M., Yan, Y., Xu, W., Gil-Rodríguez, M.C., Clark, D., Hakonarson, H., Halbach, S.,

Michelis, L.D., Rampuria, A., Rossier, E., Spranger, S., Van Maldergem, L., Lynch, S.A., Gillessen-Kaesbach, G., Lüdecke, H.J., Ramsay, R.G., McKay, M.J., Krantz, I.D., Xu, H., Horsfield, J.A., Kaiser, F.J., 2012 (2012a). RAD21 mutations cause a human cohesinopathy. Am. J. Hum. Genet. 90, 1014–1027. doi:10.1016/j.ajhg.2012.04.019

Deardorff, M.A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K.E., De Baere, E., Decroos, C., Di Donato, N., Ernst, S., Francey, L.J., Gyftodimou, Y., Hirashima, K., Hullings, M., Ishikawa, Y., Jaulin, C., Kaur, M., Kiyono, T., Lombardi, P.M., Magnaghi-Jaulin, L., Mortier, G.R., Nozaki, N., Petersen, M.B., Seimiya, H., Siu, V.M., Suzuki, Y., Takagaki, K., Wilde, J.J., Willems, P.J., Prigent, C., Gillessen-Kaesbach, G., Christianson, D.W., Kaiser, F.J., Jackson, L.G., Hirota, T., Krantz, I.D., Shirahige, K., 2012 (2012b). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature 489, 313–7. doi:10.1038/nature11316

Deardorff, M.A., Porter, N.J., Christianson, D.W., 2016. Structural aspects of HDAC8 mechanism and dysfunction in Cornelia de Lange syndrome spectrum disorders. Protein Sci. doi:10.1002/pro.3030

Decroos, C., Bowman, C.M., Moser, J.A.S., Christianson, K.E., Deardorff, M.A., Christianson, D.W., 2014. Compromised structure and function of HDAC8 mutants identified in Cornelia de Lange Syndrome spectrum disorders. ACS Chem. Biol. 9, 2157–2164. doi:10.1021/cb5003762

Decroos, C., Christianson, N.H., Gullett, L.E., Bowman, C.M., Christianson, K.E., Deardorff, M.A., Christianson, D.W., 2015. Biochemical and Structural Characterization of HDAC8 Mutants Associated with Cornelia de Lange Syndrome Spectrum Disorders. Biochemistry 54, 6501–6513. doi:10.1021/acs.biochem.5b00881

Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., Stephens, M., Gilad, Y., Pritchard, J.K., 2012. DNase-I sensitivity QTLs are a major determinant of human expression variation. Nature 482, 390–394. doi:10.1038/nature10808

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380. doi:10.1038/nature11082

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. doi:10.1038/nature11247

Erogullari, A., Hollstein, R., Seibler, P., Braunholz, D., Koschmidder, E., Depping, R., Eckhold, J., Lohnau, T., Gillessen-Kaesbach, G., Grünewald, A., Rakovic, A., Lohmann, K., Kaiser, F.J., 2014. THAP1, the gene mutated in DYT6 dystonia, autoregulates its own expression. Biochim. Biophys. Acta - Gene Regul. Mech. 1839, 1196–1204. doi:10.1016/j.bbagrm.2014.07.019

Freed, D., Stevens, E.L., Pevsner, J., 2014. Somatic mosaicism in the human genome. Genes (Basel). doi:10.3390/genes5041064

Frosi, Y., Haering, C.H., 2015. Control of chromosome interactions by condensin complexes. Curr. Opin. Cell Biol. doi:10.1016/j.ceb.2015.05.008

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826. doi:10.1038/nbt.2623

Gervasini, C., Russo, S., Cereda, A., Parenti, I., Masciadri, M., Azzollini, J., Melis, D., Aravena, T., Doray, B., Ferrarini, A., Garavelli, L., Selicorni, A., Larizza, L., 2013. Cornelia de Lange individuals with new and recurrent SMC1A mutations enhance delineation of mutation repertoire and phenotypic spectrum. Am. J. Med. Genet. Part A 161, 2909–2919. doi:10.1002/ajmg.a.36252

Gillis, L.A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A.D., Li, H., Devoto, M., Jackson, L.G., Krantz, I.D., 2004. NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am. J. Hum. Genet. 75, 610–623. doi:10.1086/424698

Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.M., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H.G., de Vries, B.B.A., Kleefstra, T., Brunner, H.G., Vissers, L.E.L.M., Veltman, J.A., 2014. Genome sequencing identifies major causes of severe intellectual disability. Nature 511, 344–347. doi:10.1038/nature13394

Gillespie, P.J., Hirano, T., 2004. Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. Curr. Biol. 14, 1598–1603. doi:10.1016/j.cub.2004.07.053

Goldstein, J.H.R., Tim-aroon, T., Shieh, J., Merrill, M., Deeb, K.K., Zhang, S., Bass, N.E., Bedoyan, J.K., 2015. Novel SMC1A frameshift mutations in children with developmental delay and epilepsy. Eur. J. Med. Genet. 58, 562–568. doi:10.1016/j.ejmg.2015.09.007

Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M., 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol. doi:10.1186/gb-2003-4-9-117

Grozeva, D., Carss, K., Spasic-Boskovic, O., Parker, M.J., Archer, H., Firth, H. V., Park, S.M., Canham, N., Holder, S.E., Wilson, M., Hackett, A., Field, M., Floyd, J.A.B., Hurles, M., Lucy Raymond, F., 2014. De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. Am. J. Hum. Genet. 94, 618–624. doi:10.1016/j.ajhg.2014.03.006

Gruber, S., Haering, C.H., Nasmyth, K., 2003. Chromosomal cohesin forms a ring. Cell 112, 765–777. doi:10.1016/S0092-8674(03)00162-4

Gullerova, M., Proudfoot, N.J., 2008. Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S. pombe. Cell 132, 983–995. doi:10.1016/j.cell.2008.02.040

Haberland, M., Mokalled, M.H., Montgomery, R.L., Olson, E.N., 2009. Epigenetic control of skull morphogenesis by histone deacetylase 8. Genes Dev. 23, 1625–1630. doi:10.1101/gad.1809209

Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., Merkenschlager, M., 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature 460, 410–3. doi:10.1038/nature08079

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

Haering, C.H., Jessberger, R., 2012. Cohesin in determining chromosome architecture. Exp. Cell Res. doi:10.1016/j.yexcr.2012.03.016

Hakimi, M.-A., Bochar, D. a, Schmiesing, J. a, Dong, Y., Barak, O.G., Speicher, D.W., Yokomori, K., Shiekhattar, R., 2002. A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature 418, 994–998. doi:10.1038/nature01024

Halal, F., Preus, M., 1979. The hand profile on de Lange syndrome: diagnostic criteria. Am J Med Genet 3, 317–323.

Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R., Davies, B., Higgs, D.R., 2017. Tissue-specific CTCF–cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat. Cell Biol. 19. doi:10.1038/ncb3573

Harakalova, M., van den Boogaard, M.-J., Sinke, R., van Lieshout, S., van Tuil, M.C., Duran, K., Renkens, I., Terhal, P. a., de Kovel, C., Nijman, I.J., van Haelst, M., Knoers, N.V. a. M., van Haaften, G., Kloosterman, W., Hennekam, R.C.M., Cuppen, E., Ploos van Amstel, H.K., 2012. X-exome sequencing identifies a HDAC8 variant in a large pedigree with X-linked intellectual disability, truncal obesity, gynaecomastia, hypogonadism and unusual face. J. Med. Genet. 49, 539–543. doi:10.1136/jmedgenet-2012-100921

Hirano, T., Mitchison, T.J., 1994. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. Cell 79, 449–458. doi:10.1016/0092-8674(94)90254-2

Hirota, T., 2004. Distinct functions of condensin I and II in mitotic chromosome assembly. J. Cell Sci. 117, 6435–6445. doi:10.1242/jcs.01604

Hoppman-Chaney, N., Jang, J.S., Jen, J., Babovic-Vuksanovic, D., Hodge, J.C., 2012. In-frame multi-exon deletion of SMC1A in a severely affected female with Cornelia de Lange Syndrome. Am. J. Med. Genet. Part A 158 A, 193–198. doi:10.1002/ajmg.a.34360

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832. doi:10.1038/nbt.2647

Huang, C.E., Milutinovich, M., Koshland, D., 2005. Rings, bracelet or snaps: fashionable alternatives for Smc complexes. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360, 537–542. doi:10.1098/rstb.2004.1609

Huang, K., Jia, J., Wu, C., Yao, M., Li, M., Jin, J., Jiang, C., Cai, Y., Pei, D., Pan, G., Yao, H., 2013. Ribosomal RNA gene transcription mediated by the master genome regulator protein CCCTC-binding factor (CTCF) is negatively regulated by the condensin complex. J. Biol. Chem. 288, 26067–26077. doi:10.1074/jbc.M113.486175

Hudson, D.F., Vagnarelli, P., Gassmann, R., Earnshaw, W.C., 2003. Condensin is required for nonhistone protein assembly and structural integrity of vertebrate mitotic chromosomes. Dev. Cell 5, 323–336. doi:10.1016/S1534-5807(03)00199-0

Huisman, S.A., Redeker, E.J.W., Maas, S.M., Mannens, M.M., Hennekam, R.C.M., 2013. High rate of mosaicism in individuals with Cornelia de Lange syndrome. J. Med. Genet. 50, 339–44. doi:10.1136/jmedgenet-2012-101477

Huisman, S., Mulder, P.A., Redeker, E., Bader, I., Bisgaard, A.M., Brooks, A., Cereda, A., Cinca, C., Clark, D., Cormier-Daire, V., Deardorff, M.A., Diderich, K., Elting, M., van Essen, A., Patrick, D.F., Gervasini, C., Gillessen-Kaesbach, G., Girisha, K.M., Hilhorst-Hofstee, Y., Hopman, S., Horn, D., Isrie, M., Jansen, S., Jespersgaard, C., Kaiser, F.J., Kaur, M., Kleefstra, T., Krantz, I.D., Lakeman, P., Landlust, A., Lessel, D., Michot, C., Moss, J., Noon, S.E., Oliver, C., Parenti, I., Pie, J., Ramos, F.J., Rieubland, C., Russo, S., Selicorni, A., Tümer, Z., Vorstenbosch, R., Wenger, T.L., van Balkom, I., Piening, S., Wierzba, J., Hennekam, R.C., 2017. Phenotypes and genotypes in individuals with SMC1A variants. Am. J. Med. Genet. Part A. doi:10.1002/ajmg.a.38279

Ivanov, D., Nasmyth, K., 2005. A topological interaction between cohesin rings and a circular minichromosome. Cell 122, 849–860. doi:10.1016/j.cell.2005.07.018

Izumi, K., Nakato, R., Zhang, Z., Edmondson, A.C., Noon, S., Dulik, M.C., Rajagopalan, R., Venditti, C.P., Gripp, K., Samanich, J., Zackai, E.H., Deardorff, M.A., Clark, D., Allen, J.L., Dorsett, D., Misulovin, Z., Komata, M., Bando, M., Kaur, M., Katou, Y., Shirahige, K., Krantz, I.D., 2015. Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat. Genet. 47, 338–344. doi:10.1038/ng.3229

Jackson, L., Kline, A.D., Barr, M.A., Koch, S., 1993. de Lange syndrome: A clinical review of 310 individuals, in: American Journal of Medical Genetics. pp. 940–946. doi:10.1002/ajmg.1320470703

Jahnke, P., Xu, W., Wülling, M., Albrecht, M., Gabriel, H., Gillessen-Kaesbach, G., Kaiser, F.J., 2008. The Cohesin loading factor NIPBL recruits histone deacetylases to mediate local chromatin modifications. Nucleic Acids Res. 36, 6450–6458. doi:10.1093/nar/gkn688

Jansen, S., Kleefstra, T., Willemsen, M.H., de Vries, P., Pfundt, R., Hehir-Kwa, J.Y., Gilissen, C., Veltman, J.A., de Vries, B.B.A., Vissers, L.E.L.M., 2016. De novo loss-of-function mutations in X-linked SMC1A cause severe ID and therapy-resistant epilepsy in females: expanding the phenotypic spectrum. Clin. Genet. 90, 413–419. doi:10.1111/cge.12729

Je, E.M., Yoo, N.J., Lee, S.H., 2014. Mutational and expressional analysis of SMC2 gene in gastric and colorectal cancers with microsatellite instability. APMIS 122, 499–504. doi:10.1111/apm.12193

Jensen, L.J., Jensen, T.S., De Lichtenberg, U., Brunak, S., Bork, P., 2006. Co-evolution of transcriptional and post-translational cell-cycle regulation. Nature 443, 594–597. doi:10.1038/nature05186

Jones, W.D., Dafou, D., McEntagart, M., Woollard, W.J., Elmslie, F. V., Holder-Espinasse, M., Irving, M., Saggar, A.K., Smithson, S., Trembath, R.C., Deshpande, C., Simpson, M.A., 2012. De novo mutations in MLL cause Wiedemann-Steiner syndrome. Am. J. Hum. Genet. 91, 358–364. doi:10.1016/j.ajhg.2012.06.008

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J., Young, R.A., 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435. doi:10.1038/nature09380

Kaiser, F.J., Ansari, M., Braunholz, D., Gil-Rodríguez, M.C., Decroos, C., Wilde, J.J., Fincher, C.T., Kaur, M., Bando, M., Amor, D.J., Atwal, P.S., Bahlo, M., Bowman, C.M., Bradley, J.J., Brunner, H.G., Clark, D., Campo, M. Del, Di Donato, N., Diakumis, P., Dubbs, H., Dyment, D.A., Eckhold, J., Ernst, S., Ferreira, J.C., Francey, L.J., Gehlken, U., Guillén-Navarro, E., Gyftodimou, Y., Hall, B.D., Hennekam, R., Hudgins, L., Hullings, M., Hunter, J.M., Yntema, H., Innes, A.M., Kline, A.D., Krumina, Z., Lee, H., Leppig, K., Lynch, S.A., Mallozzi, M.B., Mannini, L., Mckee, S., Mehta, S.G., Micule, I., Consortium, C.R.C., Mohammed, S., Moran, E., Mortier, G.R., Moser, J.A.S., Noon, S.E., Nozaki, N., Nunes, L., Pappas, J.G., Penney, L.S., Pérez-Aytés, A., Petersen, M.B., Puisac, B., Revencu, N., Roeder, E., Saitta, S., Scheuerle, A.E., Schindeler, K.L., Siu, V.M., Stark, Z., Strom, S.P., Thiese, H., Vater, I., Willems, P., Williamson, K., Wilson, L.C., Hakonarson, H., Quintero-Rivera, F., Wierzba, J., Musio, A., Gillessen-Kaesbach, G., Ramos, F.J., Jackson, L.G., Shirahige, K., Pié, J., Christianson, D.W., Krantz, I.D., Fitzpatrick, D.R., Deardorff, M.A., 2014. Loss-of-function HDAC8 mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. Hum. Mol. Genet. 23, 2888–2900. doi:10.1093/hmg/ddu002

