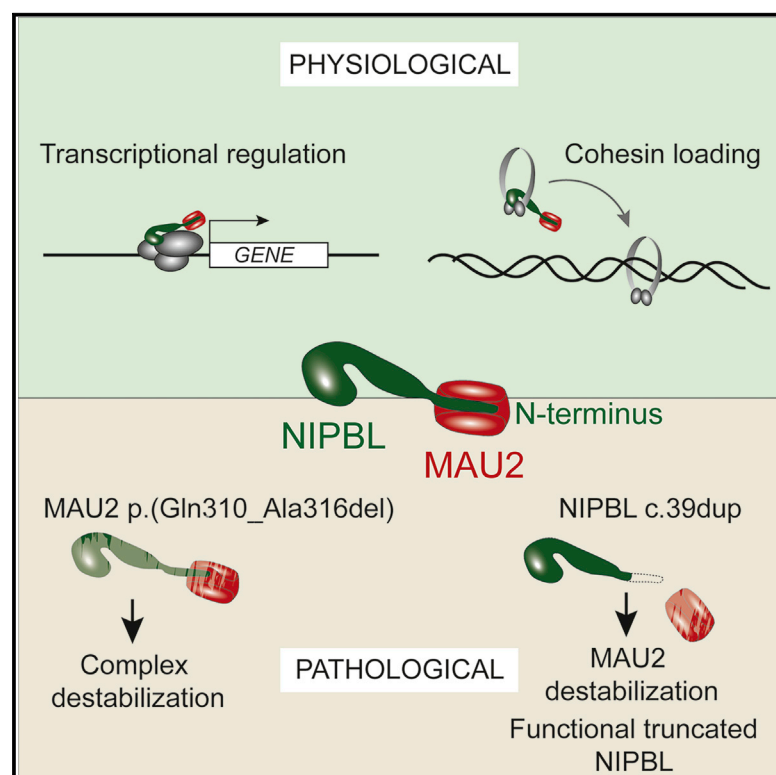


# MAU2 and NIPBL Variants Impair the Heterodimerization of the Cohesin Loader Subunits and Cause Cornelia de Lange Syndrome

## Graphical Abstract



## Authors

Ilaria Parenti, Farah Diab, Sara Ruiz Gil, ..., Erwan Watrin, Frank J. Kaiser, Kerstin S. Wendt

## Correspondence

frank.kaiser@uk-essen.de (F.J.K.), k.wendt@erasmusmc.nl (K.S.W.)

## In Brief

Parenti et al. describe mutations causative for CdLS that affect the interaction between NIPBL and MAU2, the two subunits of the cohesin loader complex. This study further reveals a protective mechanism that rescues NIPBL expression in the presence of early truncating variants by using alternative translation initiation site.

## Highlights

- The NIPBL-MAU2 interaction is required for the stability of both proteins
- A pathogenic MAU2 deletion impairs dimerization and causes NIPBL haploinsufficiency
- Early frameshift mutations in NIPBL induce alternative translation initiation
- Translation re-initiation rescues NIPBL expression and renders MAU2 dispensable



## Article

# MAU2 and NIPBL Variants Impair the Heterodimerization of the Cohesin Loader Subunits and Cause Cornelia de Lange Syndrome

Ilaria Parenti,<sup>1,2</sup> Farah Diab,<sup>3,4</sup> Sara Ruiz Gil,<sup>1</sup> Eskeatnaf Mulugeta,<sup>5</sup> Valentina Casa,<sup>5</sup> Riccardo Berutti,<sup>6</sup> Rutger W.W. Brouwer,<sup>7</sup> Valerie Dupé,<sup>3,4</sup> Juliane Eckhold,<sup>1,8</sup> Elisabeth Graf,<sup>6</sup> Beatriz Puisac,<sup>9</sup> Feliciano Ramos,<sup>9</sup> Thomas Schwarzmayr,<sup>6</sup> Macarena Moronta Gines,<sup>5</sup> Thomas van Staveren,<sup>5</sup> Wilfred F.J. van IJcken,<sup>7</sup> Tim M. Strom,<sup>6,10</sup> Juan Pié,<sup>9</sup> Erwan Watrin,<sup>3,4,12</sup> Frank J. Kaiser,<sup>1,8,11,12,\*</sup> and Kerstin S. Wendt<sup>5,12,13,\*</sup>

<sup>1</sup>Sektion für Funktionelle Genetik am Institut für Humangenetik Lübeck, Universität zu Lübeck, Lübeck, Germany

<sup>2</sup>Institute of Science and Technology (IST) Austria, Klosterneuburg, Austria

<sup>3</sup>Centre National de la Recherche Scientifique, UMR6290, Rennes, France

<sup>4</sup>Institut de Génétique et Développement de Rennes, Université de Rennes, Rennes, France

<sup>5</sup>Department of Cell Biology, Erasmus MC, Rotterdam, the Netherlands

<sup>6</sup>Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany

<sup>7</sup>Erasmus MC, University Medical Center Rotterdam, Department of Cell Biology, Center for Biomimics, the Netherlands

<sup>8</sup>Institut für Humangenetik, Universitätsklinikum Essen, Universität Duisburg-Essen, Essen, Germany

<sup>9</sup>Unit of Clinical Genetics and Functional Genomics, Department of Pharmacology-Physiology and Paediatrics, School of Medicine, University of Zaragoza, CIBERER-GCV02 and ISS-Aragon, 50009 Zaragoza, Spain

<sup>10</sup>Institute of Human Genetics, Technische Universität München, Munich, Germany

<sup>11</sup>DZHK e.V. (German Center for Cardiovascular Research), Partner Site Hamburg/Kiel/Lübeck, Lübeck, Germany

<sup>12</sup>These authors contributed equally

<sup>13</sup>Lead Contact

\*Correspondence: [frank.kaiser@uk-essen.de](mailto:frank.kaiser@uk-essen.de) (F.J.K.), [k.wendt@erasmusmc.nl](mailto:k.wendt@erasmusmc.nl) (K.S.W.)

<https://doi.org/10.1016/j.celrep.2020.107647>

## SUMMARY

The NIPBL/MAU2 heterodimer loads cohesin onto chromatin. Mutations in *NIPBL* account for most cases of the rare developmental disorder Cornelia de Lange syndrome (CdLS). Here we report a *MAU2* variant causing CdLS, a deletion of seven amino acids that impairs the interaction between MAU2 and the NIPBL N terminus. Investigating this interaction, we discovered that MAU2 and the NIPBL N terminus are largely dispensable for normal cohesin and NIPBL function in cells with a NIPBL early truncating mutation. Despite a predicted fatal outcome of an out-of-frame single nucleotide duplication in *NIPBL*, engineered in two different cell lines, alternative translation initiation yields a form of NIPBL missing N-terminal residues. This form cannot interact with MAU2, but binds DNA and mediates cohesin loading. Altogether, our work reveals that cohesin loading can occur independently of functional NIPBL/MAU2 complexes and highlights a novel mechanism protective against out-of-frame mutations that is potentially relevant for other genetic conditions.

## INTRODUCTION

Cohesin is a highly conserved protein complex essential for cell survival. In humans, the complex is composed of three structural core subunits named SMC1A, SMC3, and RAD21, which together form a ring-shaped structure that topologically encircles DNA (Haering et al., 2008; Murayama and Uhlmann, 2014). This ability allows cohesin to carry out a large spectrum of chromatin-related functions, including sister chromatid cohesion, DNA repair, transcriptional regulation, and three-dimensional organization of chromatin (Zhu and Wang, 2019). In order to accomplish these essential tasks, cohesin needs to interact with chromatin. Cohesin's binding to DNA depends on a heterodimer formed by NIPBL and MAU2, also known as cohesin loader or kollerin complex (Ciosk et al., 2000; Watrin et al.,

2006). Importantly, NIPBL is also known to play a role in transcriptional regulation independently of cohesin (Zuin et al., 2014). The interaction with cohesin is mediated by the C terminus of NIPBL (Chao et al., 2015), and mutations affecting this domain result in reduced interaction between NIPBL and RAD21 (Kikuchi et al., 2016). The interaction between NIPBL and MAU2 is mediated by their respective N termini; precisely, the first 38 amino acids of NIPBL interact with amino acids 32–71 of MAU2 (Braunholz et al., 2012; Chao et al., 2015). Recent evidence suggested that MAU2 is required for the correct folding of the N terminus of NIPBL and that NIPBL is unstable in the absence of MAU2 (Chao et al., 2015; Hinshaw et al., 2015). RNA interference experiments similarly demonstrated that depletion of NIPBL greatly reduces the cellular levels of MAU2 (Watrin et al., 2006), consistent with the notion that the



physical association of these two proteins is required for their stability.

Variations in the *NIPBL* gene are the most frequent cause of Cornelia de Lange syndrome (CdLS; Online Mendelian Inheritance in Man [OMIM] #122470), a multisystem developmental disorder characterized by growth retardation, intellectual disability, developmental delay, limb anomalies, and distinctive facial features (Kline et al., 2007). CdLS was recently grouped into the category of the so-called transcriptopathies, as a global dysregulation of gene expression was detected in cells of CdLS patients (Yuan et al., 2015). *NIPBL* haploinsufficiency is reported in approximately 70% of patients (Borck et al., 2004; Gillis et al., 2004; Selicorni et al., 2007), while up to 15% of cases are collectively accounted for by mutations in cohesin genes (*SMC1A*, *SMC3*, *RAD21*, and *HDAC8*) and additional transcriptional regulators or chromatin remodelers (Deardorff et al., 2012; Gervasini et al., 2013; Gil-Rodríguez et al., 2015; Kaiser et al., 2014; Parenti et al., 2016a, 2016b, 2017; Yuan et al., 2015). How the mutations in *NIPBL* or the other CdLS genes lead to the observed phenotypes is not fully understood, but it is clear that misregulated gene expression plays an important role (Liu et al., 2009; Mills et al., 2018). Interestingly, mutations in *MAU2* have never been reported so far, although the protein encoded by this gene is essential for the stability of *NIPBL*.

Here we describe the first pathogenic variant affecting *MAU2*. The mutation, an in-frame deletion resulting in the loss of seven amino acids, leads to impaired heterodimerization of *MAU2* with *NIPBL*, hence affecting the stability of both proteins. The clinical features of the patient with the deletion in *MAU2* overlap with those of patients with CdLS carrying mutations in *NIPBL*, suggesting a common pathogenic mechanism. To expand upon this observation, we delved into the relationship between *NIPBL* and *MAU2*, with particular attention on the respective interaction domains.

Truncating mutations in *NIPBL* are mostly associated with a more severe phenotype and with a higher frequency of limb reductions or malformations in comparison with missense substitutions (Selicorni et al., 2007). In apparent contradiction to this observation, some patients with truncating mutations affecting the N terminus of *NIPBL*, its *MAU2*-interacting region, present with mild to moderate phenotypes and do not always display dramatic reduction of the upper limbs (Borck et al., 2004; Nizon et al., 2016; Oliveira et al., 2010). This observation raises the interesting possibility that the molecular consequences of *NIPBL*-truncating mutations might depend on their position along the protein sequence. To test this hypothesis, we used genome editing to generate a homozygous early truncating mutation in *NIPBL* in two different cell lines. By this, we observed the expression of an N-terminally truncated *NIPBL* by the use of an alternative translation start site. This N-terminally truncated *NIPBL* is able to mediate cohesin loading and does not depend on *MAU2* for its stability. Actually, *MAU2* levels are dramatically reduced in the edited cells. Analogously, a dramatic reduction of *MAU2* but not of *NIPBL* was observed in the cell line of a CdLS patient with a *NIPBL* splice site variant that should lead to an early truncation of *NIPBL*. Altogether, our data point to the existence of a protective mechanism that dodges the deleterious effects of certain *NIPBL* mutations and might contribute to the

phenotypical variability associated with different *NIPBL*-truncating variants.

## RESULTS

### Whole-Genome Sequencing Identifies the First *MAU2* Mutation

Trio whole-genome sequencing was performed on a group of 15 patients with clinical diagnosis of CdLS and negative for mutations in the five known CdLS genes. One putative pathogenic variant in *MAU2* was identified in a patient with a severe phenotype and typical CdLS facial dysmorphism (for a detailed clinical description, see Human Subjects in STAR Methods).

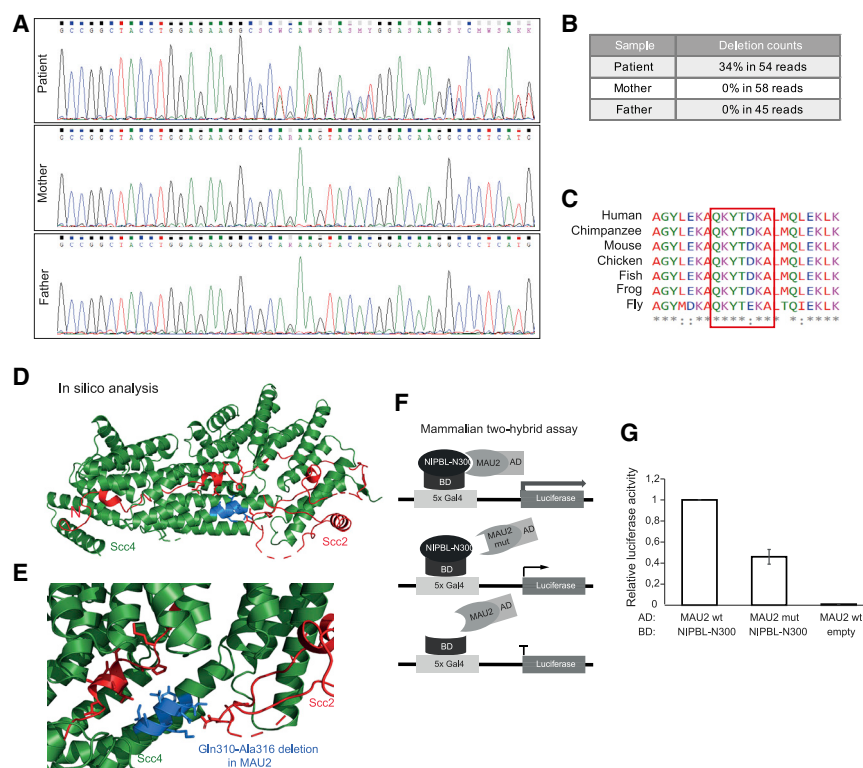
The mutation, an in-frame deletion of 21 nucleotides in *MAU2*, results in the loss of seven amino acids: RefSeq NM\_015329, c.927\_947del, p.(Gln310\_Ala316del). The variant was found in 34% of the sequencing reads, and Sanger sequencing confirmed its *de novo* origin (Figures 1A and 1B). Comparison of the *MAU2* protein sequence across seven species showed that the deletion affects a region that is highly conserved in vertebrate and non-vertebrate animal species (Figure 1C). To date, *de novo* missense and nonsense mutations in *MAU2* are significantly underrepresented in exome sequences (Hinshaw et al., 2015). *MAU2* is categorized among the haploinsufficient genes, with a pLI (probability of intolerance) score of 1, suggesting that the gene is extremely intolerant to loss-of-function variations (Lek et al., 2016). Thus, the mutation identified in the patient of our cohort might represent the first pathogenic variant in *MAU2*.

### The *MAU2* Deletion Impairs the Heterodimerization with *NIPBL*

To investigate how the identified deletion could affect *MAU2* protein structure and its ability to bind *NIPBL*, we performed an *in silico* analysis of the co-crystal structure of the *S. cerevisiae* cohesin loader complex Scc2/*NIPBL*-Scc4/*MAU2* (4XDN) (Hinshaw et al., 2015).

On the basis of a previous comparison of the *MAU2*/Scc4 protein sequence from yeast to human (Watrin et al., 2006), we mapped the deleted region on the Scc2/*NIPBL*-Scc4/*MAU2* crystal structure from *S. cerevisiae* (Figure 1D). The deletion affects highly conserved amino acids in a helix that is part of the TPR array that envelops the N terminus of *NIPBL*. The deletion removes part of a helix that contacts directly the Scc2/*NIPBL* N terminus, likely leading to a distortion of the *MAU2*-*NIPBL* interaction (Figure 1E).

To test this prediction, we introduced the mutation into expression plasmids to perform quantitative mammalian two-hybrid interaction assays in HEK293 cells. Constructs encoding the N-terminal 300 amino acids of *NIPBL* coupled to the DNA binding domain of Gal4 were tested for interaction with the wild-type or mutant full-length *MAU2*, bound to the activation domain of NF-κB (Figure 1F). The seven amino acid deletion in *MAU2* caused a reduction of the heterodimerization activity to 46% (Figure 1G). The diminishing effect of the deletion in *MAU2* on its interaction with *NIPBL* was further confirmed with a yeast two-hybrid assay. Yeast transformed with mutant *MAU2* showed a slower growth on selective medium, indicative



**Figure 1. Characterization of the MAU2 Deletion Identified in a Patient with CdLS**

(A and B) Trio Sanger sequencing (A) and counts of the genome sequencing run (B) show the *de novo* origin of the MAU2 p.(Gln310\_Ala316del) mutation.

(C) Alignment of MAU2 orthologs from different species shows conservation of the deleted amino acids (red square) from human to fly.

(D and E) *In silico* analysis of the cohesin loader co-crystal structure from *S. cerevisiae* (4XDN) (Hinshaw et al., 2015). (D) The N terminus of Scc2/NIPBL (red) is enveloped by the TPR repeats of Scc4/MAU2 (green). The helix fragment corresponding to the deletion Gln310-Ala316 of the patient is depicted in light blue. (E) The deleted residues likely interact with residues in the Scc2/NIPBL N terminus and with regions further into the structure.

(F and G) Mammalian two-hybrid assay. (F) Schematic representation of the two-hybrid assay. Full-length MAU2 wild-type (WT) or the MAU2 deletion mutant shown in (C) (mut) were conjugated to the activation domain (AD) of the mouse NF- $\kappa$ B. The first 300 amino acids of NIPBL (NIPBL-N300) were conjugated to the binding domain (BD) of the GAL4 gene. The BD domain alone served as negative control. (G) Results of the assay indicate ~50% reduction of the interaction between NIPBL and MAU2 in the presence of the deletion. Data represent three independent experiments  $\pm$  SEM.

of a weaker interaction with NIPBL, in comparison with wild-type MAU2 (Figure S1).

Taken together, these *in vitro* interaction assays support the *in silico* predictions and demonstrate that the MAU2 deletion impairs the interaction between the two subunits of the kollerin complex.

### Gene Expression Changes in MAU2 Mutant Cells and NIPBL Haploinsufficient Cells Are Similar

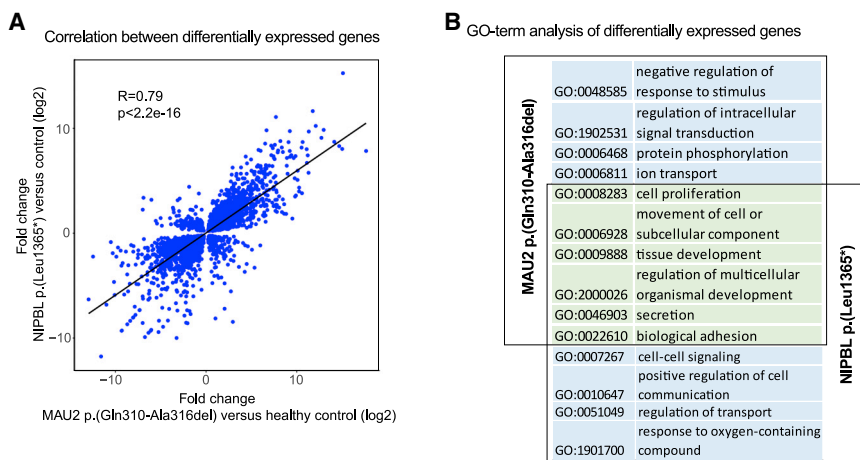
Previous assays demonstrated that the interaction between NIPBL and MAU2 is impaired in the cells of the CdLS patient carrying the in-frame deletion in MAU2 p.(Gln310\_Ala316del). Because this interaction is essential for NIPBL stability (Chao et al., 2015; Hinshaw et al., 2015), we hypothesized that the pathogenic mechanism associated with this mutation in MAU2 is the resulting instability of NIPBL, which leads to its haploinsufficiency.

To substantiate our hypothesis, we compared the transcriptome of fibroblasts from the MAU2 p.(Gln310\_Ala316del) patient (male) with fibroblasts from a NIPBL-haploinsufficient patient (mutation NIPBL p.(Leu1365\*), female) and healthy control fibroblasts (female). Notably, the patient carrying the nonsense NIPBL mutation displays a severe CdLS phenotype, in line with its associated NIPBL haploinsufficiency (Braunholz et al., 2015). All samples were sequenced in triplicates and genes locating at the sex chromosomes were omitted from the analysis. The transcriptomes of the two patients were compared with that of the control fibroblasts, and the differentially expressed genes of the patient cells were subsequently correlated.

Despite the limitation of this dataset, gene expression changes clearly correlated between the two patients. This indicates that overall gene expression is altered in a similar way in cells with MAU2 and NIPBL mutations (Figure 2A). Among the genes differentially expressed in both patients ( $p < 0.05$ ), we found 291 genes that were previously found to be misregulated in lymphoblastoid cells (LCLs) from CdLS patients (Liu et al., 2009). Gene Ontology (GO) analysis of differentially expressed genes in MAU2 mutant fibroblasts and NIPBL mutant fibroblasts indicates categories of molecular and cellular functions that are affected in both cells, such as cell proliferation, cellular movement, tissue development, and regulation of multicellular organismal development (Figure 2; Table S1).

Cells derived from CdLS patients also show increased sensitivity to DNA damage (Enervald et al., 2013; Vrouwe et al., 2007). Accordingly, NIPBL was shown to be involved in DNA damage repair, in both homologous recombination repair, the pathway involving cohesin (Vrouwe et al., 2007), and non-homologous end joining (Enervald et al., 2013), the main pathway during G1 phase of the cell cycle. To assess whether the diagnosis of CdLS of the MAU2 patient could be confirmed also at the DNA repair level, we induced DNA double-strand breaks by gamma irradiation and quantified the number of  $\gamma$ H2AX foci over time as measure of the breaks and their repair (Figure S2). Repair occurred with similar kinetics in the control and the different patient cells and was complete 24 h after irradiation. However, as the cultured skin fibroblasts were mostly in G1 phase, our results are informative only about the non-homologous end-joining pathway.





**Figure 2. Transcriptome Analysis of MAU2 and NIPBL Mutant Cells**

(A) Scatterplot showing the correlation between the fold changes of MAU2 mutant fibroblasts versus control and fibroblasts with NIPBL haploinsufficiency versus control. The Pearson correlation coefficient (R) and the p value (p) are shown in the graph.

(B) Gene Ontology term analysis of differentially expressed genes shows overlap in biological processes between MAU2 and NIPBL mutant patient fibroblasts. Details for this analysis, including p values, are shown in Table S1.

### An Early NIPBL Frameshift Mutation Leads to N-Terminally Truncated NIPBL and Loss of MAU2

MAU2 confers stability to the NIPBL protein through its interaction with the N terminus of NIPBL (Chao et al., 2015; Watrin et al., 2006). To test whether MAU2 could have additional functions besides the mere role of “NIPBL chaperone,” we tried to obtain a stable N-terminally truncated NIPBL that would not require MAU2 for its stability.

For this purpose, we performed CRISPR-Cas9 genome editing in HEK293 cells with single guide RNAs (sgRNAs) targeting exon 2 of *NIPBL* to obtain early truncating mutations through the error-prone non-homologous end-joining DNA repair mechanism. This resulted in two independently isolated homozygous clones with the same out-of-frame duplication of T (c.39dup), predicted to result in the null allele p.(Ala14Cysfs\*5) (Figure 3A). Despite the presence of a homozygous loss-of-function mutation in *NIPBL*, no difference between wild-type and genome-edited cells with regard to cell viability and proliferation was observed.

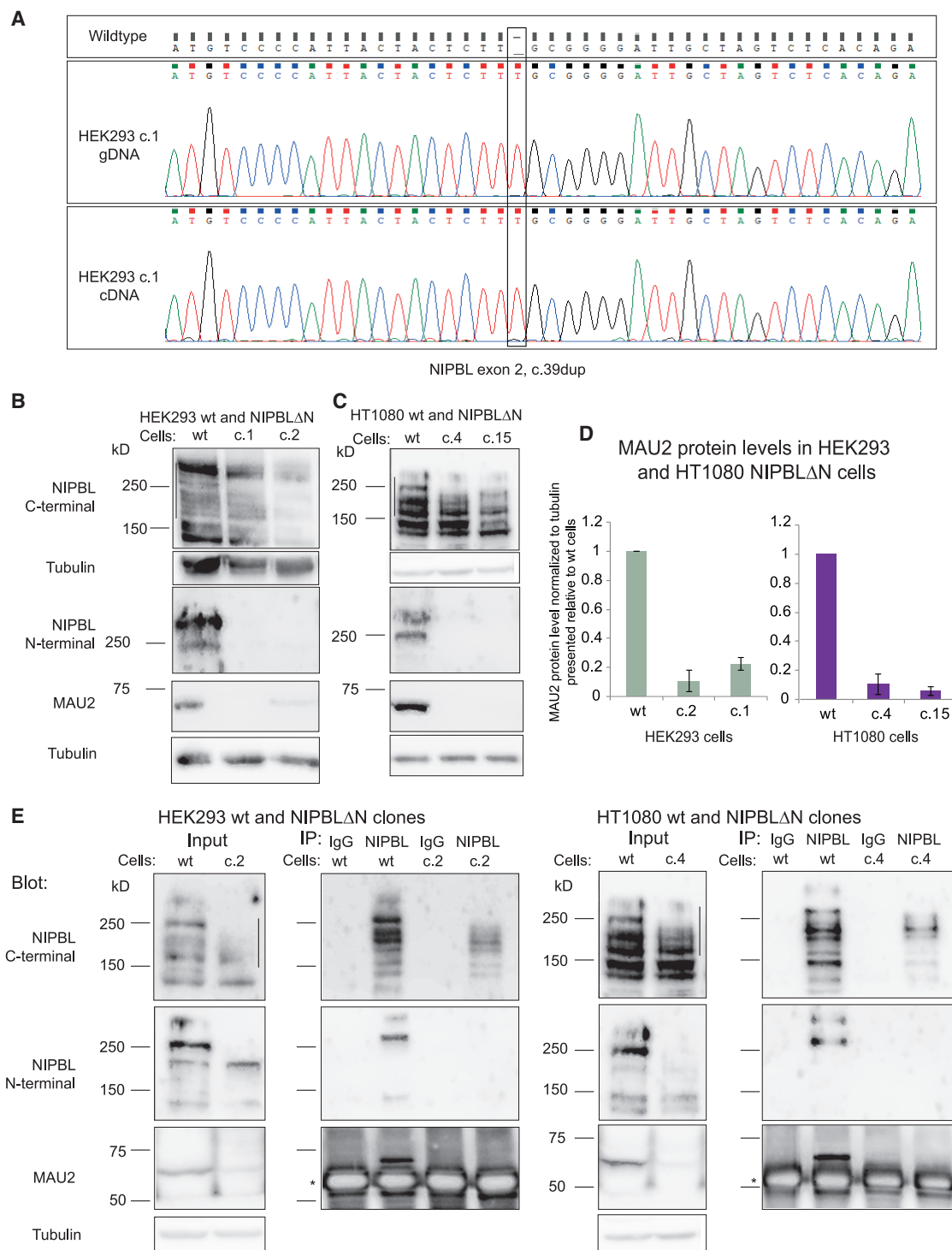
The amount of cohesin complexes present in these cells seems to be unaffected, as the levels of the cohesin subunit SMC3 are unchanged (Figure S3A). However, we observed differences for the kollerin complex subunits NIPBL and MAU2. NIPBL expression was assessed with two different antibodies, one raised against its N terminus and one recognizing its C terminus. The specificity of the antibodies was tested by knocking down MAU2 and NIPBL (Figure S3B). The C-terminal antibodies always produce multiple bands, some specific for NIPBL but some unspecific (Figure S3B). The knockdown experiment also clearly confirmed that depletion of MAU2 leads to a strong reduction of NIPBL and vice versa (Watrin et al., 2006).

In both HEK293 clones carrying the homozygous out-of-frame duplication in *NIPBL*, the N-terminal antibody could not detect any protein. However, we were able to detect NIPBL in the edited clones using the C-terminal antibody, suggesting that only the N terminus of the protein is missing and that an N-terminally truncated form of NIPBL (NIPBLΔN) is expressed in the NIPBL (c.39dup) clones (hereafter called NIPBLΔN cells), possibly resulting from the use of an alternative translation initiation site (TIS). Notably, for the detection of the C terminus of NIPBL, a

combination of antibodies recognizing different NIPBL isoforms was used, thus partially explaining the multiple bands observed in the western blot. Additional degradation products of NIPBL as well as possible isoforms arising from a physiological alternative translation could account for the additional bands observed (Figures 3B and 3C; Figure S3B).

NIPBLΔN appeared to be less abundant than full-length NIPBL in the parental cells, suggesting a lower efficiency of the alternative translation or a reduced stability of NIPBLΔN. Importantly, only a very faint signal could be detected for MAU2, suggesting an almost total loss of MAU2 in NIPBLΔN cells (Figure 3B). To exclude that the observed rescue of NIPBL could be specific for HEK293 cells, we introduced the same out-of-frame duplication of T in the fibrosarcoma cell line HT1080 and observed again the presence of NIPBLΔN and the profound reduction of MAU2 (Figure 3C). Quantification of the residual MAU2 protein in both NIPBLΔN cell types indicated that MAU2 expression is reduced approximately to 15% in genome-edited clones (Figure 3D). Co-immunoprecipitation experiments of MAU2 with anti-NIPBL antibodies in HEK293 and HT1080 clones subsequently showed that the residual MAU2 is unable to interact with NIPBLΔN (Figure 3E).

The presence of NIPBLΔN protein in the genome-edited cells suggested that the mutant transcript escapes nonsense-mediated mRNA decay. Accordingly, Sanger sequencing of the cDNA obtained from NIPBLΔN cells could detect only the mutant allele (Figure 3A). To expand upon this observation, we measured *NIPBL* transcript levels in control and in NIPBLΔN cells using real-time PCR with two different primer pairs covering the boundaries between *NIPBL* exons 2 and 3 and exons 40 and 41. Both primer pairs detected a 1.5- to 1.9-fold increase in *NIPBL* mRNA in both NIPBLΔN clones compared with parental cells (Figure S4A). These results are in line with the observation that the total *NIPBL* transcript levels never drop below 60% in cells derived from patients with different *NIPBL* mutations (Kaur et al., 2016). This could be due to an incomplete degradation of the RNA originating from the mutant allele by nonsense-mediated mRNA decay (Zuin et al., 2017) or to an upregulation of the intact allele as a mean to compensate for the downregulation of the mutant allele (Kaur et al., 2016). Our results show increased levels of *NIPBL* c.39dup transcripts, suggesting a strong upregulation of both mutant alleles, and the absence of



**Figure 3. Cells with NIPBL Out-of-Frame Duplication Express an N-Terminally Truncated NIPBL and Lack MAU2**

(A) Electropherograms of genomic DNA (gDNA) and cDNA of clone 1 (c.1) of the HEK293 edited cells (NIPBLΔN cells), showing homozygous duplication of a T in exon 2 of NIPBL.

(B and C) Blots with nuclear extracts of HEK293 WT and NIPBLΔN clones (c.1 and c.2) (B) and of HT1080 WT and NIPBLΔN clones (c.4 and c.15) (C) probed with antibodies against the N terminus (N-terminal) and the C terminus (C-terminal) of NIPBL. In NIPBLΔN cells, NIPBL can only be detected with the C-terminal antibody that produces multiple bands (bands showing NIPBL are indicated with a line). Detection of MAU2 is only possible in WT cells.

(legend continued on next page)

nonsense-mediated mRNA decay. Subsequent analysis of *MAU2* revealed no significant differences in the total amount of *MAU2* transcript in NIPBLΔN cells in comparison with wild-type cells (Figure S4A). We therefore concluded that *MAU2* is correctly transcribed and translated and that the protein is subsequently degraded because of its inability to interact with NIPBL. To test which known protein degradation pathway might be responsible for this degradation, wild-type and NIPBLΔN cells were treated with the proteasome inhibitor MG132, the lysosomal inhibitor NH<sub>4</sub>Cl, or the autophagy inhibitor bafilomycin. However, none of these inhibitors could restore the expression of *MAU2* in NIPBLΔN cells, ruling them out for the observed *MAU2* degradation (Figures S4B–S4D).

### Identification of Alternative TISs

To address whether an alternative TIS would allow initiation of NIPBL translation in NIPBLΔN cells, we established an *in vitro* transcription-coupled translation (IVTT) assay. A fragment of *NIPBL* cDNA comprising the first 651 bp was inserted into a pcDNA3.1B vector in frame with a 3xFLAG tag at the 3' end of the coding sequence. In parallel, the same construct was generated with the duplication of T (c.39dup). Of importance here, the 3' 3xFLAG tag sequence is in frame only with the wild-type sequence and its canonical ATG but not with that of the mutant sequence harboring the duplication of T (Figures 4A and 4B). The resulting plasmids were used as templates in IVTT reactions that were subsequently analyzed using western blotting. A FLAG signal was detected in the IVTT products of the wild-type as well as of the mutant plasmid (Figure 4C), indicating that the reticulocyte lysate system of the IVTT reaction indeed used an alternative TIS in the presence of the out-of-frame duplication. Notably, the wild-type and mutant IVTT proteins displayed only a very small difference in size (less than 2 kDa), hence indicating that the alternative TIS must be in close proximity to the canonical ATG. The first ATG after the duplication of T would represent the most plausible alternative TIS. However, the first ATG is located 99 amino acids downstream of the duplication. Such a distance would result in a substantial size difference between wild-type and mutant IVTT products. Because no such size difference was observed between the two proteins, we concluded that the potential alternative TIS had to be located upstream of the mutation and might result from the shift of the canonical reading frame (Figure 4A). We identified two new alternative TISs (aTISs) after shifting the reading frame of one base pair (aTIS1 and aTIS2; Figure 4B). To determine which of these putative alternative TISs was actually used, we mutated each aTIS (ATG>ACG) and performed IVTT (Figure 4C). A FLAG signal was detected after mutagenesis of aTIS1, whereas the signal was dramatically reduced for the mutated aTIS2. This indicates that aTIS2 might be used for the initiation of translation of NIPBL in the presence of the out-of-frame duplication. The weak FLAG

signal detected upon mutagenesis of aTIS2 suggests an inefficient usage of aTIS1. Accordingly, mutagenesis of both aTIS fully abolished the signal.

The use of aTIS2 to initiate translation of NIPBLΔN in mutant cells restores the correct frame of the protein after the duplication of T, producing an isoform of NIPBL characterized by a different Nterminus and a correct C terminus. Precisely, only the first 12 amino acids differ from wild-type NIPBL.

However, we cannot rule out that, in the presence of a full-length transcript, additional aTIS might be used from the ribosome to ensure maximal NIPBL yield. In any case, the alternative isoforms are all predicted to be unable to interact with *MAU2*, as the first 38 amino acids of NIPBL are mediating this interaction (Braunholz et al., 2012). Consistently, NIPBLΔN cells showed a strong reduction of *MAU2*, most likely caused by its inability to interact with NIPBL.