Kaur, M., Mehta, D., Noon, S.E., Deardorff, M.A., Zhang, Z., Krantz, I.D., 2016. NIPBL expression levels in CdLS probands as a predictor of mutation type and phenotypic severity. Am. J. Med. Genet. Part C Semin. Med. Genet. 172, 163–170. doi:10.1002/ajmg.c.31495

Kawauchi, S., Calof, A.L., Santos, R., Lopez-Burks, M.E., Young, C.M., Hoang, M.P., Chua, A., Lao, T., Lechner, M.S., Daniel, J.A., Nussenzweig, A., Kitzes, L., Yokomori, K., Hallgrimsson, B., Lander, A.D., 2009. Multiple organ system defects and transcriptional dysregulation in the Nipbl+/- mouse, a model of Cornelia de Lange syndrome. PLoS Genet. 5. doi:10.1371/journal.pgen.1000650

Kim, J.H., Zhang, T., Wong, N.C., Davidson, N., Maksimovic, J., Oshlack, A., Earnshaw, W.C., Kalitsis, P., Hudson, D.F., 2013. Condensin I associates with structural and gene regulatory regions in vertebrate chromosomes. Nat. Commun. 4. doi:10.1038/ncomms3537

Kim, J.S., Krasieva, T.B., LaMorte, V., Malcolm, A., Taylor, R., Yokomori, K., 2002. Specific recruitment of human cohesin to laser-induced DNA damage. J. Biol. Chem. 277, 45149–45153. doi:10.1074/jbc.M209123200

Kim, K.D., Tanizawa, H., Iwasaki, O., Noma, K.I., 2016. Transcription factors mediate condensin recruitment and global chromosomal organization in fission yeast. Nat. Genet. 48, 1242–1252. doi:10.1038/ng.3647

Kline, A.D., Krantz, I.D., Sommer, A., Kliewer, M., Jackson, L.G., FitzPatrick, D.R., Levin, A. V., Selicorni, A., 2007. Cornelia de Lange syndrome: Clinical review, diagnostic and scoring systems, and anticipatory guidance. Am. J. Med. Genet. Part A. doi:10.1002/ajmg.a.31757

Klopocki, E., Ott, C.E., Benatar, N., Ullmann, R., Mundlos, S., Lehmann, K., 2008. A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. J. Med. Genet. 45, 370–375. doi:10.1136/jmg.2007.055699

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

Kuechler, A., Zink, A.M., Wieland, T., Lüdecke, H.-J., Cremer, K., Salviati, L., Magini, P., Najafi, K., Zweier, C., Czeschik, J.C., Aretz, S., Endele, S., Tamburrino, F., Pinato, C., Clementi, M., Gundlach, J., Maylahn, C., Mazzanti, L., Wohlleber, E., Schwarzmayr, T., Kariminejad, R., Schlessinger, A., Wieczorek, D., Strom, T.M., Novarino, G., Engels, H., 2015. Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. Eur. J. Hum. Genet. 23, 753–760. doi:10.1038/ejhg.2014.165

Laugsch, M., Seebach, J., Schnittler, H., Jessberger, R., 2013. Imbalance of SMC1 and SMC3 Cohesins Causes Specific and Distinct Effects. PLoS One 8. doi:10.1371/journal.pone.0065149

Lavoie, B.D., Hogan, E., Koshland, D., 2004. In vivo requirements for rDNA chromosome condensation reveal two cell-cycle-regulated pathways for mitotic chromosome folding. Genes Dev. 18, 76–87. doi:10.1101/gad.1150404

Lebrun, N., Lebon, S., Jeannet, P.-Y., Jacquemont, S., Billuart, P., Bienvenu, T., 2015. Early-onset encephalopathy with epilepsy associated with a novel splice site mutation in SMC1A. Am. J. Med. Genet. A. doi:10.1002/ajmg.a.37364

de Lichtenberg, U., Jensen, L.J., Brunak, S., Bork, P., 2005. Dynamic complex formation during the yeast cell cycle. Science 307, 724–727. doi:10.1126/science.1105103

Liu, J., Krantz, I.D., 2009. Cornelia de Lange syndrome, cohesin, and beyond. Clin. Genet. doi:10.1111/j.1399-0004.2009.01271.x

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., Spinner, N.B., Vega, H., Jackson, L.G., Shirahige, K., Krantz, I.D., 2009 (2009a). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS Biol. 7, e1000119. doi:10.1371/journal.pbio.1000119

Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A., Uhlmann, F., 2014. The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. Nat. Genet. 46, 1147–51. doi:10.1038/ng.3080

Losada, A., Hirano, T., 2001. Shaping the metaphase chromosome: Coordination of cohesion and condensation. BioEssays. doi:10.1002/bies.1133

Losada, A., Hirano, T., 2005. Dynamic molecular linkers of the genome: The first decade of SMC proteins. Genes Dev. doi:10.1101/gad.1320505

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S.A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., Mundlos, S., 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012–1025. doi:10.1016/j.cell.2015.04.004

Lupski, J.R., Belmont, J.W., Boerwinkle, E., Gibbs, R.A., 2011. Clan genomics and the complex architecture of human disease. Cell. doi:10.1016/j.cell.2011.09.008

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., Church, G.M., 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838. doi:10.1038/nbt.2675

Mannini, L., Liu, J., Krantz, I.D., Musio, A., 2010. Spectrum and consequences of SMC1A mutations: The unexpected involvement of a core component of cohesin in human disease. Hum. Mutat. doi:10.1002/humu.21129

Mannini, L., Cucco, F., Quarantotti, V., Krantz, I.D., Musio, A., 2013. Mutation Spectrum and Genotype-Phenotype Correlation in Cornelia de Lange Syndrome. Hum. Mutat. 34, 1589–1596. doi:10.1002/humu.22430

Martin, C.A., Murray, J.E., Carroll, P., Leitch, A., Mackenzie, K.J., Halachev, M., Fetit, A.E., Keith, C., Bicknell, L.S., Fluteau, A., Gautier, P., Hall, E.A., Joss, S., Soares, G., Silva, J., Bober, M.B., Duker, A., Wise, C.A., Quigley, A.J., Phadke, S.R., Wood, A.J., Vagnarelli, P., Jackson, A.P., 2016. Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis. Genes Dev. 30, 2158–2172. doi:10.1101/gad.286351.116

Mattick, J.S., 2001. Non-coding RNAs: The architects of eukaryotic complexity. EMBO Rep. 2, 986–991. doi:10.1093/embo-reports/kve230

Mehta, G.D., Kumar, R., Srivastava, S., Ghosh, S.K., 2013. Cohesin: Functions beyond sister chromatid cohesion. FEBS Lett. doi:10.1016/j.febslet.2013.06.035

Mehta, D., Vergano, S.A.S., Deardorff, M., Aggarwal, S., Barot, A., Johnson, D.M., Miller, N.F., Noon, S.E., Kaur, M., Jackson, L., Krantz, I.D., 2016. Characterization of limb differences in children with Cornelia de Lange Syndrome. Am. J. Med. Genet. Part C Semin. Med. Genet. doi:10.1002/ajmg.c.31498

Merkenschlager, M., Odom, D.T., 2013. CTCF and cohesin: Linking gene regulatory elements with their targets. Cell. doi:10.1016/j.cell.2013.02.029

Michaelis, C., Ciosk, R., Nasmyth, K., 1997. Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. Cell 91, 35–45. doi:10.1016/S0092-8674(01)80007-6

Minajigi, A., Froberg, J.E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., Lee, J.T., 2015. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. Science (80-.). 349, 1DUIMMY. doi:10.1126/science.aab2276