### NIPBLΔN Cells Do Not Display Cohesion Defects

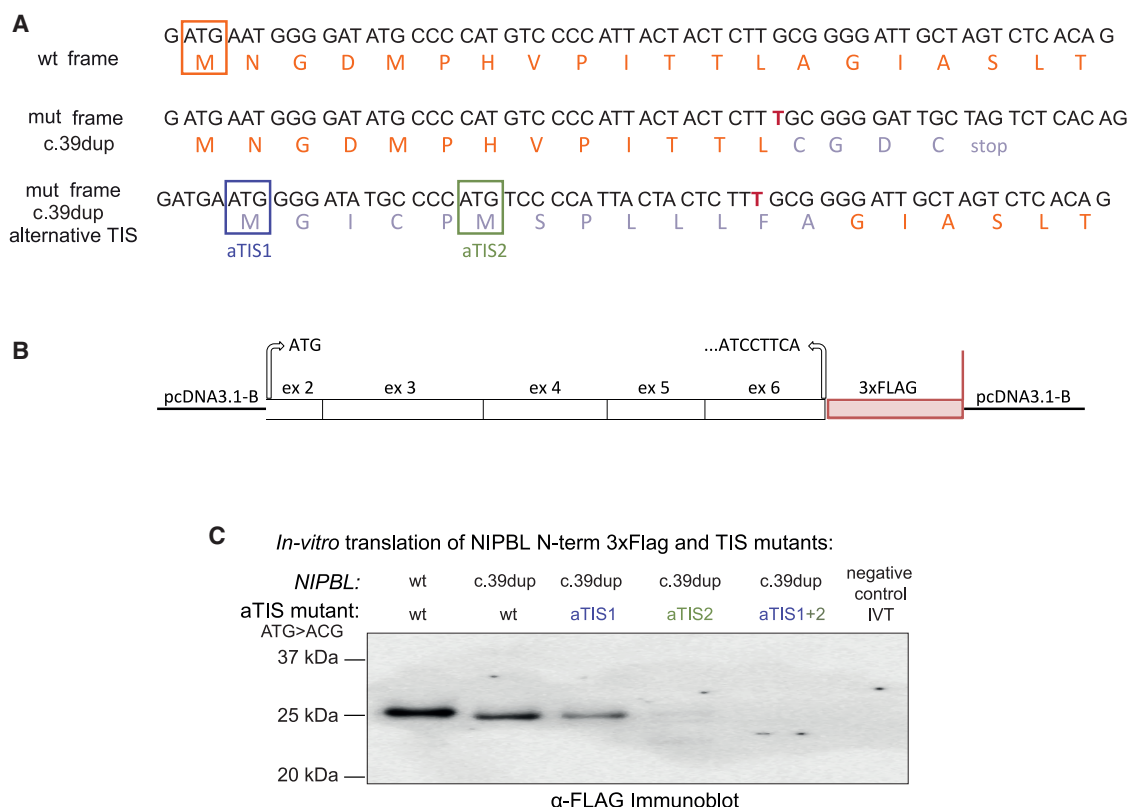
NIPBL is well known for its role as cohesin loader. The canonical function of the cohesin complex is the maintenance of sister chromatid cohesion during both mitosis and meiosis. To address whether the NIPBLΔN protein is capable to support cohesin loading without *MAU2*, we first tested cohesin's function in sister chromatid cohesion. We analyzed NIPBLΔN and control cells by spreading and Giemsa staining of mitotic chromosomes (Figure 5A). Normal cohesion was observed in 95% of HEK293 control cells and 92% and 87% in NIPBLΔN c.1 and c.2 cells, respectively. Differences were statistically significant only for clone c.2 ( $p = 0.199$  for c.1,  $p = 0.001$  for c.2) (Figure 5B). For HT1080 cells, we observed normal cohesion in 98.1% of wild-type cells and in 99.7% and 99.4% of cells of NIPBLΔN c.4 and c.15 clones ( $p = 0.018$  for c.4 and  $p = 0.122$  for c.15) (Figure 5C). We also determined the mitotic indices in the HEK293 and HT1080 clones and observed around 3% of mitotic cells in all cell lines with only a slight increase to 4.5% in the HEK293 NIPBLΔN c.2 clone (Figures 5D and 5E). In summary, we do observe for HEK293 NIPBLΔN cells a slight increase in defective sister chromatid cohesion and an increased mitotic index for one clone. This is not recapitulated by the HT1080 cells. Therefore, we concluded that the differences in HEK293 cells might be due to clonal variation and that NIPBLΔN expression in combination with reduction of *MAU2* levels leave chromatid cohesion largely unaffected in mitotic cells.

Further analysis of sister chromatid cohesion in the HEK293 cells was performed by measuring the distance between paired fluorescence *in situ* hybridization (FISH) signals in interphase cells. We used a FISH probe that maps on chromosome 21q22.2, a locus that is tetrasomic in HEK293 cells and therefore labels four pairs of sister chromatids in G2 cells (Figure 5F). In wild-type cells, the average interchromatid distance was 0.57 μm. The presence of NIPBLΔN in both *NIPBL* c.39dup

(D) Quantification of *MAU2* protein levels on western blots of total cell lysates from HEK293 and HT1080 NIPBLΔN cells. Data are shown as mean of three independent experiments ± SEM.

(E) Immunoprecipitation (IP) of NIPBL from nuclear extracts of HEK293 and HT1080 WT and NIPBLΔN cells and detection of NIPBL and *MAU2*. NIPBL was pulled down from WT and NIPBLΔN cells while *MAU2* co-precipitated only in WT cells, meaning that the residual *MAU2* is unable to interact with NIPBLΔN. Note that the strong signals in the *MAU2* immunoblot are the IP antibodies.

Note that the blot for NIPBL C-terminal in HT1080 WT and clones are identical between (C) and (E); in (E) one lane of the clone not used for IP was blanked out.



**Figure 4. An Alternative Translation Initiation Site in NIPBL**

(A) An N-terminally truncated NIPBL (NIPBLΔN) is produced in HEK293 c.39dup cells. A very small size difference was observed in the western blot between wild-type and mutant proteins, indicating close proximity of the alternative translation initiation site (TIS) to the canonical ATG. Two new ATGs appear in the sequence around the T duplication (red) after shifting the reading frame by one base pair (aTIS1 and aTIS2).

(B) To test the existence of an alternative translation start site in NIPBLΔN cells, the first 651 bp of NIPBL (wild-type [WT] or with the c.39dup) were cloned into a vector for IVTT with a 3xFLAG tag that was designed in frame with the WT sequence. Plasmids carrying mutations of the new putative aTIS, in addition to the duplication of T, were also generated.

(C) All plasmids were used as templates in IVTT reactions, and the IVTT products were analyzed by western blotting using anti-FLAG antibodies. A FLAG signal was detected in the IVTT products of the WT NIPBL sequence and of the NIPBL c.39dup mutant, confirming that an alternative TIS is used in the presence of the T duplication. FLAG was detectable with the aTIS1 mutation but was much reduced in the aTIS2 mutation, pointing to aTIS2 as aTIS for NIPBLΔN. Mutation of both aTIS fully abolished the FLAG signal.

clones did not change the distance between FISH signals significantly (Figure 5G), indicating that sister chromatid cohesion in postreplicative cells is not altered by the NIPBL c.39dup mutation.

Altogether, these results indicate that sister chromatid cohesion during interphase and mitosis is largely unaffected by the presence of NIPBLΔN and a strong reduction of MAU2 levels.

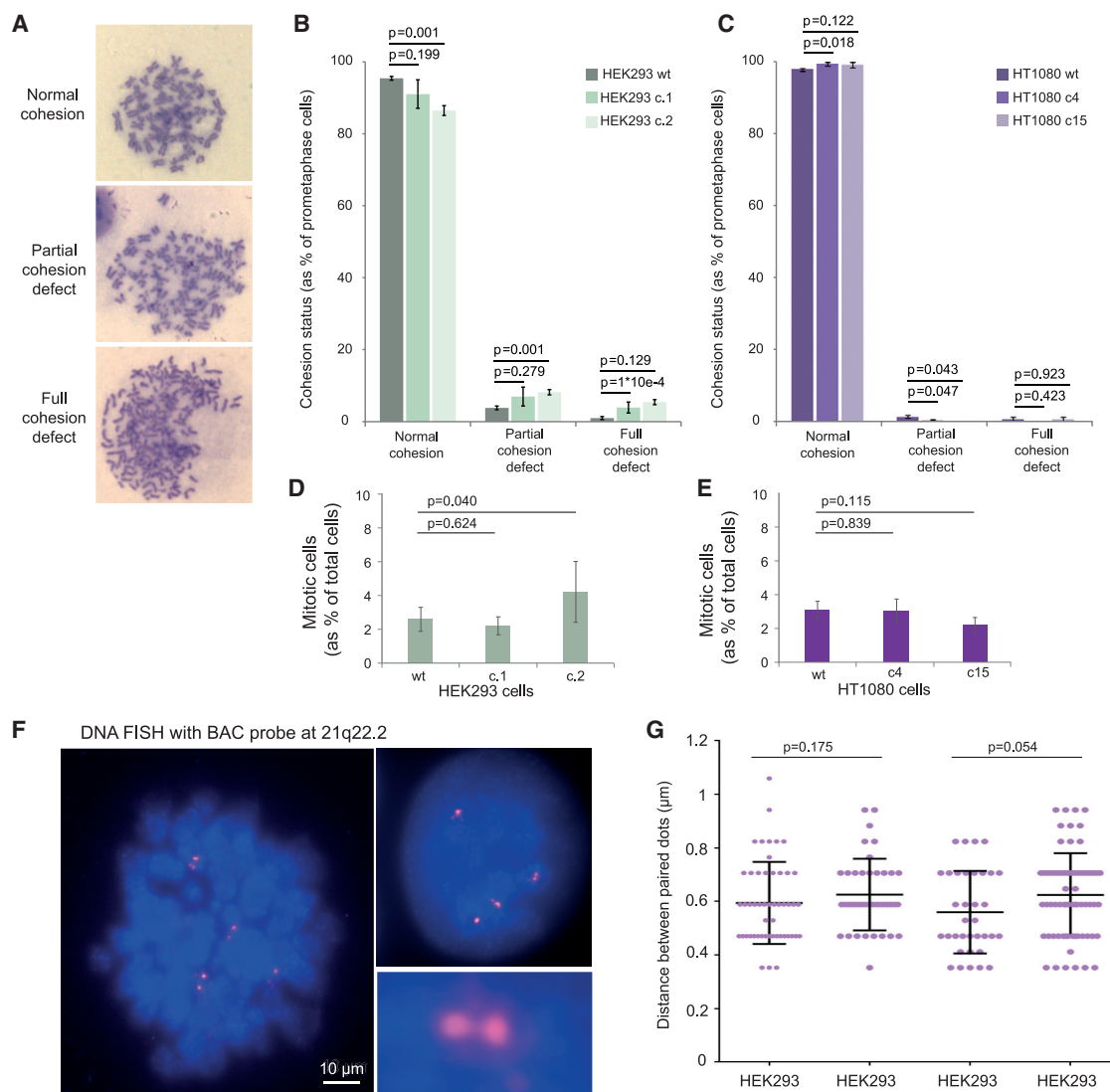
### Cohesin Binding to Chromatin Is Not Affected in NIPBLΔN Cells

To evaluate whether and how the MAU2 deficiency and the absence of the N terminus of NIPBL affect the abundance and chromatin-binding ability of NIPBL and cohesin, we fractionated wild-type and NIPBLΔN cells (HEK293 c.1 and c.2) into total, soluble and chromatin-bound fractions (Figure 6A). A modest reduction of NIPBL and the strong reduction of MAU2 were observed in all fractions of both NIPBLΔN clones. No difference in the total amount of chromatin-bound cohesin was observed

between the two cell lines, indicating that cohesin complexes can be loaded onto chromatin by N-terminally truncated NIPBL and in the absence of MAU2.

Next, we tested whether the localization of NIPBL and cohesin on chromatin is affected and performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) for NIPBL and the cohesin subunit SMC3. We called 23,194 SMC3 peaks in wild-type cells and 21,984 peaks in NIPBLΔN cells (Figure 6B; Figure S5A). The heatmaps comparing the peaks signals in wild-type cells with the signal intensity in NIPBLΔN cells showed a slightly reduced signal in the NIPBLΔN cells, an effect that is more pronounced at weaker peaks. The distribution of the peaks over different genomic features, however, is unchanged (Figure 6C). For NIPBL we called 2,681 peaks for wild-type cells and 2,505 and 2,695 peaks for NIPBLΔN clones 1 and 2, respectively (Figure 6B; Figure S5B). The identified peaks do not show a change in feature distribution (Figure 6C). To identify also NIPBL binding sites where NIPBL is not in direct contact with DNA or





**Figure 5. NIPBL $\Delta$ N Cells Do Not Display Cohesion Defects**

(A) Representative images of the categories to cluster the chromosome spreads. For normal cohesion, chromosomes contain a primary constriction and sister chromatids are tightly connected at their centromeres (upper panel). For partial cohesion defects, chromosomes lack a primary constriction and sister chromatids are abnormally spaced while still closely opposed to each other (middle panel). Complete separation of sister chromatids indicates full cohesion defect (bottom panel).

(B and C) Frequency of the different cohesion phenotypes determined in HEK293 cells (B) and HT1080 cells (C), expressed as a percentage of total prometaphase cells. Data are shown as means of three independent experiments  $\pm$  SEM. p values obtained with the two-tailed, unpaired t test are indicated.

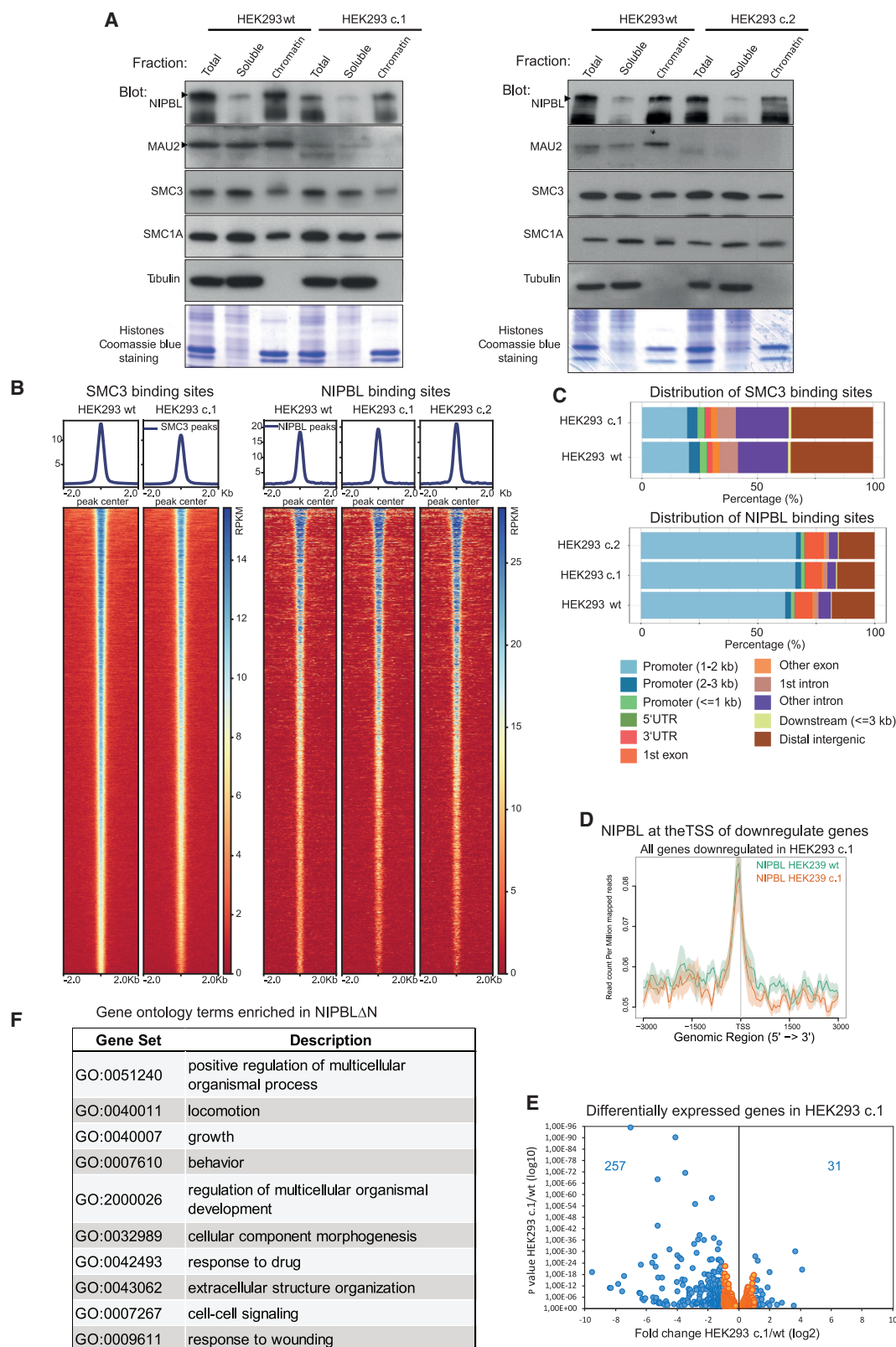
(D and E) Mitotic indices of wild-type and NIPBL $\Delta$ N mutant clones in HEK293 (D) and HT1080 (E) cells. Data are shown as means of three independent experiments  $\pm$  SEM. p values obtained with the two-tailed, unpaired t test are indicated.

(F) Representative image of interphase DNA FISH performed on NIPBL $\Delta$ N cells. Scale bar represents 10  $\mu$ m.

(G) Quantification of the interphase DNA FISH shows that the distance between paired FISH signals is not affected, confirming the absence of cohesion defects. p values obtained with the two-tailed, unpaired t test are indicated.

weakly bound to it, we used a ChIP-seq protocol involving protein-protein crosslinking in addition to formaldehyde crosslinking. Here 14,198 peaks in wild-type cells and 9,124 peaks in NIPBL $\Delta$ N cells (Figure S5B) were called. Heatmaps comparing the peak signals in wild-type cells with NIPBL $\Delta$ N cells (Figure S5C) show that on all wild-type peaks we can still obtain signal in NIPBL $\Delta$ N cells, indicating that a large number of the

peaks in NIPBL $\Delta$ N is still present but just fell below the peak cut-off. This is also supported by the notion that the overall distribution of the peaks over different genomic features is not altered (Figure S5D). The results obtained by ChIP-seq were subsequently confirmed by ChIP-qPCR for SMC3 and NIPBL at different genomic loci, confirming that cohesin- and NIPBL-binding sites are largely unchanged in NIPBL $\Delta$ N cells (Figure S6).



(legend on next page)

Altogether, these experiments indicate that NIPBLΔN is able to bind chromatin and to mediate normal cohesin loading onto DNA despite the absence of MAU2. However, although the position of the sites along the genome is not altered, a lower occupancy was observed, consistent with the reduction of the chromatin-bound NIPBL fraction (Figure 6A).

### NIPBLΔN Cells Show Moderate Changes in Gene Expression

NIPBL can influence gene expression directly or by loading the cohesin complex on chromatin (Zuin et al., 2014). To test whether the truncation of NIPBL triggers transcriptional changes, we performed RNA sequencing for HEK293 wild-type cells and two NIPBLΔN clones, three replicates each. We observed a number of differentially expressed genes that correlate well between the two clones (Figure S7A). By correlating the transcriptome data with the ChIP-seq signals for NIPBL, we could observe that the coverage of the transcription start site (TSS) of downregulated genes with NIPBL seems to be largely unchanged (Figure 6D). Signals of the cohesin subunit SMC3 was observed only at the TSS of downregulated genes, and these signals appeared to be slightly reduced in the NIPBLΔN cells (Figure S7C). This is consistent with the slightly reduced overall SMC3 binding (Figure 6B). Applying stringent criteria to the RNA sequencing data ( $p < 0.05$ ; consistency between clones,  $|\text{fold change [FC]}| > 2$ ), we observed 31 upregulated genes and 257 downregulated genes in the HEK293 NIPBLΔN cells (Figure 6E; Tables S3 and S4). Only two of the upregulated genes but 41 downregulated genes have a NIPBL site near the promoter ( $\pm 1$  kb around the TSS). GO term analysis showed a broad range of terms related to cell growth and development (Figure 6F), which would be consistent with the involvement of NIPBL and cohesin in development. In summary, these results show that the rescue of kollerin complex functions by alternative translation might not be complete, as a number of genes seem to depend on the presence of a wild-type NIPBL with intact N terminus and MAU2.

### Cells of a CdLS Patient with an Early Splice Site Mutation in NIPBL Show MAU2 Instability

To show that the observed rescue mechanism for NIPBL early truncation mutations that leads to an alternative N terminus and renders MAU2 largely dispensable is relevant for CdLS patients, we searched for cases in which this mechanism might play a role.

We obtained LCLs from a patient with the clinical diagnosis of mild CdLS (AG0885) that harbors a heterozygous early splice site

variant in *NIPBL* (c.358+1G>C) (hereafter termed NIPBL c.358+1G>C cells). Similar mutations in two patients with mild CdLS features have already been described and characterized (Teresa-Rodrigo et al., 2014). Here, the c.358+1G>A substitution led to the out-of-frame skipping of exon 4 of *NIPBL*, predicted to result in a premature truncation of NIPBL after amino acid 77 (NIPBL truncation p.[Ile77Metfs\*5]). We could confirm the skipping of *NIPBL* exon 4 in the NIPBL c.358+1G>C cells by amplification and sequencing of cDNA (data not shown) (Figure 7A). Transcripts with such an out-of-frame deletion are thought to be the target of nonsense-mediated mRNA decay, thereby leading to *NIPBL* haploinsufficiency. To test the relevance of nonsense-mediated mRNA decay, we determined the total and allele-specific *NIPBL* transcript levels in the NIPBL c.358+1G>C cells and two healthy controls (LCLs AG0089 and AG0096) (Figure 7B). Total *NIPBL* transcripts are not reduced in the NIPBL c.358+1G>C cells (Figure 7C) and the allele-specific real-time PCR suggested that the wild-type and the mutant allele are equally expressed in these cells (Figure 7D). Altogether, these results corroborate the hypothesis that the *NIPBL* mutant transcript does not undergo nonsense-mediated mRNA decay in the NIPBL c.358+1G>C cells.

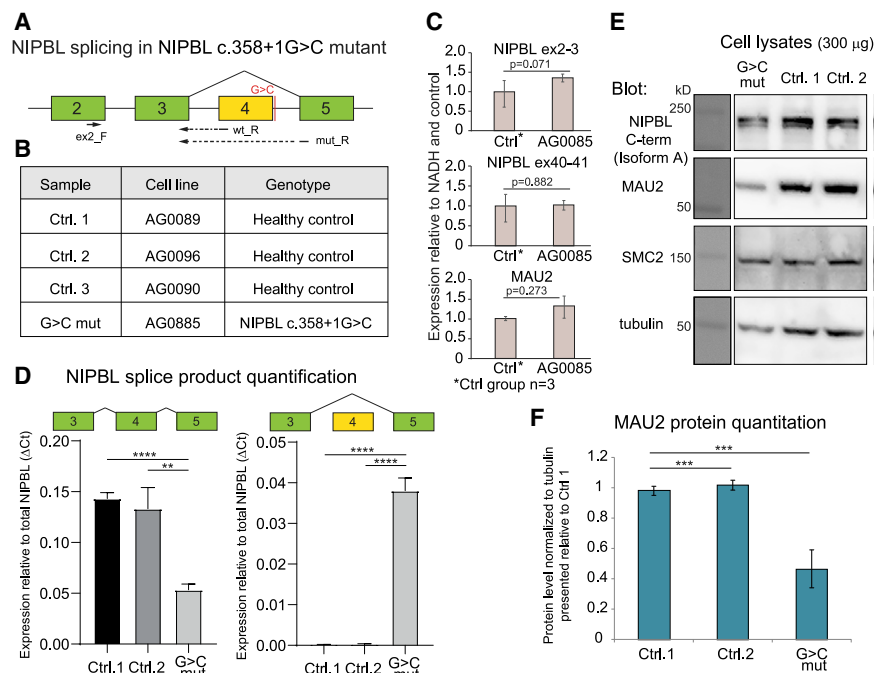
To investigate the effect of this mutation on protein levels, we analyzed NIPBL and MAU2 protein levels in nuclear extracts from NIPBL c.358+1G>C cells and healthy controls (Figure 7E). Despite the splice site mutation, NIPBL protein levels detected with an antibody against the C terminus of NIPBL (isoform A) appeared to be largely unchanged, whereas MAU2 protein levels are reduced up to 50% in NIPBL c.358+1G>C cells (Figure 7F).

These observations are in line with our results from NIPBLΔN cells (HEK293 and HT1080). An alternative translation initiation could prevent the reduction of NIPBL levels in NIPBL c.358+1G>C cells. The resulting N-terminal truncation is too small relative the full-length NIPBL protein to cause a visible molecular weight change. However, as demonstrated in the NIPBLΔN cells, an N-terminal truncation would affect the N-terminal NIPBL residues interacting with MAU2. This would lead to reduced MAU2 protein levels, as observed.

From a phenotypical point of view, the patient with NIPBL c.358+1G>C mutation and the two previously described patients with the splicing mutation affecting nucleotide c.358+1 display mild CdLS features (Teresa-Rodrigo et al., 2014; David Fitzpatrick, personal communication). Patients with truncating mutations occurring later in the protein sequence are generally characterized by moderate to severe phenotypes and often present with limb anomalies (Selicorni et al., 2007). Hence, the

### Figure 6. Cohesin Binding to Chromatin Is Not Affected in NIPBLΔN Cells

(A) Fractionation of wild-type and NIPBLΔN cells (c.1 and c.2) in soluble and chromatin-bound proteins shows a small reduction of NIPBL and loss of MAU2 in the soluble pool but also on chromatin in NIPBLΔN cells. No differences in the amounts of soluble or chromatin-bound cohesin are observed. (B) The heatmaps depict normalized ChIP coverage observed in the SMC3 and NIPBL ChIP-seq experiments performed with formaldehyde crosslinking for wild-type and NIPBLΔN cells. The heatmaps are centered on the peaks observed in wild-type cells, and the color intensity relates to normalized counts (RPKM). (C) Distribution of the peaks observed in the SMC3 and NIPBL ChIP-seq experiments over different genomic features (promoters, introns, exons, etc.) for wild-type and NIPBLΔN cells. No striking differences are visible between wild-type and NIPBLΔN cells. (D) Plot showing the NIPBL coverage at promoters of genes downregulated in NIPBLΔN cells. (E) Volcano plot showing the genes differentially expressed in NIPBLΔN clone c.1. Only genes with the criterion  $p$  value  $< 0.05$  and consistency between the analyzed clones are shown. Genes with  $|\text{FC}| < 2$  are colored orange. Genes with  $|\text{FC}| > 2$  are displayed in blue, and the numbers of genes in this category are shown. (F) Gene Ontology term analysis of the genes shown in (E).



**Figure 7. NIPBL Splice Site Mutation in CdLS Leads to MAU2 Instability**

(A) Scheme indicating the position of the NIPBL splice site mutation c.358+1G>C and the resulting splice product excluding exon 4. The qPCR primers used to quantify the splice products are indicated.

(B) NIPBL mutant and control LCLs used.

(C) Messenger RNA levels of NIPBL and MAU2 in CdLS patient cells with NIPBL (c.358+1G>C) mutation and controls were determined by RT-PCR/qPCR relative to the housekeeping gene NADH. For NIPBL, two intron-skipping primer pair at opposite ends of the transcript were used (exons 2 and 3 and exons 40 and 41). SDs are from qPCR replicates; p values were calculated for the three controls and one patient using the unpaired t test with Welch's correction.

(D) Quantitation of the mRNA originating from the correct splice product (left panel, primers ex2\_F+wt\_R) and of the mutant splice product (right panel, primers ex2\_F+mutR). Data are shown as mean of n = 3 experiments ± SEM. p values obtained with the two-tailed, unpaired t test are indicated as \*\*p < 0.05, \*\*\*p < 0.005, and \*\*\*\*p < 0.0005.

(E) Nuclear extracts of controls (Ctrl.1 and Ctrl.2) and of NIPBL c.358+1G>C mutant cells (G>C mut)

were analyzed by western blotting with antibodies against NIPBL (C-terminal antibody detecting isoform A), MAU2, and SMC2 and tubulin as loading controls. (F) Quantitation of the MAU2 protein levels relative to tubulin and to Ctrl.1. \*\*p < 0.05 (two-tailed, unpaired t test from n = 3 experiments).

translation-initiation-based rescue mechanism seemingly protects patients from the otherwise more severe consequences arising from NIPBL haploinsufficiency.

## DISCUSSION

MAU2 (Scc4 in *S. cerevisiae*) is a ubiquitously expressed protein that interacts with NIPBL (Scc2 in *S. cerevisiae*) (Ciosk et al., 2000; Seitan et al., 2006; Watrin et al., 2006) in a complex termed kollerin. The interaction with NIPBL is crucial for MAU2 protein stability and vice versa (Watrin et al., 2006). Kollerin is involved in cohesin loading and transcriptional regulation (Ciosk et al., 2000; Zuin et al., 2014). Mutations in NIPBL are the primary genetic cause of CdLS, whereas alterations of its binding partner MAU2 have not been reported so far (Mannini et al., 2013). Here we describe the first genetic variant in MAU2 causing a CdLS phenotype. The observed p.(Gln310\_Ala316del) variant in MAU2 impairs the interaction with NIPBL, possibly by distorting the three-dimensional structure of MAU2. Consistently, reduced affinity was observed between MAU2 and NIPBL with *in vitro* assays in the presence of the in-frame MAU2 deletion detected in the CdLS patient. The deletion in MAU2 lies in a “highly conserved surface patch” described by Hinshaw et al. (2015) in budding yeast Scc4. In yeast, mutations in this patch lead to a plasmid segregation phenotype and reduced cohesin loading at the centromere, while cohesin loading on chromosome arms remained unaffected (Hinshaw et al., 2015). However, centromere organization is fundamentally different between yeast and human.

Therefore, we propose that this deletion exerts its deleterious effects in a way comparable with mutations that directly affect

the NIPBL coding sequence and leads to NIPBL haploinsufficiency. In agreement with this proposition, the patient's phenotype resembles the phenotype of the spectrum of NIPBL mutations. The commonalities between the MAU2 and NIPBL patients include typical CdLS facial dysmorphism, microcephaly, intellectual disability, and delayed motor and verbal development. Additionally, these patients also share similar changes in gene expression patterns. A particular resemblance stands out when comparing the phenotype of the MAU2 p.(Gln310\_Ala316del) patient with a patient carrying the p.(Gly15Arg) substitution in NIPBL, a mutation shown to interrupt the interaction with MAU2 (Bot et al., 2017; Braunholz et al., 2012). This resemblance further supports the hypothesis that the deletion in MAU2 and the NIPBL missense mutation are associated with similar downstream molecular consequences.

By further investigating the repercussions of early truncations in NIPBL, we found that cells expressing an N-terminally truncated form of NIPBL (NIPBLΔN) are fully viable, although the interaction with MAU2 that depends on the first 38 residues (Braunholz et al., 2012) is impaired. As consequence, these cells have nearly completely lost the MAU2 protein, in line with the observation that depletion of NIPBL by RNA interference reduces MAU2 levels (Watrin et al., 2006). Similar observations have been made by Haarhuis et al. (2017).

Molecular analyses of these cells indicated that NIPBLΔN is still able to bind chromatin and to mediate cohesin loading onto DNA despite the almost total loss of MAU2. This observation is consistent with overexpression assays that demonstrate that a N-terminally truncated form of NIPBL that is not capable to interact with MAU2 can still be recruited to DNA damage sites (Bot et al., 2017). In addition, it was reported that, *in vitro*, Scc2



alone displayed DNA binding properties indistinguishable from that of the kollerin complex and was able to mediate cohesin loading (Murayama and Uhlmann, 2014), although another study showed that Scc4 is required in budding yeast for *in vivo* recruitment of Scc2 to chromatin (Chao et al., 2015). It was suggested that the Scc2 N terminus, together with Scc4, is necessary to localize kollerin to chromatin (Chao et al., 2015). Furthermore, the N terminus of NIPBL was reported to be in contact with SMC1A, while the interaction to RAD21 (Scc1 in *S. cerevisiae*) is mediated by the C-terminal HEAT-repeat domain of NIPBL (Chao et al., 2017). In NIPBLΔN cells, cohesin and NIPBL DNA-binding sites are largely unchanged, indicating that NIPBLΔN is fully capable to load and position cohesin. Cohesin loaded by NIPBLΔN also proved to be able to accomplish the main function of the complex for centromere and chromosome arm cohesion.

Does MAU2 have more functions besides stabilizing the NIPBL N terminus in human cells? So far, NIPBLΔN can fulfill the roles of the kollerin complex despite its inability to interact with MAU2. The specific roles shown for centromeric cohesin in yeast is unlikely to be relevant in mammals. However, our transcriptome analysis in NIPBLΔN cells revealed a number of gene expression changes, mostly downregulation, which cannot be explained by reduced NIPBL coverage at promoters. Izumi et al. (2019) similarly concluded that the NIPBL N terminus is important for the transcriptional regulatory role of NIPBL. In this view, MAU2 could have a role in fine-tuning the expression of genes by modulating kollerin's interaction with other transcription factors localizing to these promoters (Zuin et al., 2014).

CdLS is characterized by extensive clinical variability, and patients present with a wide range of phenotypes (Kline et al., 2007). Heterozygous out-of-frame deletions and insertions or nonsense variants in NIPBL are associated mainly with severe phenotypes and with the presence of limb malformations (Selicorni et al., 2007). However, truncating mutations affecting exons 2–9 of NIPBL are associated with a lower frequency of limb reductions or malformations and often result in phenotypes that are milder than expected in comparison with truncating mutations affecting exons 11–47 (Borck et al., 2004; Nizon et al., 2016; Oliveira et al., 2010). Our genome editing experiments demonstrated that cells with early truncations in NIPBL adopt a protective mechanism on the basis of alternative translation initiation in order to minimize the otherwise deleterious effects of mutations, thus offering an explanation for the milder phenotypes of these patients. Translation re-initiation is a widespread mechanism for escaping nonsense-mediated mRNA decay in human genes (Lindeboom et al., 2016). In most cases, the distance between the original start codon and the premature termination codon can be considered a strong predictor of the ability to escape nonsense-mediated mRNA decay (Lindeboom et al., 2016). In our cells, the choice of the new TIS appears to be dependent on the nucleotide sequence surrounding the variant of interest, thereby rendering this protective mechanism mutation-specific. It is conceivable that the CdLS cells carrying the NIPBL splicing mutation that leads to skipping of exon four (c.358+1G>C), and that is predicted to insert an early stop codon after 5 amino acids, might use a different aTIS than the one used by the NIPBLΔN cells, which harbor a frameshift mutation in

exon 2. This rescue mechanism ensures the synthesis of the major part of NIPBL at the expense of its N terminus, which appears to be dispensable for cell survival. Congruent with this hypothesis, cap analysis of gene expression (CAGE) data suggest the existence of an alternative TSS in exon 10 of NIPBL (<http://fantom3.gsc.riken.jp/>). A recent study also demonstrated that cells could tolerate disruptive mutations up to exon 10, indicating that the region between exons 11 and 47 alone is able to accomplish the main tasks of NIPBL and is therefore essential for cell survival (Haarhuis et al., 2017). Additionally, the existence of an N-terminal truncation of NIPBL has been proved in physiological conditions in yeast, where the cell cycle-dependent cleavage of the first 150 amino acids regulates the spatiotemporal association of cohesin with chromatin (Woodman et al., 2014). In the case we present here, however, alternative translation initiation was observed only in the presence of a deleterious mutation in the NIPBL transcript. Mutations in MAU2 that impair kollerin complex stability, like the one reported in this paper, would not be expected to trigger such protective mechanism.