Misulovin, Z., Schwartz, Y.B., Li, X.Y., Kahn, T.G., Gause, M., MacArthur, S., Fay, J.C., Eisen, M.B., Pirrotta, V., Biggin, M.D., Dorsett, D., 2008. Association of cohesin and Nipped-B with transcriptionally active regions of the Drosophila melanogaster genome. Chromosoma 117, 89–102. doi:10.1007/s00412-007-0129-1

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., Grewal, S.I.S., 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature 516, 432–5. doi:10.1038/nature13833

Murakami-Tonami, Y., Kishida, S., Takeuchi, I., Katou, Y., Maris, J.M., Ichikawa, H., Kondo, Y., Sekido, Y., Shirahige, K., Murakami, H., Kadomatsu, K., 2014. Inactivation of SMC2 shows a synergistic lethal response in MYCN-amplified neuroblastoma cells. Cell Cycle 13, 1115–1131. doi:10.4161/cc.27983

Musio, A., Marrella, V., Sobacchi, C., Rucci, F., Fariselli, L., Giliani, S., Lanzi, G., Notarangelo, L.D., Delia, D., Colombo, R., Vezzoni, P., Villa, A., 2005. Damaging-agent sensitivity of Artemis-deficient cell lines. Eur. J. Immunol. 35, 1250–1256. doi:10.1002/eji.200425555

Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S., Vezzoni, P., Larizza, L., 2006. X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. Nat. Genet. 38, 528–530. doi:10.1038/ng1779

Nasmyth, K., Haering, C.H., 2005. The Structure and Function of SMC and Kleisin Complexes. Annu. Rev. Biochem 74, 595–648. doi:10.1146/annurev.biochem.74.082803.133219

Nasmyth, K., Haering, C.H., 2009. Cohesin: Its Roles and Mechanisms. Annu. Rev. Genet. 43, 525–558. doi:10.1146/annurev-genet-102108-134233

Nativio, R., Wendt, K.S., Ito, Y., Huddleston, J.E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.M., Murrell, A., 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. PLoS Genet. 5. doi:10.1371/journal.pgen.1000739

Neuwald, A.F., Hirano, T., 2000. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res. 10, 1445–1452. doi:10.1101/gr.147400

Niu, D.-M., Huang, J.-Y., Li, H.-Y., Liu, K.-M., Wang, S.-T., Chen, Y.-J., Udaka, T., Izumi, K., Kosaki, K., 2006. Paternal gonadal mosaicism of NIPBL mutation in a father of siblings with Cornelia de Lange syndrome. Prenat. Diagn. 26, 1054–1057. doi:10.1002/pd.1554

Ohta, S., Montaňo-Gutierrez, L.F., Alves, F.D.L., Ogawa, H., Toramoto, I., Sato, N., Morrison, C.G., Takeda, S., Hudson, D.F., Rappsilber, J., Earnshaw, W.C., 2016. Proteomics analysis with a nano Random Forest approach reveals novel functional interactions regulated by SMC complexes on mitotic chromosomes. Mol. Cell. Proteomics 2802–2818. doi:10.1074/mcp.M116.057885

Osipovich, A.B., Gangula, R., Vianna, P.G., Magnuson, M.A., 2016. Setd5 is essential for mammalian development and the co-transcriptional regulation of histone acetylation. Development 143, 4595–4607. doi:10.1242/dev.141465

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G., Merkenschlager, M., 2008. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell 132, 422–433. doi:10.1016/j.cell.2008.01.011

Parenti, I., Rovina, D., Masciadri, M., Cereda, A., Azzollini, J., Picinelli, C., Limongelli, G., Finelli, P., Selicorni, A., Russo, S., Gervasini, C., Larizza, L., 2014. Overall and allele-specific expression of the SMC1A gene in female Cornelia de Lange syndrome patients and healthy controls. Epigenetics 9, 973–979. doi:10.4161/epi.28903

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., Jensen, T.H., 2008. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. Science (80-.). 322, 1851–1854. doi:10.1126/science.1164096

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680. doi:10.1016/j.cell.2014.11.021

Rao, S.S.P., Huang, S.-C., Hilaire, B.G.S., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., Huang, X., Shamim, M.S., Shin, J., Turner, D., Ye, Z., Omer, A.D., Robinson, J.T., Schlick, T., Bernstein, B.E., Casellas, R., Lander, E.S., Aiden, E.L., 2017. Cohesin Loss Eliminates All Loop Domains. Cell 171, 305–320.e24. doi:10.1016/j.cell.2017.09.026

Revenkova, E., Focarelli, M.L., Susani, L., Paulis, M., Bassi, M.T., Mannini, L., Frattini, A., Delia, D., Krantz, I., Vezzoni, P., Jessberger, R., Musio, A., 2009. Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. Hum. Mol. Genet. 18, 418–427. doi:10.1093/hmg/ddn369

Riedel, C.G., Gregan, J., Gruber, S., Nasmyth, K., 2004. Is chromatin remodeling required to build sister-chromatid cohesion? Trends Biochem. Sci. doi:10.1016/j.tibs.2004.06.007

Rincon-Arano, H., Halow, J., Delrow, J.J., Parkhurst, S.M., Groudine, M., 2012. UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions. Cell 151, 1214–1228. doi:10.1016/j.cell.2012.11.009

Rollins, R.A., Morcillo, P., Dorsett, D., 1999. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. Genetics 152, 577–593.

Rollins, R.A., Korom, M., Aulner, N., Martens, A., Dorsett, D., 2004. Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Mol. Cell. Biol. 24, 3100–11. doi:10.1128/MCB.24.8.3100-3111.2004

Saitoh, N., Goldberg, I.G., Wood, E.R., Earnshaw, W.C., 1994. ScII: An abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. J. Cell Biol. 127, 303–318. doi:10.1083/jcb.127.2.303

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A., Song, Y., El-Fishawy, P., Murtha, R.C., Choi, M., Overton, J.D., Bjornson, R.D., Carriero, N.J., Meyer, K.A., Bilguvar, K., Mane, S.M., Šestan, N., Lifton, R.P., Günel, M., Roeder, K., Geschwind, D.H., Devlin, B., State, M.W., 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237–241. doi:10.1038/nature10945

Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., Odom, D.T., 2010. A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res. 20, 578–588. doi:10.1101/gr.100479.109

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., Spitz, F., 2017. Two independent modes of chromatin organization revealed by cohesin removal. Nature 551, 51–56. doi:10.1038/nature24281

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., Sharp, P.A., 2008. Divergent Transcription from Active Promoters. Science (80-.). 322, 1849–1851. doi:10.1126/science.1162253

Shintomi, K., Hirano, T., 2011. The relative ratio of condensing I to II determines chromosome shapes. Genes Dev. 25, 1464–1469. doi:10.1101/gad.2060311

Simonis, N., Gonze, D., Orsi, C., van Helden, J., Wodak, S.J., 2006. Modularity of the Transcriptional Response of Protein Complexes in Yeast. J. Mol. Biol. 363, 589–610. doi:10.1016/j.jmb.2006.06.024

Singh, R.K., Mandal, T., Balsubramanian, N., Viaene, T., Leedahl, T., Sule, N., Cook, G., Srivastava, D.K., 2011. Histone deacetylase activators: N-acetylthioureas serve as highly potent and isozyme selective activators for human histone deacetylase-8 on a fluorescent substrate. Bioorganic Med. Chem. Lett. 21, 5920–5923. doi:10.1016/j.bmcl.2011.07.080

Sirmaci, A., Spiliopoulos, M., Brancati, F., Powell, E., Duman, D., Abrams, A., Bademci, G., Agolini, E., Guo, S., Konuk, B., Kavaz, A., Blanton, S., Digilio, M.C., Dallapiccola, B., Young, J., Zuchner, S., Tekin, M., 2011. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. Am. J. Hum. Genet. doi:10.1016/j.ajhg.2011.06.007

Slavin, T.P., Lazebnik, N., Clark, D.M., Vengoechea, J., Cohen, L., Kaur, M., Konczal, L., Crowe, C.A., Corteville, J.E., Nowaczyk, M.J., Byrne, J.L., Jackson, L.G., Krantz, I.D., 2012. Germline mosaicism in Cornelia de Lange syndrome. Am. J. Med. Genet. Part A 158 A, 1481–1485. doi:10.1002/ajmg.a.35381

Spielmann, M., Mundlos, S., 2016. Looking beyond the genes: The role of non-coding variants in human disease. Hum. Mol. Genet. doi:10.1093/hmg/ddw205

Tan, K., Shlomi, T., Feizi, H., Ideker, T., Sharan, R., 2007. Transcriptional regulation of protein complexes within and across species. Proc. Natl. Acad. Sci. U. S. A. 104, 1283–1288. doi:10.1073/pnas.0606914104

Teresa-Rodrigo, M.E., Eckhold, J., Puisac, B., Dalski, A., Gil-Rodríguez, M.C., Braunholz, D., Baquero, C., Hernández-Marcos, M., de Karam, J.C., Ciero, M., Santos-Simarro, F., Lapunzina, P., Wierzba, J., Casale, C.H., Ramos, F.J., Gillessen-Kaesbach, G., Kaiser, F.J., Pié, J., 2014. Functional characterization of NIPBL physiological splice variants and eight

splicing mutations in patients with Cornelia de Lange syndrome. Int. J. Mol. Sci. 15, 10350–10364. doi:10.3390/ijms150610350

Terret, M.-E., Sherwood, R., Rahman, S., Qin, J., Jallepalli, P. V, 2009. Cohesin acetylation speeds the replication fork. Nature 462, 231–4. doi:10.1038/nature08550

Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., Raychaudhuri, S., 2013. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat. Genet. 45, 124–130. doi:10.1038/ng.2504

Tonkin, E.T., Wang, T.-J., Lisgo, S., Bamshad, M.J., Strachan, T., 2004. 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. 36, 636–641. doi:10.1038/ng1363

Velmeshev, D., Magistri, M., Faghihi, M.A., 2013. Expression of non-protein-coding antisense RNAs in genomic regions related to autism spectrum disorders. Mol. Autism 4, 32. doi:10.1186/2040-2392-4-32

Watrin, E., Peters, J.-M., 2006 (2006a). Cohesin and DNA damage repair. Exp. Cell Res. 312, 2687–2693. doi:10.1016/j.yexcr.2006.06.024

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K., Peters, J.M., 2006 (2006b). Human Scc4 Is Required for Cohesin Binding to Chromatin, Sister-Chromatid Cohesion, and Mitotic Progression. Curr. Biol. 16, 863–874. doi:10.1016/j.cub.2006.03.049

Watrin, E., Peters, J.-M., 2009. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. EMBO J. 28, 2625–35. doi:10.1038/emboj.2009.202

Watrin, E., Kaiser, F.J., Wendt, K.S., 2016. Gene regulation and chromatin organization: Relevance of cohesin mutations to human disease. Curr. Opin. Genet. Dev. doi:10.1016/j.gde.2015.12.004

Weedon, M.N., Cebola, I., Patch, A.-M., Flanagan, S.E., De Franco, E., Caswell, R., Rodríguez-Seguí, S.A., Shaw-Smith, C., Cho, C.H.-H., Lango Allen, H., Houghton, J.A.L., Roth, C.L., Chen, R., Hussain, K., Marsh, P., Vallier, L., Murray, A., International Pancreatic Agenesis Consortium, Ellard, S., Ferrer, J., Hattersley, A.T., 2014. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat. Genet. 46, 61–4. doi:10.1038/ng.2826

Weichert, J., Schröer, A., Beyer, D.A., Gillessen-Kaesbach, G., Stefanova, I., 2011. Cornelia de Lange syndrome: antenatal diagnosis in two consecutive pregnancies due to rare gonadal mosaicism of NIPBL gene mutation. J. Matern. Fetal. Neonatal Med. 24, 978–82. doi:10.3109/14767058.2010.531312

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto,

N., Maeshima, K., Shirahige, K., Peters, J.-M., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796–801. doi:10.1038/nature06634

Wieczorek, D., Bögershausen, N., Beleggia, F., Steiner-Haldenstätt, S., Pohl, E., Li, Y., Milz, E., Martin, M., Thiele, H., Altmüller, J., Alanay, Y., Kayserili, H., Klein-Hitpass, L., Böhringer, S., Wollstein, A., Albrecht, B., Boduroglu, K., Caliebe, A., Chrzanowska, K., Cogulu, O., Cristofoli, F., Czeschik, J.C., Devriendt, K., Dotti, M.T., Elcioglu, N., Gener, B., Goecke, T.O., Krajewska-Walasek, M., Guillén-Navarro, E., Hayek, J., Houge, G., Kilic, E., Simsek-Kiper, P.Ö., López-González, V., Kuechler, A., Lyonnet, S., Mari, F., Marozza, A., Dramard, M.M., Mikat, B., Morin, G., Morice-Picard, F., Özkinay, F., Rauch, A., Renieri, A., Tinschert, S., Eda Utine, G., Vilain, C., Vivarelli, R., Zweier, C., Nürnberg, P., Rahmann, S., Vermeesch, J., Lüdecke, H.J., Zeschnigk, M., Wollnik, B., 2013. A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Hum. Mol. Genet. 22, 5121–5135. doi:10.1093/hmg/ddt366

Wieczorek, D., Newman, W.G., Wieland, T., Berulava, T., Kaffe, M., Falkenstein, D., Beetz, C., Graf, E., Schwarzmayr, T., Douzgou, S., Clayton-Smith, J., Daly, S.B., Williams, S.G., Bhaskar, S.S., Urquhart, J.E., Anderson, B., O'Sullivan, J., Boute, O., Gundlach, J., Czeschik, J.C., Van Essen, A.J., Hazan, F., Park, S., Hing, A., Kuechler, A., Lohmann, D.R., Ludwig, K.U., Mangold, E., Steenpaß, L., Zeschnigk, M., Lemke, J.R., Lourenco, C.M., Hehr, U., Prott, E.C., Waldenberger, M., Böhmer, A.C., Horsthemke, B., O'Keefe, R.T., Meitinger, T., Burn, J., Lüdecke, H.J., Strom, T.M., 2014. Compound Heterozygosity of Low-Frequency Promoter Deletions and Rare Loss-of-Function Mutations in TXNL4A Causes Burn-McKeown Syndrome. Am. J. Hum. Genet. 95, 698–707. doi:10.1016/j.ajhg.2014.10.014

Woods, S.A., Robinson, H.B., Kohler, L.J., Agamanolis, D., Sterbenz, G., Khalifa, M., 2014. Exome sequencing identifies a novel EP300 frame shift mutation in a patient with features that overlap cornelia de lange syndrome. Am. J. Med. Genet. Part A 164, 251–258. doi:10.1002/ajmg.a.36237

Woodward, J., Taylor, G.C., Soares, D.C., Boyle, S., Sie, D., Read, D., Chathoth, K., Vukovic, M., Tarrats, N., Jamieson, D., Campbell, K.J., Blyth, K., Acosta, J.C., Ylstra, B., Arends, M.J., Kranc, K.R., Jackson, A.P., Bickmore, W.A., Wood, A.J., 2016. Condensin II mutation causes T-cell lymphoma through tissue-specific genome instability. Genes Dev. 30, 2173–2186. doi:10.1101/gad.284562.116