In summary, our data strongly indicate the existence of a protective mechanism preventing a total loss of NIPBL gene product by the use of alternative start codons in transcripts with early truncating mutations. Such translation re-initiation could account for a proportion of patients who carry truncating mutations in NIPBL and exhibit unexpectedly mild phenotypes, but might also apply to other genetic conditions. A more exhaustive understanding of the molecular mechanisms triggering this rescue could help define a better genotype-phenotype correlation and could open perspectives for translational approaches.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human subjects
  - Detailed characterization of the MAU2 patient
  - Cell lines
- METHOD DETAILS
  - CRISPR/Cas9
  - Protein isolation and quantification
  - RNA isolation, cDNA synthesis and Real-Time PCR
  - Protein degradation pathways
  - *In vitro* transcription-coupled translation
  - Subcellular fractionation
  - siRNA knockdown of MAU2 and NIPBL
  - ChIP-sequencing
  - ChIP-qPCR
  - Precocious sister chromatid separation assays by chromosome spreading
  - Determination of the mitotic index
  - Fluorescence *in situ* hybridization (FISH)

- Whole Genome Sequencing
- Mammalian two-hybrid assay
- Yeast two-hybrid assay
- Gamma irradiation and  $\gamma$ -H2AX foci count
- RNA-sequencing
- Gene ontology term enrichment analysis

## ● QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.107647>.

## ACKNOWLEDGMENTS

We thank Jaulin's team for access to the microscope, Koichi Tanaka and Kim Nasmyth for sharing the NIPBL antibody, and Jan-Michael Peters for sharing the MAU2 antibody. We thank David Fitzpatrick and Morad Ansari for sharing the AG0885 LCLs used to prove MAU2 instability in CdLS patients with early truncations. We thank Ifigeneia Thomopoulou for help with western blotting and Maarten Fornerod for advice on transcriptome analyses. This work was funded by the German Federal Ministry of Education and Research (BMBF) (CHROMATIN-Net to F.J.K.), by the German Research Foundation (DFG) (research unit FOR2488 to F.J.K.), by E-Rare-2 TARGET-CdLS (to F.J.K., K.S.W., and E.W.), and by the Medical Faculty of the University of Lübeck (J09-2017 to I.P.). Work in the lab of K.S.W. was funded by Dutch Cancer Society (KWF) grant EMCR 2015-7857 and by the by Netherlands Organization of Scientific Research (NWO-BBOL) grant 737.016.014.

## AUTHOR CONTRIBUTIONS

Conceptualization, I.P., E.W., F.J.K., and K.S.W.; Investigation, I.P., F.D., S.R.G., V.C., V.D., J.E., T.v.S., E.W., and K.S.W.; Formal Analysis, I.P., F.D., E.M., R.B., R.W.W.B., V.D., T.S., T.M.S., M.M.G., E.W., F.J.K., and K.S.W.; Resources, R.B., E.G., B.P., F.R., T.S., W.F.J.v.I.J., T.M.S.; Supervision, E.W., F.J.K., and K.S.W.; Writing – Original Draft, I.P. and K.S.W.; Writing – Review & Editing, I.P., E.W., F.J.K., and K.S.W.; Funding Acquisition, E.W., F.J.K., and K.S.W.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 6, 2019

Revised: September 30, 2019

Accepted: April 24, 2020

Published: May 19, 2020

## REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Borck, G., Redon, R., Sanlaville, D., Rio, M., Prieur, M., Lyonnet, S., Vekemans, M., Carter, N.P., Munnich, A., Colleaux, L., and Cormier-Daire, V. (2004). NIPBL mutations and genetic heterogeneity in Cornelia de Lange syndrome. *J. Med. Genet.* 41, e128–e128.
- Bot, C., Pfeiffer, A., Giordano, F., Edara, D.M., Dantuma, N.P., and Ström, L. (2017). Independent mechanisms recruit the cohesin loader protein NIPBL to sites of DNA damage. *J. Cell Sci.* 130, 1134–1146.
- Braunholz, D., Hullings, M., Gil-Rodríguez, M.C., Fincher, C.T., Mallozzi, M.B., Loy, E., Albrecht, M., Kaur, M., Limon, J., Rampuria, A., et al. (2012). Isolated NIPBL missense mutations that cause Cornelia de Lange syndrome alter MAU2 interaction. *Eur. J. Hum. Genet.* 20, 271–276.
- Braunholz, D., Obieglo, C., Parenti, I., Pozojevic, J., Eckhold, J., Reiz, B., Braenne, I., Wendt, K.S., Watrin, E., Vodopietz, J., et al. (2015). Hidden mutations in Cornelia de Lange syndrome limitations of sanger sequencing in molecular diagnostics. *Hum. Mutat.* 36, 26–29.
- Chao, W.C.H., Murayama, Y., Muñoz, S., Costa, A., Uhlmann, F., and Singleton, M.R. (2015). Structural studies reveal the functional modularity of the Scc2-Scc4 cohesin loader. *Cell Rep.* 12, 719–725.
- Chao, W.C.H., Murayama, Y., Muñoz, S., Jones, A.W., Wade, B.O., Purkiss, A.G., Hu, X.-W., Borg, A., Snijders, A.P., Uhlmann, F., and Singleton, M.R. (2017). Structure of the cohesin loader Scc2. *Nat. Commun.* 8, 13952.
- Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* 5, 243–254.
- Deardorff, M.A., Wilde, J.J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Mönnich, M., Yan, Y., Xu, W., Gil-Rodríguez, M.C., et al. (2012). RAD21 mutations cause a human cohesinopathy. *Am. J. Hum. Genet.* 90, 1014–1027.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Envervald, E., Du, L., Visnes, T., Björkman, A., Lindgren, E., Wincet, J., Borck, G., Colleaux, L., Cormier-Daire, V., van Gent, D.C., et al. (2013). A regulatory role for the cohesin loader NIPBL in nonhomologous end joining during immunoglobulin class switch recombination. *J. Exp. Med.* 210, 2503–2513.
- Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* 7, 1728–1740.
- Gervasini, C., Russo, S., Cereda, A., Parenti, I., Masciadri, M., Azzollini, J., Melis, D., Aravena, T., Doray, B., Ferrarini, A., et al. (2013). Cornelia de Lange individuals with new and recurrent SMC1A mutations enhance delineation of mutation repertoire and phenotypic spectrum. *Am. J. Med. Genet. A.* 161A, 2909–2919.
- 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., et al. (2015). De novo heterozygous mutations in SMC3 cause a range of Cornelia de Lange syndrome-overlapping phenotypes. *Hum. Mutat.* 36, 454–462.
- Gillis, L.A., McCallum, J., Kaur, M., DeScipio, C., Yaeger, D., Mariani, A., Kline, A.D., Li, H.H., Devoto, M., Jackson, L.G., and 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.
- Haack, T.B., Kopajtich, R., Freisinger, P., Wieland, T., Rorbach, J., Nicholls, T.J., Baruffini, E., Walther, A., Danhauser, K., Zimmermann, F.A., et al. (2013). ELAC2 mutations cause a mitochondrial RNA processing defect associated with hypertrophic cardiomyopathy. *Am. J. Hum. Genet.* 93, 211–223.
- Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. *Cell* 169, 693–707.e14.
- Haering, C.H., Farcas, A.-M., Arumugam, P., Metson, J., and Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. *Nature* 454, 297–301.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Hinshaw, S.M., Makrantonis, V., Kerr, A., Marston, A.L., and Harrison, S.C. (2015). Structural evidence for Scc4-dependent localization of cohesin loading. *eLife* 4, e06057.
- Izumi, K., Akiyama, K., Fujiki, K., Masuda, K., Nakato, R., Otsubo, A., Bando, M., and Shirahige, K. (2019). Dissecting the functions of NIPBL using genome

editing: the importance of the N-terminus of NIPBL in transcriptional regulation. *bioRxiv*. <https://doi.org/10.1101/616086>.

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., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.

Kaiser, F.J., Ansari, M., Braunholz, D., Concepción Gil-Rodríguez, M., Decroos, C., Wilde, J.J., Fincher, C.T., Kaur, M., Bando, M., Amor, D.J., et al.; Care4Rare Canada Consortium; University of Washington Center for Mendelian Genomics (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.

Kaur, M., Mehta, D., Noon, S.E., Deardorff, M.A., Zhang, Z., and Krantz, I.D. (2016). NIPBL expression levels in CdLS probands as a predictor of mutation type and phenotypic severity. *Am. J. Med. Genet. C. Semin. Med. Genet.* 172, 163–170.

Kikuchi, S., Borek, D.M., Otwinowski, Z., Tomchick, D.R., and Yu, H. (2016). Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proc. Natl. Acad. Sci. U S A* 113, 12444–12449.

Kline, A.D., Krantz, I.D., Sommer, A., Kliever, M., Jackson, L.G., FitzPatrick, D.R., Levin, A.V., and Selicorni, A. (2007). Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. *Am. J. Med. Genet. A* 143A, 1287–1296.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47 (W1), W199–W205.

Lindeboom, R.G.H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* 48, 1112–1118.

Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M.A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., et al. (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol.* 7, e1000119.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC(T)</sup> Method. *Methods* 25, 402–408.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.

Mannini, L., Cucco, F., Quarantotti, V., Krantz, I.D., and Musio, A. (2013). Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. *Hum. Mutat.* 34, 1589–1596.

Mills, J.A., Herrera, P.S., Kaur, M., Leo, L., McEldrew, D., Tintos-Hernandez, J.A., Rajagopalan, R., Gagne, A., Zhang, Z., Ortiz-Gonzalez, X.R., and Krantz, I.D. (2018). NIPBL<sup>+/-</sup> haploinsufficiency reveals a constellation of transcriptome disruptions in the pluripotent and cardiac states. *Sci. Rep.* 8, 1056.

Murayama, Y., and Uhlmann, F. (2014). Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* 505, 367–371.

Nizon, M., Henry, M., Michot, C., Baumann, C., Bazin, A., Bessières, B., Blesson, S., Cordier-Alex, M.-P., David, A., Delahaye-Duriez, A., et al. (2016). A series of 38 novel germline and somatic mutations of NIPBL in Cornelia de Lange syndrome. *Clin. Genet.* 89, 584–589.

Oliveira, J., Dias, C., Redeker, E., Costa, E., Silva, J., Reis Lima, M., den Dunnen, J.T., and Santos, R. (2010). Development of NIPBL locus-specific data-

base using LOVD: from novel mutations to further genotype-phenotype correlations in Cornelia de Lange Syndrome. *Hum. Mutat.* 31, 1216–1222.

Parenti, I., Gervasini, C., Pozojevic, J., Graul-Neumann, L., Azzollini, J., Braunholz, D., Watrin, E., Wendt, K.S., Cereda, A., Cittaro, D., et al. (2016a). Broadening of cohesinopathies: exome sequencing identifies mutations in ANKRD11 in two patients with Cornelia de Lange-overlapping phenotype. *Clin. Genet.* 89, 74–81.

Parenti, I., Gervasini, C., Pozojevic, J., Wendt, K.S., Watrin, E., Azzollini, J., Braunholz, D., Buiting, K., Cereda, A., Engels, H., et al. (2016b). Expanding the clinical spectrum of the 'HDAC8-phenotype' - implications for molecular diagnostics, counseling and risk prediction. *Clin. Genet.* 89, 564–573.

Parenti, I., Teresa-Rodrigo, M.E., Pozojevic, J., Ruiz Gil, S., Bader, I., Braunholz, D., Bramswig, N.C., Gervasini, C., Larizza, L., Pfeiffer, L., et al. (2017). Mutations in chromatin regulators functionally link Cornelia de Lange syndrome and clinically overlapping phenotypes. *Hum. Genet.* 136, 307–320.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308.

Schmitz, J., Watrin, E., Lénárt, P., Mechtler, K., and Peters, J.-M. (2007). Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr. Biol.* 17, 630–636.

Seitan, V.C., Banks, P., Laval, S., Majid, N.A., Dorsett, D., Rana, A., Smith, J., Bateman, A., Krpic, S., Hostert, A., et al. (2006). Metazoan Scc4 homologs link sister chromatid cohesion to cell and axon migration guidance. *PLoS Biol.* 4, e242.

Selicorni, A., Russo, S., Gervasini, C., Castronovo, P., Milani, D., Cavalleri, F., Bentivegna, A., Masciadri, M., Domi, A., Divizia, M.T., et al. (2007). Clinical score of 62 Italian patients with Cornelia de Lange syndrome and correlations with the presence and type of NIPBL mutation. *Clin. Genet.* 72, 98–108.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15, 284.

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., et al. (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.

van den Berg, D.L.C., Azzarelli, R., Oishi, K., Martynoga, B., Urbán, N., Dekkers, D.H.W., Demmers, J.A., and Guillemot, F. (2017). Nipbl interacts with Zfp609 and the integrator complex to regulate cortical neuron migration. *Neuron* 93, 348–361.

Vrouwe, M.G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B.C., Bhuiyan, Z.A., Redeker, E.J., Mannens, M.M., Mullenders, L.H.F., Pastink, A., and Darroudi, F. (2007). Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. *Hum. Mol. Genet.* 16, 1478–1487.

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K., and Peters, J.-M. (2006). Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. *Curr. Biol.* 16, 863–874.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCTC-binding factor. *Nature* 451, 796–801.

Woodman, J., Fara, T., Dzieciatkowska, M., Trejo, M., Luong, N., Hansen, K.C., and Megee, P.C. (2014). Cell cycle-specific cleavage of Scc2 regulates its cohesin deposition activity. *Proc. Natl. Acad. Sci. U S A* 111, 7060–7065.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383.

Yuan, B., Pehlivan, D., Karaca, E., Patel, N., Chang, W.-L., Gambin, T., Gonzaga-Jauregui, C., Sutton, V.R., Yesil, G., Bozdogan, S.T., et al. (2015). Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. *J. Clin. Invest.* **125**, 636–651.

Zhu, Z., and Wang, X. (2019). Roles of cohesin in chromosome architecture and gene expression. *Semin. Cell Dev. Biol.* **90**, 187–193.

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., and Wendt, K.S. (2014). A cohesin-independent role for NIPBL at promoters provides insights in CdLS. *PLoS Genet.* **10**, e1004153.

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., et al. (2017). Regulation of the cohesin-loading factor NIPBL: Role of the lncRNA NIPBL-AS1 and identification of a distal enhancer element. *PLoS Genet.* **13**, e1007137.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit anti-MAU2	Abcam	Cat#ab183033; RRID:AB_2783830
Rabbit anti-MAU2	Gift from Jan-Michael Peters	N/A
Rat anti-NIPBL	Absea	Cat#010702F01; RRID:AB_2235936
Mouse anti-NIPBL	Santa Cruz	Cat#sc-374625; RRID:AB_10989775
Rabbit anti-SMC1A	Abcam	Cat#ab21583; RRID:AB_2192477
Rabbit anti-RAD21	Cell Signaling	Cat#4321; RRID:AB_1904106
Mouse anti-tubulin	Sigma Aldrich	Cat#T5201; RRID:AB_609915
Mouse anti-GAL4 DBD	Santa Cruz	Cat#sc-510; RRID:AB_627655
Rabbit anti-NFkB	Santa Cruz	Cat#sc-372; RRID:AB_632037
Mouse anti-GAL4 AD	Clontech	Cat#630402
Goat anti-rabbit	Thermo Fisher Scientific	Cat#31460; RRID:AB_228341
Goat anti-mouse	Thermo Fisher Scientific	Cat#31340; RRID:AB_228339
Goat anti-rat	Thermo Fisher Scientific	Cat#31470; RRID:AB_228356
Rabbit anti-SMC3 for ChIP	Custom	N/A
Rabbit anti-NIPBL for ChIP	Bethyl Laboratories	A301-779A; RRID:AB_1211232
Rabbit anti-NIPBL for ChIP	Originally gift from Koichi Tanaka and Kim Nasmyth and but re-created by ABSEA (China) using the same epitope	N/A
<b>Biological Samples</b>		
Patient-derived fibroblasts	Clinical Genetics Unit, Service of Paediatrics, Hospital “Lozano Blesa” Medical School	N/A
Patient-derived LCLs	David Fitzpatrick, The University of Edinburgh, Edinburgh	N/A
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Puromycin	Thermo Fisher Scientific	Cat#A11138-03; CAS: 53-79-2
MG132	Sigma Aldrich	Cat#M7449; CAS: 133407-82-6
Ammonium chloride	Merck Millipore	Cat#101145; CAS: 12125-02-9
Bafilomycin	Sigma Aldrich	Cat#B1793; CAS: 88899-55-2
Formaldehyde	Sigma Aldrich	Cat#252549; CAS: 50-00-0
DSG	Thermo Fisher Scientific	Cat#20593; CAS: 79642-50-5
Nocodazole	Sigma Aldrich	M1404; CAS: 31430-18-9
<b>Critical Commercial Assays</b>		
BCA protein assay kit	Thermo Fisher Scientific	23225
ReliaPrep RNA Cell Miniprep system	Promega	Z6011
Superscript III Reverse Transcriptase	Thermo Fisher Scientific	18080093
qPCRBIO Probe Mix Hi-ROX assay	PCR Biosystems	PB20.26-05
TnT Quick coupled transcription/translation system	Promega	L1170
NEXTflex ChIP-seq kit	BioO Scientific	NOVA-5143-01
DNeasy Blood & Tissue kit	QIAGEN	69506
Lipofectamine RNAiMAX	Thermo Fisher	13778100
TruSeq DNA PCR-Free kit	Illumina	20015963