Wu, N., Ming, X., Xiao, J., Wu, Z., Chen, X., Shinawi, M., Shen, Y., Yu, G., Liu, J., Xie, H., Gucev, Z.S., Liu, S., Yang, N., Al-Kateb, H., Chen, J., Zhang, J., Hauser, N., Zhang, T., Tasic, V., Liu, P., Su, X., Pan, X., Liu, C., Wang, L., Shen, J., Shen, J., Chen, Y., Zhang, T., Zhang, J., Choy, K.W., Wang, J., Wang, Q., Li, S., Zhou, W., Guo, J., Wang, Y., Zhang, C., Zhao, H., An, Y., Zhao, Y., Wang, J., Liu, Z., Zuo, Y., Tian, Y., Weng, X., Sutton, V.R., Wang, H., Ming, Y., Kulkarni, S., Zhong, T.P., Giampietro, P.F., Dunwoodie, S.L., Cheung, S.W., Zhang, X., Jin, L., Lupski, J.R., Qiu, G., Zhang, F., 2015. *TBX6* Null Variants and a Common Hypomorphic Allele in Congenital Scoliosis. N. Engl. J. Med. 372, 341–350. doi:10.1056/NEJMoa1406829

Yan, J., Saifi, G.M., Wierzba, T.H., Withers, M., Bien-Willner, G.A., Limon, J., Stankiewicz, P., Lupski, J.R., Wierzba, J., 2006. Mutational and genotype-phenotype correlation analyses in 28 Polish patients with Cornelia de Lange syndrome. Am. J. Med. Genet. Part A 140, 1531–1541. doi:10.1002/ajmg.a.31305

Yazdi, P.T., Wang, Y., Zhao, S., Patel, N., Lee, E.Y.H.P., Qin, J., 2002. SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. Genes Dev. 16, 571–582. doi:10.1101/gad.970702

Yuan, B., Pehlivan, D., Karaca, E., Patel, N., Charng, W.L., Gambin, T., Gonzaga-Jauregui, C., Sutton, V.R., Yesil, G., Bozdogan, S.T., Tos, T., Koparir, A., Koparir, E., Beck, C.R., Gu, S., Aslan, H., Yuregir, O.O., Rubeaan, K. Al, Alnaqeb, D., Alshammari, M.J., Bayram, Y., Atik, M.M., Aydin, H., Geckinli, B.B., Seven, M., Ulucan, H., Fenercioglu, E., Ozen, M., Jhangiani, S., Muzny, D.M., Boerwinkle, E., Tuysuz, B., Alkuraya, F.S., Gibbs, R.A., Lupski, J.R., 2015. Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. J. Clin. Invest. 125, 636–651. doi:10.1172/JCI77435

Zhang, C., Kuang, M., Li, M., Feng, L., Zhang, K., Cheng, S., 2016. SMC4, which is essentially involved in lung development, is associated with lung adenocarcinoma progression. Sci. Rep. 6, 34508. doi:10.1038/srep34508

Zhang, F., Lupski, J.R., 2015. Non-coding genetic variants in human disease. Hum. Mol. Genet. doi:10.1093/hmg/ddv259

Zhang, J., Shi, X., Li, Y., Kim, B.J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y., Zhang, P., Kim, S.T., Pan, X., Qin, J., 2008. Acetylation of Smc3 by Eco1 Is Required for S Phase Sister Chromatid Cohesion in Both Human and Yeast. Mol. Cell 31, 143–151. doi:10.1016/j.molcel.2008.06.006

Zhang, N., Kuznetsov, S.G., Sharan, S.K., Li, K., Rao, P.H., Pati, D., 2008. A handcuff model for the cohesin complex. J. Cell Biol. 183, 1019–1031. doi:10.1083/jcb.200801157

Zuin, J., Dixon, J.R., van der Reijden, M.I.J.A., Ye, Z., Kolovos, P., Brouwer, R.W.W., van de Corput, M.P.C., van de Werken, H.J.G., Knoch, T.A., van IJcken, W.F.J., Grosveld, F.G., Ren, B., Wendt, K.S., 2014 (2014a). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci USA 111, 996–1001. doi:10.1073/pnas.1317788111

Zuin, J., Franke, V., van IJcken, W.F.J., van der Sloot, A., Krantz, I.D., van der Reijden, M.I.J.A., Nakato, R., Lenhard, B., Wendt, K.S., 2014 (2014b). A Cohesin-Independent Role for NIPBL at Promoters Provides Insights in CdLS. PLoS Genet. 10. doi:10.1371/journal.pgen.1004153

Abbreviations

A-Adenine

ASO- antisense oligonucleotides

ATP- adenosine triphosphate

bp- base pair

BSA- bovine serum albumin

C- Cytosine

CdLS- Cornelia de Lange Syndrome

cDNA- complementary DNA

CGH- Comparative Genomic Hybridization

CHX- cycloheximide

CNVs- copy number variations

CRISPR- Clustered Regularly Interspaced Short Palindromic Repeats

CSS- Coffin-Siris Syndrome

CTCF- CCCTC-binding factor

DMEM- Dulbecco´s Modified Eagle´s Medium

DMSO- dimethyl sulfoxide

DNA- deoxyribonucleic acid

dNTP- deoxyribonucleotide triphosphate

ddNTPs- dideoxynucleotide triphosphate

EDTA- Ethylenediaminetetraacetic acid

ENCODE- Encyclopedia of DNA Elements

FBS- fetal bovine serum

G- Guanine

GERD- gastroesophageal reflux disease

gRNA- guide RNA

GWAS- genome-wide association study

h- hour

HDAC- histone deacetylase

HRP- horseradish peroxidase

ICS- intrinsic chromosome structure

kb- kilobase

kDa- kilodalton

l- liter

LB- lysogeny broth

LCLs- lymphoblastoid cell lines

lncRNA- long non-coding RNA

LoF- loss of function

Mb- megabase

min- minute

mg- miligram

ml- milliliter

mRNA- messenger RNA

µg- microgram

µl- microliter

µM- micromolar

NGS- Next Generation Sequencing

NIPBL- Nipped-B-like

NMD- nonsense-mediated decay

PAGE- polyacrylamide gel electrophoresis

PBS- Phosphate buffered saline

PCR- polymerase chain reaction

RNA- ribonucleic acid

rpm- rounds per minute

SCC- sister chromatin cohesion

SDS- sodium dodecyl sulphate

siRNA- small interfering RNA

SMC- structural maintenance of chromosomes

SNP- single nucleotide polymorphism

SNVs- Single nucleotide variations

T- Thymine

TAD- topologically associated domain

Taq- Thermus aquaticus

TBE- tris borat EDTA

TEMED- *Tetramethylethylenediamine*

TSS- transcription start site

UCSC- University of California, Santa Cruz

V- volt

3C- chromosome conformation capturing

Gene, exon and assay	Forward primer (5`-3`)	Reverse primer (5`-3`)	Application
NIPBL ex18	CTTTATCTTCCAGGTTCTGTAGC	ATTATTGAGTGAGCTAGGTTATAC	Patient 1
NIPBL ex20	CAGCTTACCTTAGATACTGAAAAC	CAGACAGAAATGAAGAATAAATGTATGACGG	Patient 4
NIPBL ex23	GGTAGACAGATGACTGACATGTG	GAATTCATCATCTGTGAATCAGAG	Patient 2
NIPBL ex26-30 cDNA	CAGTGGTTTCGAGACACAAC	CATCATTGACTCTGCGAATC	Patient 3
NIPBL ex29	CTTGATATGCAACGAGGTG	CAATATTCTTTCAATCAGCATATC	Patient 5
NIPBL ex29 pyro	GATGGATAATTCGACTAGTG		Patient 5
SMC1A ex20 cDNA	GGAAGAGGGTAGCTCCCAG	GCTGGCTTGTAGCTGTGGAT	Case 10
HDAC8 ex8	TACCTCCAACTTGCCCTAC	TCCAATGCAAGAAAGCCAAG	Patient 6/10
HDAC8 ex8 SNaPshot	CATGTGCTCCTTTAACATGA		Patient 6/10
HDAC8 ex8-9 cDNA	GGCATCATGCAAAGAAAGATG	TAGACCACATGCTTCAGATTC	Patient 8
ANKRD11 ex9	AGGACCTGGAGATCGAGGAG	GCGGTAAAGGTTTGTGGAGA	Patient A
NIPBL Promoter F2	GCTAGCTCCAGCCTTGTCCAGCGCCTCAG	CTCGAGGTGTGTGTCTCTCTCTCGTTCCGTC	cloning
NIPBL Enhancer R1-1	CAAATGTGGTAAAATCGATAAggatccGGCTGCGGATACACCTTG	CTCAAGGGCATCGGTCGACggatccAAGGGCAGATTCTGCCTC	cloning-Gibson
NIPBL Enhancer R1-2	GGATCCTGGCGTTTCACCATGTTGTC	GTCGACGTGGAATGGAAGGTTTGTGTAG	cloning
NIPBL Enh R1 seq	CCTGCTATTCTTCAGGTGC	GCACCGTAGTCTCACACC	cohort of patients
NIPBL Enh R2 seq	GAGGATCATATGTTTAGAAGGGTC	CACTCGGGCAGCCCTGCG	cohort of patients
NIPBL Enh R3 seq	CAGTTCCAGCCGCGCTGCAG	GTCGGAGCGCAGCCAGGTGAG	cohort of patients
NIPBL Enh R4 seq	CTGCAGTGGAGTCGGGAGC	CTAAGGTAGTGTGTTGACACGATG	cohort of patients
NIPBL ex22 cDNA	CTTCAGGTTAAACAGTAGTG	CTTACTACCACATTTTTTAAGG	Patient 1
NIPBL ex22 cDNA pyro	CTCTTATGAAACAGCTATG		Patient 1
NIPBL ex10	CACAGGCATGACAATAGG	CTACTAGAGGATCTCAATG	Patient 3
NIPBL ex10 pyro	GTTTCTGAGTCACTAAGACG		Patient 3
chr9del breakpoint PCR	CTACATTGTTGACAATGGG	CAATGTCATACTGAATGGAG	Patient 1
SMC2 Promoter	gagctcCGAGAAATATTTTCAAAATATAAGCCCATTCC	ctcgagCGAGGCAAAAACAGCAACGATAG	cloning
SMC2 Enh full length	ggatccGCTGCAGCTACACTGTGAGACAGGCG	gtcgacGACTTTAGGTAGCCTAATGACTATATGCCTTGG	cloning
SMC2 Enh F1	ggatccGCTGCAGCTACACTGTGAGACAGGCG	gtcgacGTGAAGTTGCCACACTCCAGGC	cloning
SMC2 Enh F2	ggatccCTGCCATTCACCCTGGTAG	gtcgacGATTGGAATGATAGAGGTTTCTTG	cloning
SMC2 Enh F3	ggatccCTGACATGGCAGGGGAGCTG	gtcgacGACTTTAGGTAGCCTAATGACTATATGCCTTGG	cloning
CRISPR chr9del gRNA1	CGACGTAGCACCGGTGAGTCCCAAGTGTGAGAGTTTTAGAGCTAGAAATAGCAAGTTAAA		CRISPR/Cas9
CRISPR chr9del gRNA5	CGACGTAGCACCGGTATATATCTAGCTCGGCAGTTTTAGAGCTAGAAATAGCAAGTTAAA		CRISPR/Cas9
CRISPR_verification	CAAGAAGCAATGCAGGAAC	CTACTCTTAGAATTTGCTTC	CRISPR/Cas9

ex-exon; pyro-pyrosequencing; F1/F2/F3-fragment 1/2/3; R-region; Enh-enhancer

Acknowledgements

My endless gratitude goes to my supervisor Prof. Dr. Frank Kaiser for allowing me to start PhD studies in his lab, for guiding me and trusting in me. I have learned so many new things about DNA, RNA, proteins and myself during this period.

Next, many thanks to all the girls (and some guys) in the lab, who patiently showed me new methods at my beginnings and made the work fun, contributing to a great atmosphere in our lab. I am especially grateful to Dr. Diana Braunholz and Juliane Eckhold for their kindness and care, but also to Dr. Ilaria Parenti, Sara Ruiz Gil, Ronja Hollstein and Barbara Geidner. Many thanks to all my Germans from the lab (Frank, Juliane and Martin Grosse) for help with the Zusammenfassung.

I am also very grateful to Prof. Dr. Gabriele Gillessen-Kaesbach, former director of the Institute and current director of the University, for her great support, enormous clinical knowledge and for maintaining a great atmosphere at the Institute.

Many thanks to all the colleagues from the Institute, who were always willing to help with the diagnostic-related problems and questions I had.

I am incredibly grateful to our collaborators Dr. Erwan Watrin and Dr. Kerstin Wendt. Besides excellent discussions, tips and tricks, I also had an opportunity to visit Erwan´s lab and see how he does those beautiful Western blots. I have learned many lab tricks there and will always have a great admiration for him as a scientist and a person.

Speaking of great persons, Dr. Ana Westenberger, Dr. Valerija Dobričić and Dr. Marija Dulović need to be mentioned here. Ana has been helping me ever since coming to Germany was even only a thought, Val showed me how to do my ever first PCR and Marija has the title of my ever first real friend here in Germany, helping also with the statistics, cells etc.

I also have to thank Dr. Tillman Vollbrandt a thousand times, for all the cell cycle experiments and his patience, as well as Dr. Aleksandar Raković for showing me how to do CRISPR/Cas9. I am also very grateful to Dr. Ralf Werner and Dagmar Struve, for help with the pyrosequencing experiments.

Many thanks to Sander for the ´hakuna matata´ attitude, for his tolerance during my writing process and for many lovely moments.

The final but the biggest gratitude goes to my family, for their pure and everlasting love and support. There are no words to explain my love and gratitude and the least I can do is devoting this thesis to them!