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
ChIP-seq raw data	This paper	GEO: GSE122299
RNA-seq raw data	This paper	SRA: PRJNA609330
Experimental Models: Cell Lines		
HT1080	Institut für Humangenetik, Essen, Germany	N/A
HEK293	University of Luebeck, Luebeck, Germany	N/A
Experimental Models: Organisms/Strains		
Yeasts strain AH109	Clontech	PT3247-1
Oligonucleotides		
Guide RNA		
NIPBL_ex2_gRNA1_top: CACCGATCCCCG CAAGAGTAGTAAT	This paper	N/A
NIPBL_ex2_gRNA1_bottom: AAACATTACT ACTCTTGCGGGGATC	This paper	N/A
NIPBL_ex2_gRNA2_top: CACCGTCCCCATT ACTACTCTTGCG	This paper	N/A
NIPBL_ex2_gRNA2_bottom: AAACCGCAAG AGTAGTAATGGGGAC	This paper	N/A
QPCR primer		
NIPBL_ex2_F: caccattccagaaattcagg	This paper	N/A
NIPBL_wt_R: caaggttatcttcaactctATG	This paper	N/A
NIPBL_mut_R: atactgagacatcatcattcATG	This paper	N/A
siRNA:		
Control non-targeting sense CGUACGCG GAAUACUUCGAtt	Watrin et al., 2006	N/A
Control non-targeting antisense UCGAAGU AUUCCGCGUACGtt	Watrin et al., 2006	N/A
NIPBL sense GCAUCGGUAUCAAGUCCAtt	Watrin et al., 2006	N/A
NIPBL antisense UGGGACUUGAUACCGAUGCtt	Watrin et al., 2006	N/A
MAU2 sense GCAUCGGUAUCAAGUCCAtt	Watrin et al., 2006	N/A
MAU2 antisense UGGGACUUGAUACCGAUGCtt	Watrin et al., 2006	N/A
Recombinant DNA		
pSpCas9(BB)-2A-Puro (PX459) V2.0	Addgene	#62988
pcDNA 3.1/myc-His B	Thermo Fisher Scientific	V80020
pCMV-BD	Agilent Technologies	#211342
pCMV-AD	Agilent Technologies	#211343
pGBKT7	Clontech	630489
pGADT7	Clontech	630442
Software and Algorithms		
FASTQC v0.11.2	Babraham Bioinformatics	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
Bowtie2 v2.3.3.1	Langmead and Salzberg, 2012	<a href="http://bowtie-bio.sourceforge.net/">http://bowtie-bio.sourceforge.net/</a>
Trimmomatic v0.32		<a href="http://www.usadellab.org/cms/index.php?page=trimmomatic">http://www.usadellab.org/cms/index.php?page=trimmomatic</a>
Picard v1.97	Broad Institute	<a href="http://broadinstitute.github.io/picard/">http://broadinstitute.github.io/picard/</a>
MACS2 v 0.1.1.20160309	(Feng et al., 2012)	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
HOMER v4.3	(Heinz et al., 2010)	<a href="http://homer.ucsd.edu/homer/">http://homer.ucsd.edu/homer/</a>
DeepTools v3.0.2	(Ramírez et al., 2014)	<a href="https://usegalaxy.eu/">https://usegalaxy.eu/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NGS plot v2.61	(Shen et al., 2014)	<a href="https://github.com/shenlab-sinai/ngsplot">https://github.com/shenlab-sinai/ngsplot</a>
ChIPseeker v1.14.2	(Yu et al., 2015)	( <a href="http://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html">http://www.bioconductor.org/packages/release/bioc/html/ChIPseeker.html</a> )
Bcl2fastq v2.17	Illumina	<a href="https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html">https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html</a>
Bwa 0.7.8	(Li and Durbin, 2009)	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
GATK Base recalibrator	Broad Institute	<a href="https://gatkforums.broadinstitute.org/gatk">https://gatkforums.broadinstitute.org/gatk</a>
WEB-based GENE SeT Analysis Toolkit vers 2019	(Liao et al., 2019)	<a href="http://www.webgestalt.org">http://www.webgestalt.org</a>

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kerstin S. Wendt ([k.wendt@erasmusmc.nl](mailto:k.wendt@erasmusmc.nl)).

**Materials availability**

All plasmids and cell lines generated in this study are available on request from the Lead Contact without restriction.

**Data and code availability**

The accession number for the ChIP-sequencing datasets reported in this paper is Gene Expression Omnibus (GEO): GSE122299. The accession number for the RNA-sequencing data reported in this paper is SRA archive: PRJNA609330.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**Human subjects**

Whole exome sequencing was performed on a selected group of 15 patients with a clinical diagnosis of CdLS, comprising five females and ten males. Age of patients ranged from 7 to 32 years. All procedures including patients were approved by the Ethical Committee of the University of Lübeck (approval number for human studies HL07-158) and were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individuals participants included in this study.

**Detailed characterization of the MAU2 patient**

The patient with the MAU2 mutation is currently 14 years old. This patient was born at 34 weeks of gestation after caesarean section from healthy and non-consanguineous parents. Intra-uterine growth retardation was detected at the 30th week of pregnancy. At birth, Apgar scores were 5 and 8 at 1 and 5 minutes, respectively. Birth weight was 1330 g (−2.46 SD), birth length was 39.5 cm (−2.42 SD) and the occipital frontal circumference was 27 cm (−3.17 SD). He showed early feeding difficulties and developed a generalized spasticity from the very first weeks of life, which severely affected his mobility and soon required the use of a wheelchair. At the age of two years he was diagnosed with Cornelia de Lange syndrome because of his facial features and the diagnosis was confirmed by different experts independently. His facial features were typical for the syndrome and included microcephaly, brachycephaly, low anterior and posterior hairline, arched eyebrows, synophrys, long eyelashes, ptosis, flat nasal bridge, long and smooth philtrum, thin lips with downturned corners of the mouth and highly arched palate. In addition, he presented with hypertelorism, myopia, low-set and posteriorly rotated ears, small feet, clinodactyly of the fifth finger, hirsutism and cryptorchidism. He developed very early gastroesophageal reflux disease. Magnetic Resonance Imaging revealed the existence of a thin corpus callosum, a mild ventriculomegaly and periventricular cysts. Last evaluation was performed at the age of 10 years. At this age, weight was 18.8 kg (−4.5 SD), height was 121 cm (−4.96 SD) and the occipital frontal circumference was 47 cm (−4.96 SD). He presented with severe intellectual disability and delayed speech and motor development. He pronounced his first words at the age of two years but had no further verbal development. Currently, he is still not able to sit unassisted or to walk. This patient was found to carry an in-frame deletion of 21 nucleotides in MAU2, resulting in the loss of seven amino acids: RefSeq NM\_015329, c.927\_947del, p.(Gln310\_Ala316del).

### Cell lines

HEK293 cells (female) were cultured in DMEM supplemented with 10% FBS and 1% Antibiotics and grown at 37°C with 4% CO<sub>2</sub>.

HT1080 cells (male) were cultured in DMEM supplemented with 10% FBS and 1% Antibiotics and grown at 37°C with 4% CO<sub>2</sub>.

Control fibroblast and patient's fibroblasts were cultured in DMEM supplemented with 10% FBS and 1% Antibiotics and grown at 37°C with 4% CO<sub>2</sub>.

Control LCLs and patient's LCL were cultured in RPMI supplemented with 10% FBS and 1% Antibiotics and grown at 37°C with 4% CO<sub>2</sub>.

## METHOD DETAILS

### CRISPR/Cas9

Genome editing was performed as described by Ran and colleagues in 2013 (Ran et al., 2013). In detail, two different 20 nucleotides guide sequences followed by a 5'-NGG PAM and complementary to exon 2 of *NIPBL* (Biomers, Ulm, Germany, primers available upon request) were inserted into the pSpCas9(BB)-2A-Puro (PX459) V2.0 vector (#62988, Addgene, Cambridge, MA, USA). HEK293 cells were transfected with the resulting plasmids with the TaKaRa Xfect Transfection Reagent (Clontech-Takara, Saint-Germain-en-Laye, France), following the manufacturer's instructions. 24 hours post-transfection, selection of positively transfected clones was performed for 48 hours with DMEM medium supplemented with 10% FBS, 1% Antibiotics and puromycin at a final concentration of 8 µg/ml (Thermo Fisher Scientific). After selection, single clones were picked and expanded.

### Protein isolation and quantification

Cells were lysed using RIPA buffer pH 7.6 (50 mM HEPES, 1 mM EDTA, 1% NP-40, 0.5M LiCl) and a protease inhibitor cocktail (Roche, Mannheim, Germany). Protein concentration was determined through the BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's protocol with the Tristar<sup>2</sup> Multimode Reader LB942 (Berthold Technologies, Bad Wildbad, Germany).

### RNA isolation, cDNA synthesis and Real-Time PCR

RNA extraction was performed with the ReliaPrep RNA Cell Miniprep System (Promega, Mannheim, Germany) according to the manufacturer's instructions. Subsequent treatment with DNase I (RNase-free, New England Biolabs, Frankfurt am Main, Germany) was carried out on all RNA samples in order to avoid genomic DNA contaminations.

The SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) was used to retro-transcribe 2 µg of RNA with random hexamers. cDNA synthesis was performed in two independent experiments for each sample.

The expression level of the transcripts of interest was assessed by the use of the Real-Time PCR qPCR BIO Probe Mix Hi-ROX assay (PCR Biosystems, London, UK). The investigation was run on the 7300 Real-Time PCR system (Thermo Fisher Scientific). The following TaqMan gene expression assays were used for the analysis: Hs01062386\_m1, Hs00209846\_m1 and Hs01122291\_m1 (Thermo Fisher Scientific). Based on efficiency experiments, the GAPDH or NADH genes were selected as endogenous normalizer and amplified with the TaqMan gene expression assay ID Hs02758991\_g1 and Hs00190020\_m1 (Thermo Fisher Scientific). Relative gene expression was determined using the  $\Delta\Delta C_t$  method (Livak and Schmittgen, 2001).

### Protein degradation pathways

Wild-type and CRISPR/Cas9 HEK293 cells were treated with the proteasome inhibitor MG132 for 4-8 hours (10 µM; Sigma Aldrich), with the lysosome inhibitor ammonium chloride for 3-6 hours (NH<sub>4</sub>Cl, 20 mM; Millipore, Darmstadt, Germany) or with the autophagy inhibitor Bafilomycin for 3-6 hours (100 nM, Sigma Aldrich).

### In vitro transcription-coupled translation

*In Vitro* Transcription-coupled Translation (IVTT) reactions were performed with the TnT® Quick Coupled Transcription/Translation System (Promega) starting from 500 ng of plasmid, following the manufacturer's instructions.

### Subcellular fractionation

HEK293 cells were collected off the plates and washed twice with ice-cold PBS. The final cell pellet was resuspended in extraction buffer (20 mM Tris pH 7.5, 100 mM sodium chloride, 5 mM magnesium chloride, 0.2% NP-40, 10% glycerol, 0.5 mM dithiothreitol) supplemented with tablets of protease inhibitors (Roche) and phosphate inhibitors (10 mM sodium fluoride, 20 mM  $\beta$ -glycerophosphate). Cells were lysed on ice by passage through a 26-gauge needle. Lysates were incubated for 10 minutes on ice and were then centrifuged (21,000 g, 15 minutes, 4°C) for collection of the soluble protein extract. The chromatin-containing pellet was washed six times with extraction buffer (8,000 g, 3 minutes, 4°C) and then directly resuspended in sample buffer.



### siRNA knockdown of MAU2 and NIPBL

The following siRNA oligos were used to perform knockdowns for MAU2 and NIPBL:

Control non-targeting

sense CGUACGCGGAUACUUCGAtt  
antisense UCGAAGUAUCCGCGUACGtt

NIPBL

sense GCAUCGGUAUCAAGUCCCAtt  
antisense UGGGACUUGAUACCGAUGCtt

MAU2

sense GCAUCGGUAUCAAGUCCCAtt  
antisense UGGGACUUGAUACCGAUGCtt

SiRNA knockdown was performed with Lipofectamine RNAiMAX using the reverse transfection protocol. Cells were harvested for analysis at 72 hours after transfection.

### ChIP-sequencing

Chromatin immunoprecipitation for SMC3 was performed as previously described (Wendt et al., 2008). In brief, cells at 70%–80% confluence were crosslinked with 1% formaldehyde for 10 minutes and quenched with 125 mM glycine. After washing with PBS, cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 1% SDS, 10 mM EDTA, 1 mM PMSF and Complete Protease Inhibitor (Roche)) and sonicated (Diagenode Bioruptor, Seraing, Belgium) to around 500 bp DNA fragments. Debris were removed by centrifugation and the lysate was diluted 1:4 with IP dilution buffer (20 mM Tris-HCl pH 8.0, 0.15 M NaCl, 2 mM EDTA, 1% TX-100, protease inhibitors) and precleared with Affi-Prep Protein A support beads (Bio-Rad). The respective antibodies were incubated with the lysate overnight at 4°C, followed by 2 hours incubation at 4°C with blocked protein A Affiprep beads (Bio-Rad) (blocking solution: 0.1 mg/ml BSA). Beads were washed with washing buffer I (20 mM Tris-HCl pH 8.0, 0.15 M NaCl, 2 mM EDTA, 1% TX-100, 0.1% SDS, 1 mM PMSF), washing buffer II (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 2 mM EDTA, 1% TX-100, 0.1% SDS, 1 mM PMSF), washing buffer III (10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate) and TE-buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Beads were eluted twice (25 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% SDS) for 20 minutes at 65°C. The eluates were treated with proteinase K and RNase for 1 hour at 37°C and decrosslinked at 65°C overnight. The samples were further purified by phenol-chloroform extraction and ethanol-precipitated. The pellet was dissolved in TE buffer.

For NIPBL ChIP-seq experiments two different protocols were applied. Initially, we used a previously published protocol employing only formaldehyde crosslinking (FA-xlink) (Zuin et al., 2014). Subsequently, we employed a protocol (DSG/FA-xlink) adapted from van den Berg et al. (van den Berg et al., 2017) that has been shown to allow efficient detection of weaker NIPBL binding sites. For this second protocol, before crosslinking with formaldehyde, cells underwent a protein-protein crosslinking step. In brief, cells were suspended in PBS and treated for 45 minutes at room temperature under rotation with 2 mM disuccinimidyl glutarate (DSG). After three washes with PBS, cells were crosslinked with formaldehyde as described above. Sonication and pull-down were performed as described above with modifications of the beads washing steps according to Kagey et al. (Kagey et al., 2010). Beads were washed once with the IP dilution buffer, once with 20mM Tris-HCl pH8, 500mM NaCl, 2mM EDTA, 0.1% SDS, 1% Triton X-100, once with 10mM Tris-HCl pH8, 250nM LiCl, 2mM EDTA, 1% NP40 and once with TE buffer containing 50 mM NaCl.

For sequencing, the DNA libraries were prepared using the NEXTflex ChIP-Seq kit (BioO Scientific, Austin, TX, USA). These libraries were sequenced according to the Illumina TruSeq v3 protocol on an Illumina HiSeq2500 sequencer (Illumina, San Diego, CA, USA). Single reads were generated of 50 base pairs in length. The quality of DNA sequence was investigated using FASTQC (version 0.11.2), and, when necessary trimmomatic (version 0.32) was used to remove low-quality reads and regions. Quality controlled sequence was aligned to Human genome (hg19) using bowtie (version 1.0.0), and samtools (version 0.1.19) was used to remove reads with mapping quality less than 30, and to keep only aligned reads. Duplicated reads were removed, after alignment, using Picard (version 1.97). For statistics of the ChIP-sequencing see Table S6. Peaks were called using MACS (macs 2) and Peaks were filtered using P value. UCSC tracks were generated after duplicate removal, using HOMER (version 4.3) and deeptools (version 3.0.2). Heatmaps were generated using Deeptools (version 3.0.2) and NGS plot (version 2.61). Peaks were annotated to specific regions using ChIPseeker (Bioconductor package version 1.14.2). The data are accessible at GSE122299.

### ChIP-qPCR

For ChIP-qPCR experiments the ChIP for IgG control, SMC3 and NIPBL (NIPBL #1 antibodies) were performed as described above for the SMC3 ChIP-seq experiment. qPCR analysis using Platinum taq (Thermo Fisher Scientific) was performed according to the manufacturer's instructions and analyzed using a CFX96 C1000 Thermal cycler (Bio-Rad) using the qPCR primers listed in Table S2.

### Precocious sister chromatid separation assays by chromosome spreading

Cells were treated for 90 minutes with 100 ng/ml of nocodazole. Cells were collected and resuspended in 1 mL of medium. 1.5 mL of tap water was added. 6 minutes later 7 mL of Carnoy fixative (3:1, methanol: glacial acetic acid). Cells were then spread on glass slide, dried, stained with Giemsa stain and mounted in Entellan. Mitotic chromosome spreads were observed using a light microscope (DM2000 Leica, France) with a 40x dry objective.

### Determination of the mitotic index

Cells were grown on coverslips to maximal 60% confluency, fixed and mounted with Vectashield DAPI. Cells in metaphase and telophase were identified according to their morphology. Between 500 to 900 cells were counted per cell line in three different replicates.

### Fluorescence *in situ* hybridization (FISH)

DNA FISH was performed as previously described (Schmitz et al., 2007), with the exception that cells were fixed as described above and processed for FISH after spreading on glass slides. Only pairs for which the dots could be clearly resolved were considered in the analysis. Microscopic image acquisitions were performed on a Zeiss Axio Imager M2 using 63X oil-immersion objective. Image analysis was performed using ImageJ software.

### Whole Genome Sequencing

Genomic DNA isolation was performed from blood with the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Genomic DNA quantity was assessed using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific). For library preparation, 1000 ng of genomic DNA were used together with the TruSeq DNA PCR-Free Kit (Illumina) following the manufacturer's recommended protocol. The genomic DNA was fragmented to an average length of 350 bp by sonication on a Covaris E220 instrument (Covaris Inc., Woburn, MA, USA). Library preparation was performed in an automated manner using the NGS Option B (Agilent Technologies, Santa Clara, CA, USA). We assessed the library quality and absence of primer dimers by running a Bioanalyzer DNA High Sensitivity chip. Library quantification was performed using qPCR together with the Kapa Library Quantification Illumina/Light Cyclers 480 (Roche). The validated libraries were pooled in equimolar quantities and sequenced via 150 bp paired-end on an Illumina HiSeq 4000 platform following Illumina's recommended protocol. Raw data were demultiplexed with the Illumina bcl2fastq 2.17 into individual fastq files. Reads were aligned using the mem algorithm of bwa 0.7.8 and aligned to the hg19 reference into which decoy sequences had been added, PAR regions had been masked and the mitochondrial DNA had been replaced with the rCRS (revised Cambridge Reference Sequence) to match MITOMAP and most publicly available resources for mtDNA variants. Base quality scores were recalibrated using GATK BaseRecalibrator with enlarged context size for SNVs and Indels with respect to default (4 and 8 base pairs instead of 2 and 3). Variant calling was then performed using first HaplotypeCaller, to produce an individual GVCf file, and then multisample calling was performed using CombineGVCfs and then GenotypeGVCfs which produced the actual variant calls. Variant qualities and filters were then assessed with VariantRecalibrator tool using the tracks from GATK bundle and variants from GnomAD. For *de novo* variants, the Genotype Refinement workflow was applied using GenotypePosteriors to which a PED file with the family relationships, and again, a GnomAD track with allele counts and frequencies were supplied. The two subsequent steps were VariantFiltration to exclude low quality genotypes and VariantAnnotator to annotate possible *de novo* in the final VCF file.