Jelena Pozojević

Date of birth: 30.12.1986
Place of birth: Zrenjanin, Serbia
Nationality: Serbian

Education:

PhD in Human Genetics - Thesis in progress
January 2014 – January 2018
Institute of Human Genetics
University of Lübeck, Germany
Mentor: Prof. Dr. Frank Kaiser

MSc in Pharmacy - Medical Biochemistry
March 2012
Faculty of Pharmacy, University of Belgrade, Serbia
Thesis Title: Immunotherapy for allergies and asthma
Mentor: Prof. Dr. Nevena Arsenović Ranin

High school (gymnasium) diploma
June 2005
Zrenjanin Grammar School, Zrenjanin, Serbia
Graduation work in chemistry: ´Carbonyl compounds aldehydes and ketones´
Mentor: Prof. Marija Medarević

Research experience:

Section for Functional Genetics at the Institute of Human Genetics, Lübeck
January 2014-Present
PhD student
Mentor: Prof. Dr. Frank Kaiser

Institute of Genetics and Development of Rennes (IGDR), Rennes, France
May and June 2015
Internship in Cell Biology
Mentor: Dr. Erwan Watrin

Molecular Genetics Laboratory
Institute of Neurology, Clinical Center of Serbia, Belgrade, Serbia
June 2013 – October 2013
Practice as an early stage researcher
Mentor: Dr. Valerija Dobričić

IPSF Student Exchange Program
Department of Biopharmaceutics and Pharmacokinetics
Faculty of Pharmacy, University of Ljubljana, Ljubljana, Slovenia
July and August 2011
Participated in pharmacokinetic studies of drugs
Mentor: Prof. Dr. Albin Kristl

Awards: **Young Investigator Award for Outstanding Science**
at the European Conference of Human Genetics (ESHG) 2016 in Barcelona

German Society of Human Genetics (Gfh) travel grant
for attending ESHG Conference 2016 in Barcelona

Conferences: **European Conference of Human Genetics (ESHG)**
2016 in Barcelona
Talk: ´Loss of a non-coding regulatory element on chromosome 9 in a family with CdLS-like phenotype´

European Conference of Human Genetics (ESHG)
2017 in Copenhagen
Co-chaired the session: ´Intellectual Disability´

German Society of Human Genetics (GfH) conferences
2014 in Essen
2016 in Lübeck
2017 in Bochum
Talk: ´Mutations affecting coding or regulatory regions of *SMC2* cause dysregulation of condensins resulting in a phenotype reminiscent of cohesinopathies´

Treffen der Norddeutschen Humangenetiker
2016 in Lübeck
Talk: ´Loss of a non-coding regulatory element on chromosome 9 in a family with CdLS-like phenotype´

Publications:

Zuin, J., Casa, V., **Pozojevic, J.**, Kolovos, P., van den Hout, M.C.G.N., van Ijcken, W.F.J., Parenti, I., Braunholz, D., Baron, Y., Watrin, E., Kaiser, F.J., Wendt, K.S., 2017. Regulation of the cohesin-loading factor NIPBL: Role of the lncRNA NIPBL-AS1 and identification of a distal enhancer element. PLoS Genet. 13

Pozojevic J, Parenti I, Graul-Neumann L, Ruiz Gil S, Watrin E, Wendt KS, Werner R, Strom TM, Gillessen-Kaesbach G, Kaiser FJ, 2017. Novel mosaic variants in two patients with Cornela de Lange syndrome. Eur J Med Genet. [Epub ahead of print]

Symonds JD, Joss S, Metcalfe KA, Somarathi S, Cruden J, Devlin AM, Donaldson A, DiDonato N, Fitzpatrick D, Kaiser FJ, Lampe AK, Lees MM, McLellan A, Montgomery T, Mundada V, Nairn L, Sarkar A, Schallner J, **Pozojevic J**, Parenti I, Tan J, Turnpenny P, Whitehouse WP, Zuberi SM, 2017. Heterozygous truncation mutations of the SMC1A gene cause a severe early onset epilepsy with cluster seizures in females: Detailed phenotyping of 10 new cases. Epilepsia 58, 565–575

Parenti, I., Teresa-Rodrigo, M.E., **Pozojevic, J.**, Ruiz Gil, S., Bader, I., Braunholz, D., Bramswig, N.C., Gervasini, C., Larizza, L., Pfeiffer, L., Ozkinay, F., Ramos, F., Reiz, B., Rittinger, O., Strom, T.M.,

Watrin, E., Wendt, K., Wieczorek, D., Wollnik, B., Baquero-Montoya, C., Pié, J., Deardorff, M.A., Gillessen-Kaesbach, G., Kaiser, F.J., 2017. Mutations in chromatin regulators functionally link Cornelia de Lange syndrome and clinically overlapping phenotypes. Hum. Genet. 136, 307–320

Parenti, I., Gervasini, C., **Pozojevic, J**., Wendt, K.S., Watrin, E., Azzollini, J., Braunholz, D., Buiting, K., Cereda, A., Engels, H., Garavelli, L., Glazar, R., Graffmann, B., Larizza, L., Lüdecke, H.J., Mariani, M., Masciadri, M., Pié, J., Ramos, F.J., Russo, S., Selicorni, A., Stefanova, M., Strom, T.M., Werner, R., Wierzba, J., Zampino, G., Gillessen-Kaesbach, G., Wieczorek, D., Kaiser, F.J., 2016. Expanding the clinical spectrum of the "HDAC8-phenotype" - implications for molecular diagnostics, counseling and risk prediction. Clin. Genet. 89, 564–573

Teresa-Rodrigo ME, Eckhold J, Puisac B, **Pozojevic J**, Parenti I, Baquero-Montoya C, Gil-Rodríguez MC, Braunholz D, Dalski A, Hernández-Marcos M, Ayerza A, Bernal ML, Ramos FJ, Wieczorek D, Gillessen-Kaesbach G, Pié J, Kaiser FJ, 2016. Identification and functional characterization of two intronic NIPBL mutations in two patients with Cornelia de Lange syndrome. Int. J. Mol. Sci. 15, 10350–10364.

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

Bramswig, N.C., Lüdecke, H.J., Alanay, Y., Albrecht, B., Barthelmie, A., Boduroglu, K., Braunholz, D., Caliebe, A., Chrzanowska, K.H., Czeschik, J.C., Endele, S., Graf, E., Guillén-Navarro, E., Kiper, P.Ö.S., López-González, V., Parenti, I., **Pozojevic, J.**, Utine, G.E., Wieland, T., Kaiser, F.J., Wollnik, B., Strom, T.M., Wieczorek, D., 2015. Exome sequencing unravels unexpected differential diagnoses in individuals with the tentative diagnosis of Coffin–Siris and Nicolaides–Baraitser syndromes. Hum. Genet. 134, 553–568

Gil-Rodríguez, M.C., Deardorff, M.A., Ansari, M., Tan, C.A., Parenti, I., Baquero-Montoya, C., Ousager, L.B., Puisac, B., Hernández-Marcos, M., Teresa-Rodrigo, M.E., Marcos-Alcalde, I., Wesselink, J.J., Lusa-Bernal, S., Bijlsma, E.K., Braunholz, D., Bueno-Martinez, I., Clark, D., Cooper, N.S., Curry, C.J., Fisher, R., Fryer, A., Ganesh, J., Gervasini, C., Gillessen-Kaesbach, G., Guo, Y., Hakonarson, H., Hopkin, R.J., Kaur, M., Keating, B.J., Kibaek, M., Kinning, E., Kleefstra, T., Kline, A.D., Kuchinskaya, E., Larizza, L., Li, Y.R., Liu, X., Mariani, M., Picker, J.D., Pié, Á., **Pozojevic, J.**, Queralt, E., Richer, J., Roeder, E., Sinha, A., Scott, R.H., So, J., Wusik, K.A., Wilson, L., Zhang, J., Gómez-Puertas, P., Casale, C.H., Ström, L., Selicorni, A., Ramos, F.J., Jackson, L.G., Krantz, I.D., Das, S., Hennekam, R.C.M., Kaiser, F.J., Fitzpatrick, D.R., Pié, J., 2015. De novo heterozygous mutations in SMC3 cause a range of Cornelia de Lange syndrome-overlapping phenotypes. Hum. Mutat. 36, 454–462

Braunholz, D., Obieglo, C., Parenti, I., **Pozojevic, J.**, Eckhold, J., Reiz, B., Brænne, I., Wendt, K.S., Watrin, E., Vodopiutz, J., Rieder, H., Gillessen-Kaesbach, G., Kaiser, F.J., 2015. Hidden mutations in Cornelia de Lange syndrome limitations of Sanger sequencing in molecular diagnostics. Hum. Mutat. 36, 26–29