### Mammalian two-hybrid assay

A fragment of NIPBL containing amino acids 1–300 was inserted into the pCMV-BD expression plasmid (#211342, Agilent Technologies, Santa Clara, CA, USA). The full-length open reading frame of MAU2 was cloned into the pCMV-AD plasmid (#211343, Agilent Technologies). MAU2 mutant constructs containing the deletion identified in the patient of our cohort (c.927\_947del21; p.(Gln310\_Ala316del)) were generated by site-directed *in vitro* mutagenesis with the Quick Change Site-directed mutagenesis kit, according to the manufacturer's instructions (Agilent Technologies). HEK293 cells were transiently transfected in 24-well plates with FuGene-HD (Promega), according to the manufacturer's instructions. Each well was transfected with 250 ng of the pCMV-BD-NIPBL1-300aa, 250 ng of the pCMV-MAU2 wild-type or mutant constructs, 250 ng of the Firefly Luciferase reporter plasmid (Promega) and 2.5 ng of the pHRG-TK Renilla luciferase expression plasmid (Promega). Activity of Firefly and Renilla luciferases was measured with the Dual Luciferase Reporter Assay System (Promega) with the Tristar<sup>2</sup> Multimode Reader LB942 (Berthold Technologies). All measurements were performed in triplicate in at least three independent experiments. Relative luciferase activity, indicating the strength of the interaction, was determined as the triplicate average of the ratio between the Firefly and the Renilla luciferase activity.

### Yeast two-hybrid assay

The first 300 amino acids of NIPBL and the wild-type and mutant full-length MAU2 were cloned into the Matchmaker GAL4 Two-Hybrid System 3 (Clontech-Takara) pGBKT7 and pGADT7 plasmids, respectively, to obtain NIPBL–GAL4 BD or MAU2–GAL4 AD fusion proteins. Yeasts (AH109) were co-transformed with the NIPBL and MAU2 fusion proteins, according to the Matchmaker 3 manual. Growth selection assays were performed using SD agar plates lacking Trp, Ade, His and Leu to detect interacting transformants.

### **Gamma irradiation and $\gamma$ -H2AX foci count**

Patient fibroblasts seeded on a coverslip were exposed to 1 Gy  $\gamma$ -irradiation using a 300kV ceramic X-ray tube (RS320 X-ray machine, Xstrahl). Cells were fixed according to standard immunofluorescence staining protocol and stained with anti- $\gamma$ -H2AX antibodies (Invitrogen, 14-9865-80) and Alexa Fluor 488-conjugated goat anti-mouse secondary antibodies (Invitrogen, A11029). Images were acquired using an immunofluorescence microscope (Zeiss Imager Z.1). The average number of  $\gamma$ -H2AX foci for at least 50 cells per cell line per time point were counted using ImageJ/FIJI software.

### **RNA-sequencing**

Strand specific, polyA-enriched RNA sequencing was performed as described earlier ([Haack et al., 2013](#)). Briefly, RNA was isolated from whole-cell lysates using the AllPrep RNA Kit (QIAGEN) and RNA integrity number (RIN) was determined with the Agilent 2100 BioAnalyzer (RNA 6000 Nano Kit, Agilent). For library preparation, 1  $\mu$ g of RNA was poly(A) selected, fragmented, and reverse transcribed with the Elute, Prime, Fragment Mix (Illumina). A-tailing, adaptor ligation, and library enrichment were performed as described in the TruSeq Stranded mRNA Sample Prep Guide (Illumina). RNA libraries were assessed for quality and quantity with the Agilent 2100 BioAnalyzer and the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies). RNA libraries were sequenced as 150 bp paired-end runs on an Illumina HiSeq4000 platform. The STAR aligner ([Dobin et al., 2013](#)) (v 2.4.2a) with modified parameter settings (`--twopassMode = Basic`) was used for split-read alignment against the human genome assembly hg19 (GRCh37) and UCSC known-Gene annotation. To quantify the number of reads mapping to annotated genes we used HTseq-count ([Anders et al., 2015](#)) (v0.6.0). FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) values were calculated using custom scripts. Differential expression analysis was performed using the R Bioconductor package DESeq2 ([Love et al., 2014](#)).

For a statistic on the RNA-sequencing experiments see [Table S7](#).

### **Gene ontology term enrichment analysis**

Genes found to be differentially expressed ( $P$  value  $< 0.05$ ,  $|\text{Fold change}| > 2$ ) as compared to wild-type were considered in the analyses. Term enrichment analysis was performed using WEB-based GENE SeT AnaLysis Toolkit in its 2019 version ([Liao et al., 2019](#)). Over-Representation Analysis method was used on the Biological Processes functional database against genome protein-coding reference set, with default parameters). Redundancy in rendered term list was reduced by weighted set cover method.

### **QUANTIFICATION AND STATISTICAL ANALYSIS**

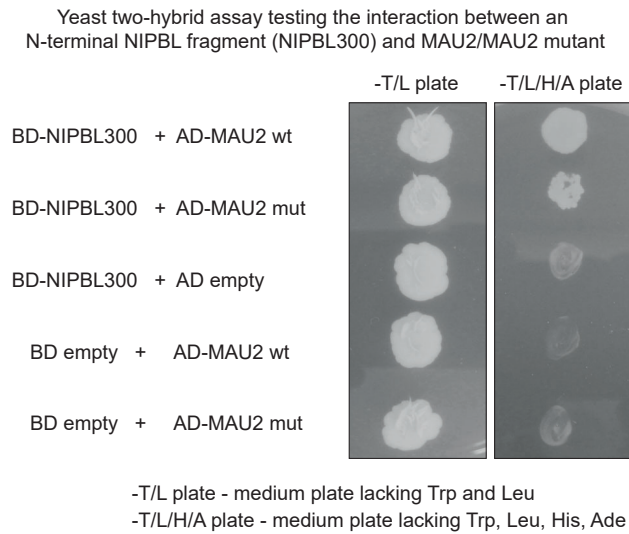
Real-Time PCR experiments were performed in triplicate for each sample; values with standard deviations exceeding 0.5% or standard errors exceeding 0.3% were excluded and the experiments were repeated. The bilateral unpaired  $t$  test was applied to compare gene expression data, and differences between experimental groups were considered significant when  $p < 0.05$ . Specific  $p$  values are stated in [Figure 3](#).

**Supplemental Information**

**MAU2 and NIPBL Variants Impair  
the Heterodimerization of the Cohesin Loader  
Subunits and Cause Cornelia de Lange Syndrome**

**Ilaria Parenti, Farah Diab, Sara Ruiz Gil, Eskeatnaf Mulugeta, Valentina Casa, Riccardo Berutti, Rutger W.W. Brouwer, Valerie Dupé, Juliane Eckhold, Elisabeth Graf, Beatriz Puisac, Feliciano Ramos, Thomas Schwarzmayr, Macarena Moronta Gines, Thomas van Staveren, Wilfred F.J. van IJcken, Tim M. Strom, Juan Pié, Erwan Watrin, Frank J. Kaiser, and Kerstin S. Wendt**





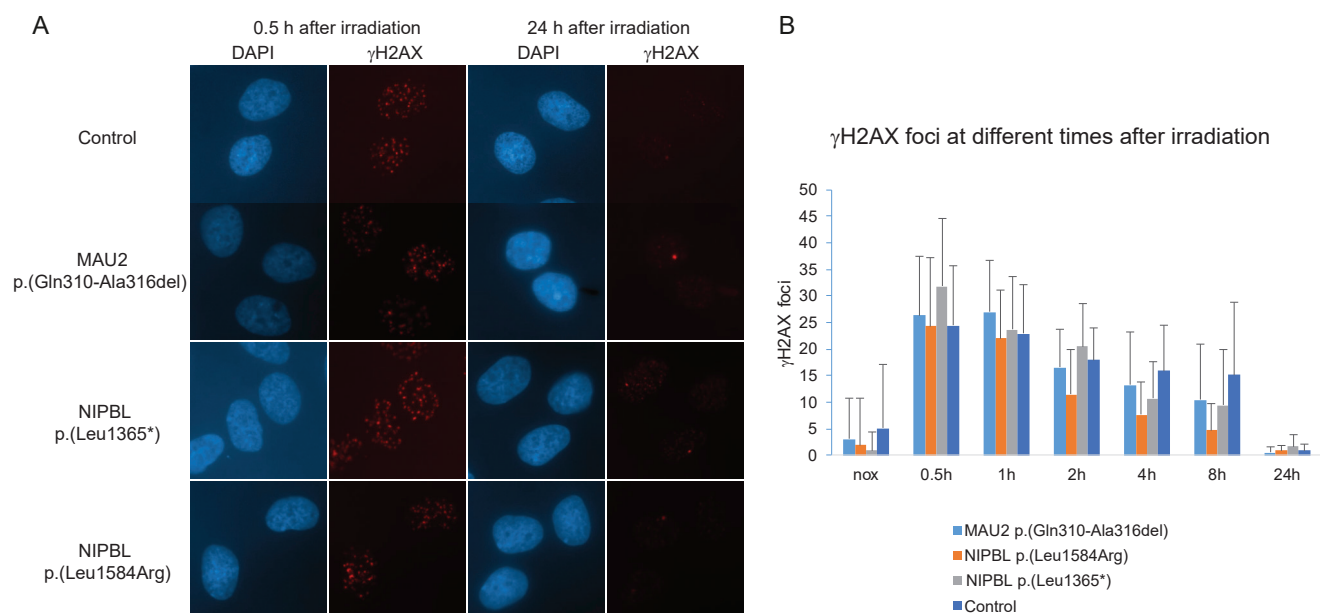
**Figure S1: Related to Figure 1.**

**The in-frame deletion p.(Gln310\_Ala316del) in MAU2 impairs its interaction with NIPBL**

The yeast two-hybrid assay shows a slower growth of yeasts on selective medium in the presence of the MAU2 deletion, supporting the results obtained with the mammalian two-hybrid assay.

The left panel shows interaction-independent growth of yeasts on Trp- and Leu-deficient plates.

The right panel illustrates the interaction-dependent growth of yeasts expressing NIPBL and MAU2 in plates lacking Trp, Leu, His and Ade.

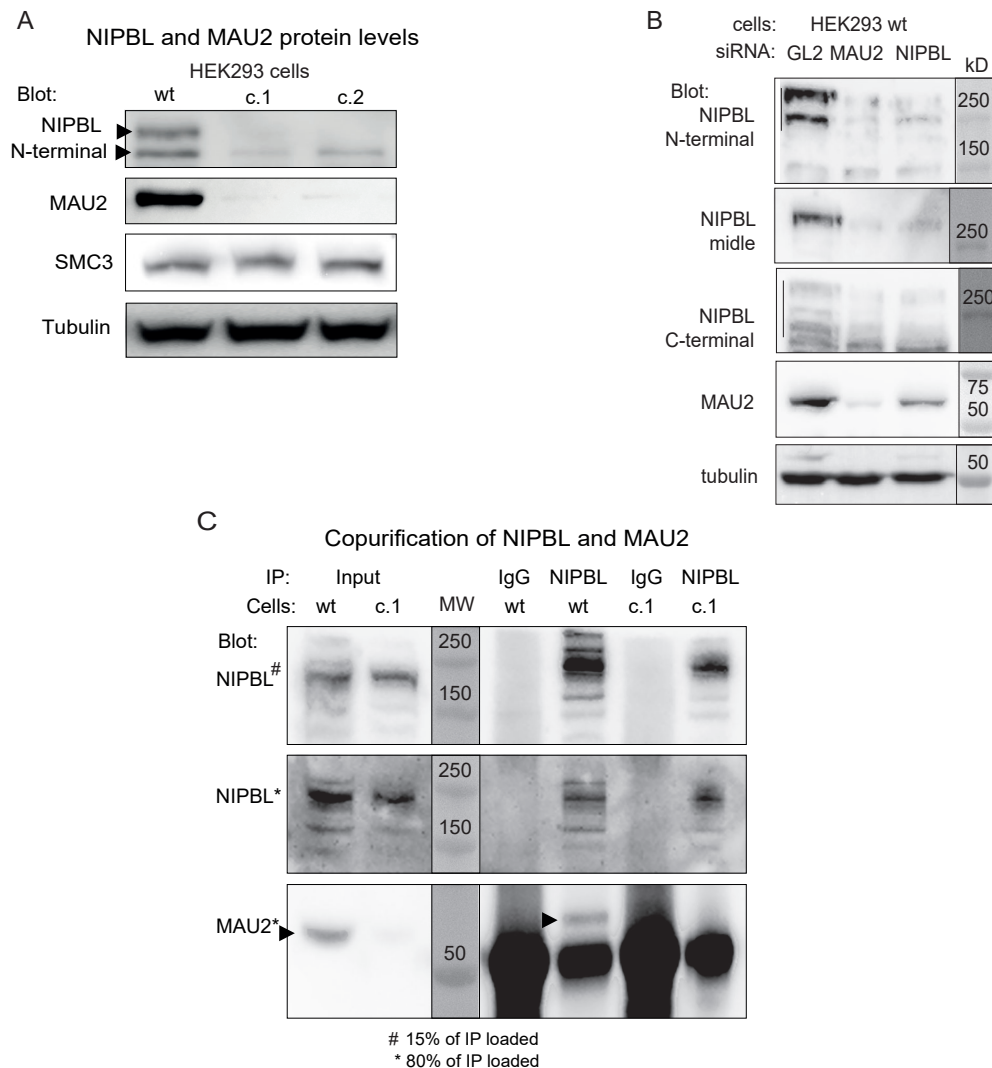


**Figure S2: Related to Figure 1.**

**Response of CdLS fibroblast cells to DNA damage**

A) Skin fibroblasts of a healthy control, from two patients with different NIPBL mutations (NIPBL p.Leu1365\*, NIPBL p.Leu1584Arg) and from the patient with deletion in MAU2 (MAU2 p.Gln310-Ala 316del) were exposed to 1 Gy gamma-irradiation. DNA double strand breaks (DSB's) were visualized by immunostaining for  $\gamma$ H2AX at different time points after exposure.

B) To assess DSB repair, the number of DSBs at different time points was quantitated as number of  $\gamma$ H2AX foci. Data are presented as means from n=3 experiments +/- SEM.



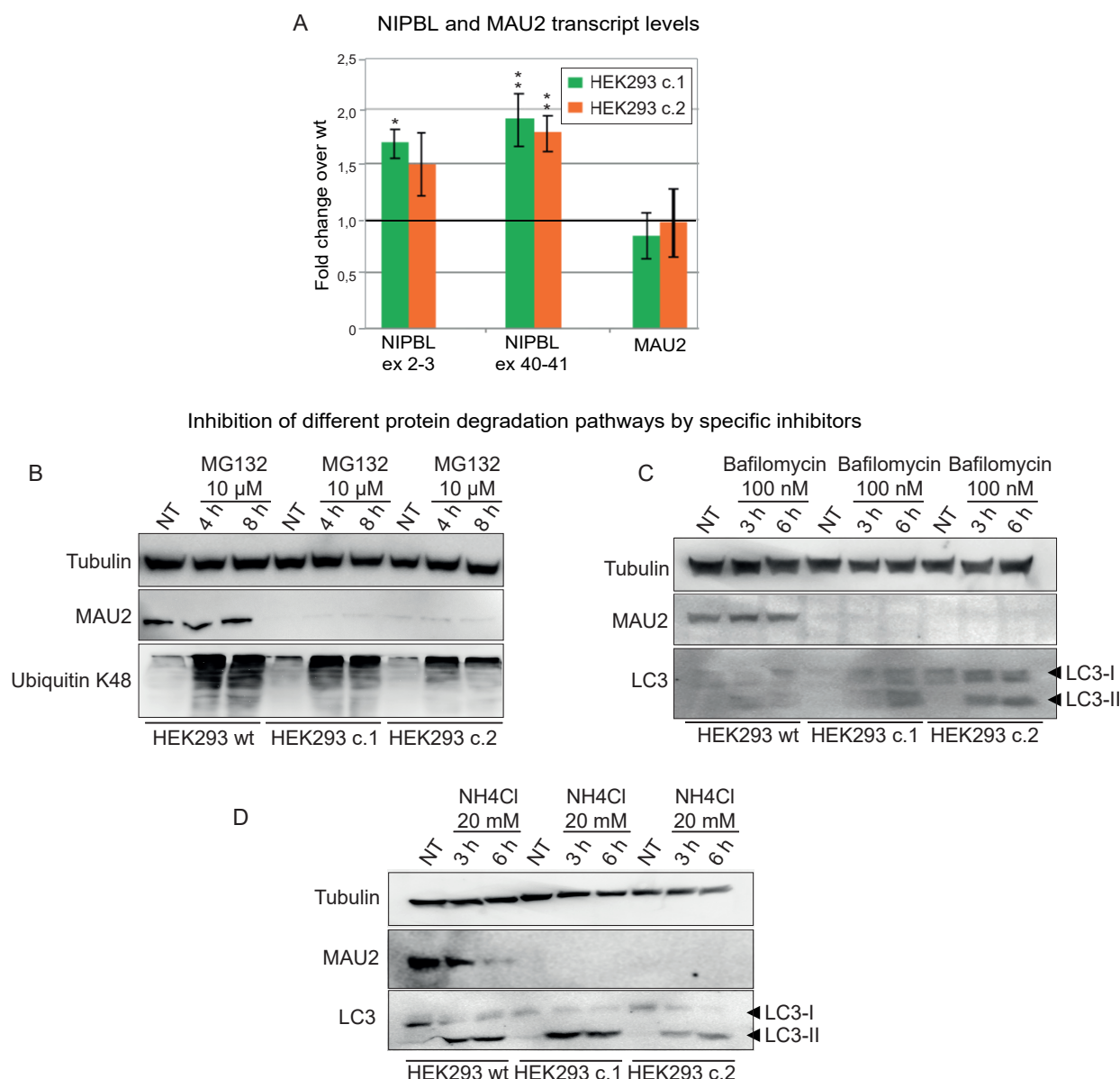
**Figure S3: Related to Figure 3.**

**NIPBL and MAU2 depletion by siRNA knockdown and quantitation of MAU2 protein levels in NIPBLΔN cells**

A) Western blots with total cell lysates from HEK293 NIPBLΔN cells, probed with antibodies against NIPBL, MAU2, SMC3 and tubulin.

B) Knockdown of MAU2 and NIPBL by siRNA in HEK293 cells. Total cell lysates were loaded and the blots probed with antibodies directed against the N-terminus and the C-terminus of NIPBL reveal which of the detected bands are NIPBL (marked with a line) and which are unspecific bands. Further, rabbit polyclonal antibodies raised against residues 744-977 from mouse Nipbl were used (NIPBL middle). Note that NIPBL depletion leads to a strong reduction of MAU2 protein levels and *vice versa*.

C) Immunoprecipitation (IP) of NIPBL from nuclear extracts of HEK293 wt and NIPBLΔN clone c.1 and detection of NIPBL and MAU2 by western blotting. NIPBL could be immunoprecipitated in wt and NIPBLΔN cells while MAU2 was only coprecipitated in wt cells, demonstrating that the residual MAU2 is unable to interact with NIPBLΔN. The strong signals in the anti-MAU2 immunoblot are caused by the IgGs used for immunoprecipitation. The arrowhead indicates MAU2. A similar blot for another HEK293 NIPBLΔN clone and for HT1080 cells is included in Figure 3D.



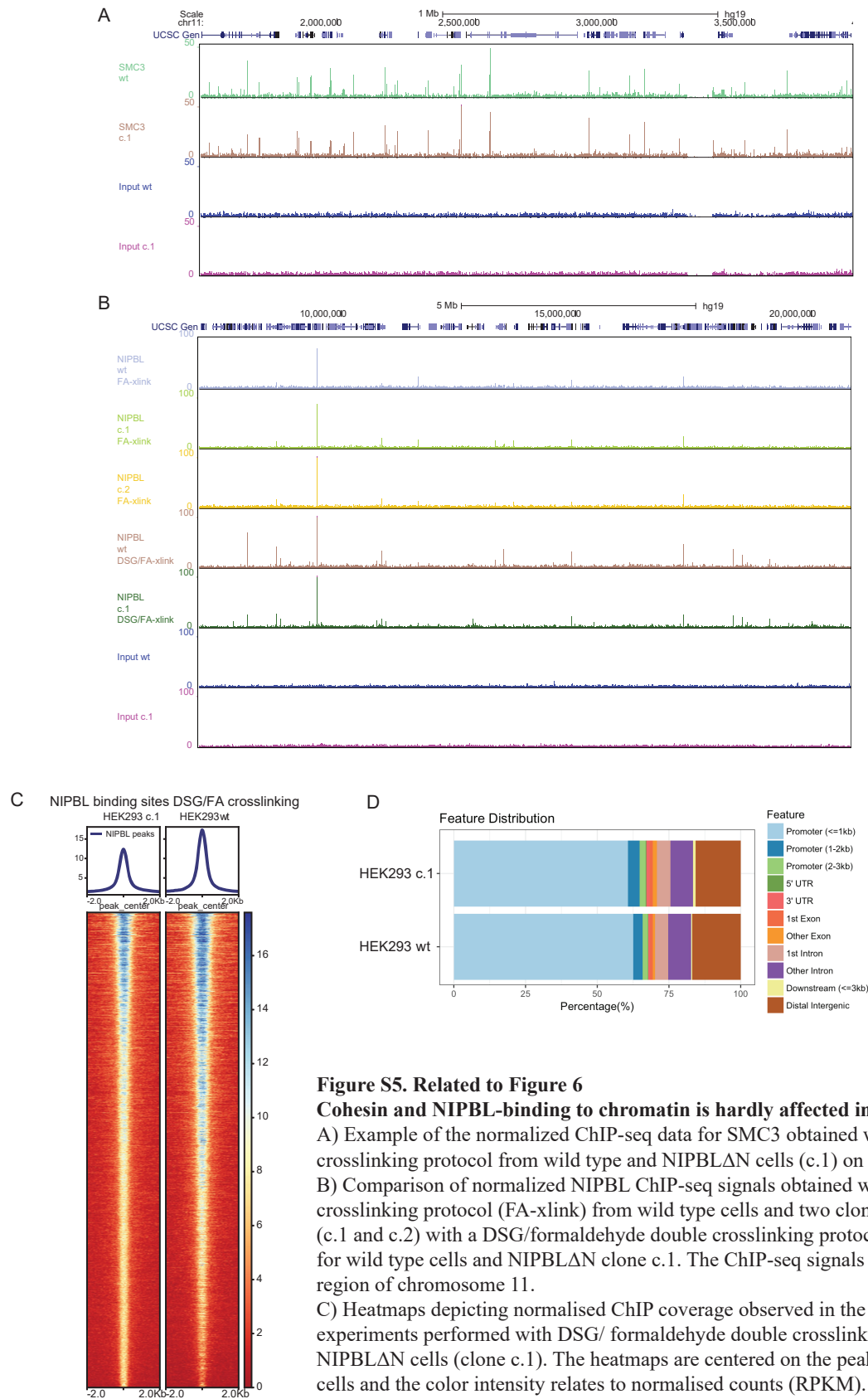
**Figure S4: Related to Figure 3.**

**MAU2 loss does not depend on the main cellular protein degradation pathways**

A) Analysis of NIPBL and MAU2 transcript levels. Expression data were normalized to GAPDH and relative mRNA levels compared to wt cells were determined using the  $\Delta\Delta C_t$  method. MAU2 expression appears to be largely unchanged in NIPBL $\Delta$ N cells (Bilateral unpaired T-test; c1  $p=0.09$ , c2  $p=0.85$ ), while two different intron-spanning primer pairs for NIPBL (ex 2-3 and ex 40-41) show increased NIPBL transcript levels (Bilateral unpaired T-test; exon 2-3: c1  $*p=0.03$ , c2  $p=0.06$ ; exon 40-41: c1  $**p=0.006$ , c2  $**p=0.001$ ).

B) MG132 treatment leads to the inhibition of the proteasome degradation pathway. Poly-ubiquitination of K48 was used as a positive control for the treatment (upon arrest of the proteasome with MG132, proteins with the ubiquitin degradation marker on K48 accumulate in the cell). Despite the success of the treatment with MG132, MAU2 expression could not be restored in the NIPBL $\Delta$ N cells, indicating that the proteasome is not responsible for MAU2 degradation.

C-D) Bafilomycin (inhibition of fusion between autophagosome and lysosome) (C) and Ammonium chloride (NH<sub>4</sub>Cl, alteration of lysosomal pH) (D) are both responsible for the arrest of the autophagy flux. The conversion of the soluble LC3-I to lipid-bound LC3-II is associated with the formation of autophagosomes. LC3-II associates with the membrane of the autophagosomes and is degraded after fusion of the autophagosome with the lysosome. When cells are treated with Ammonium chloride or Bafilomycin, the degradation of LC3-II is blocked, resulting in its accumulation. Upon treatment of the NIPBL $\Delta$ N cells with these two agents, LC3-II levels increased, whereas MAU2 expression could not be restored. Altogether, these results indicate that the autophagy flux is not responsible for the loss of MAU2.



**Figure S5. Related to Figure 6**

**Cohesin and NIPBL-binding to chromatin is hardly affected in NIPBLΔN cells**

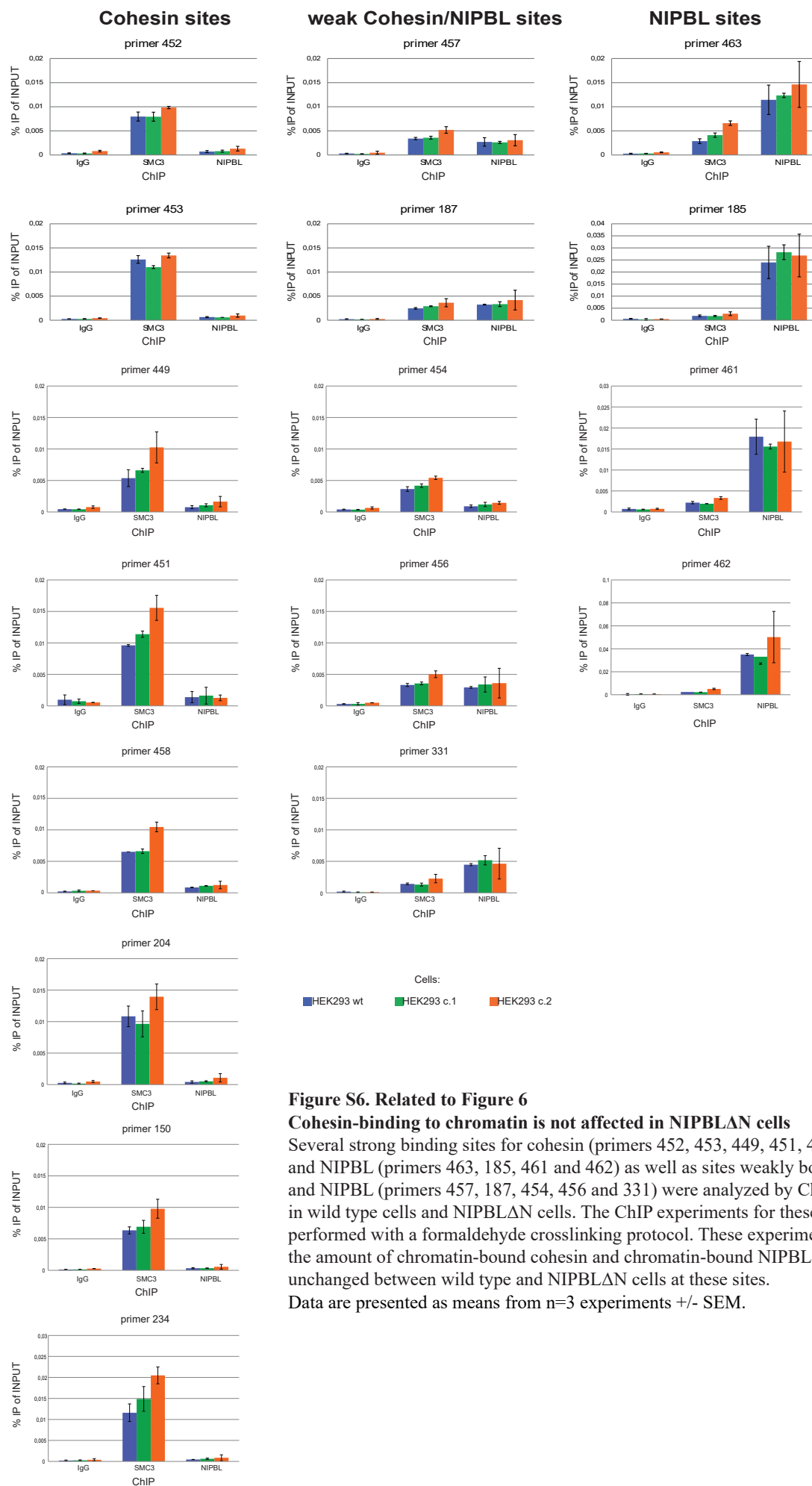
A) Example of the normalized ChIP-seq data for SMC3 obtained with a formaldehyde-crosslinking protocol from wild type and NIPBLΔN cells (c.1) on a region of chr11.

B) Comparison of normalized NIPBL ChIP-seq signals obtained with a formaldehyde-crosslinking protocol (FA-xlink) from wild type cells and two clones of NIPBLΔN cells (c.1 and c.2) with a DSG/formaldehyde double crosslinking protocol (DSG/FA-xlink) for wild type cells and NIPBLΔN clone c.1. The ChIP-seq signals displayed represent a region of chromosome 11.

C) Heatmaps depicting normalised ChIP coverage observed in the NIPBL ChIP-seq experiments performed with DSG/ formaldehyde double crosslinking for wild type and NIPBLΔN cells (clone c.1). The heatmaps are centered on the peaks observed in wild type cells and the color intensity relates to normalised counts (RPKM).

D) Distribution of the peaks observed in the SMC3 and NIPBL ChIP-seq experiments performed with a DSG/formaldehyde double crosslinking protocol (DSG/FA-xlink) over different genomic features (eg. promoters, introns, exons) for wild type and NIPBLΔN cells. No striking differences are visible between wild type and NIPBLΔN cells.



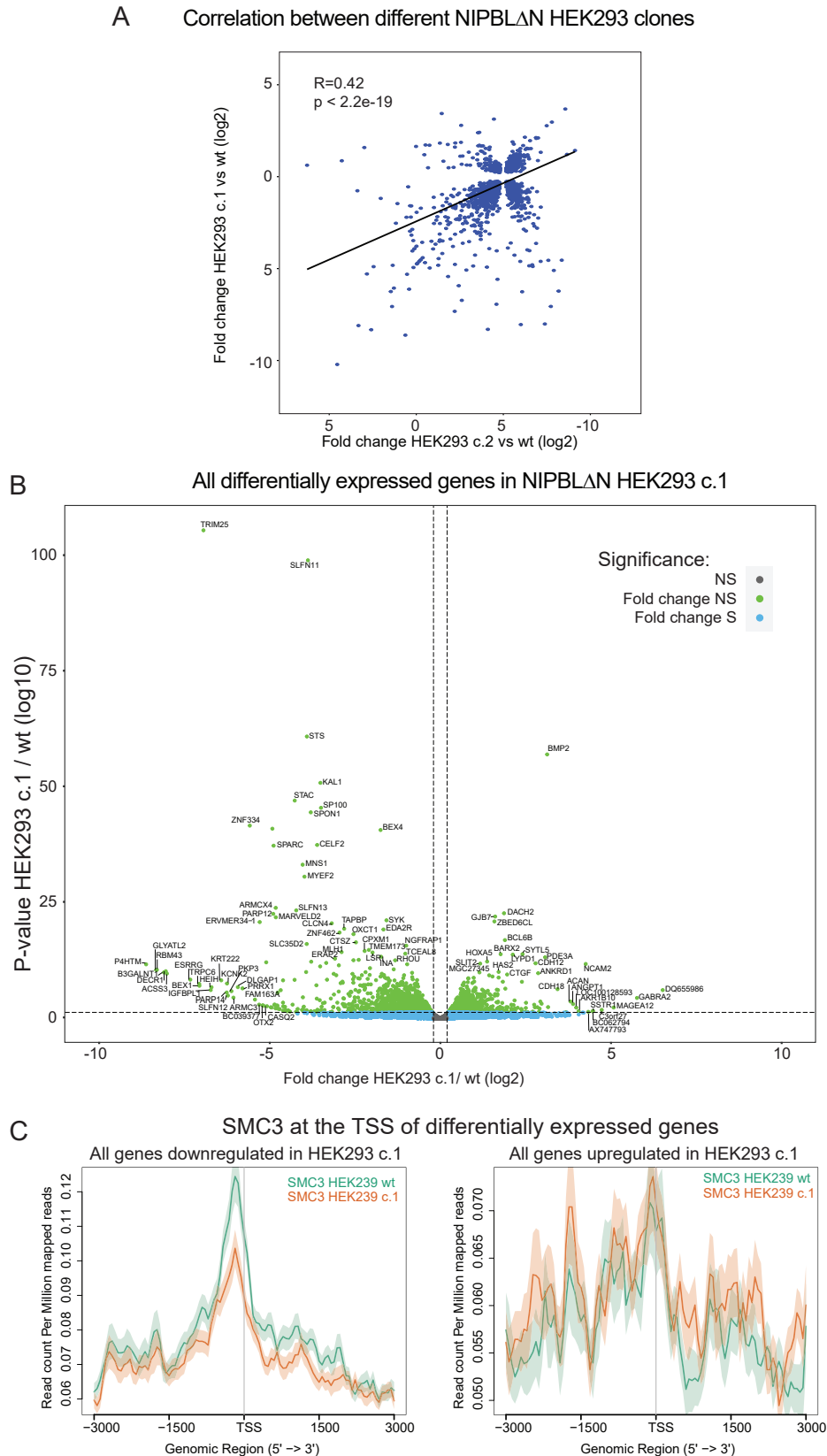


**Figure S6. Related to Figure 6**

**Cohesin-binding to chromatin is not affected in NIPBLΔN cells**

Several strong binding sites for cohesin (primers 452, 453, 449, 451, 458, 204, 150, 234) and NIPBL (primers 463, 185, 461 and 462) as well as sites weakly bound by cohesin and NIPBL (primers 457, 187, 454, 456 and 331) were analyzed by ChIP-qPCR in wild type cells and NIPBLΔN cells. The ChIP experiments for these samples were performed with a formaldehyde crosslinking protocol. These experiments indicate that the amount of chromatin-bound cohesin and chromatin-bound NIPBL is mainly unchanged between wild type and NIPBLΔN cells at these sites.

Data are presented as means from n=3 experiments +/- SEM.



**Figure S7. Related to Figure 6. Transcriptome analysis of NIPBLΔN HEK293 cells.**

A) Scatter plot showing the correlation of the fold changes between two different HEK293 cell clones (HEK293 c.1 on the x-axis and HEK293 c.2 on the y-axis). The Pearson correlation coefficient (R) and the p-value (p) are shown in the graph.

B) Volcano plot showing the distribution of up and downregulated genes in H293 cells (HEK293c.1).

Genes with p-values < 0.05 are shown as blue dots, genes with a  $|FC| < 1$  are shown as grey dots.

The green dots show genes with significant FC and P-values. Note that we show here the differentially expressed genes for the HEK293 c.1 clone. Genes that show inconsistent changes between the two analyzed clones c.1 and c.2 are not removed.

C) Plots showing the SMC3 coverage at promoters of genes downregulated or upregulated in NIPBLΔN cells

Table S1:

GO-term analysis of differentially expressed genes in MAU2 p.(Gln310-Ala316del) fibroblasts and NIPBL p.(Leu1365\*) fibroblasts versus a healthy control, related to Figure 2

Analysis with Webgestalt (<http://www.webgestalt.org/>), genes with p-value <0.05 and FC outside [-1;+1] were included. The options term redundancy reduced by weighted set cover were used. Overlapping |GO-terms are highlighted in red.

#### MAU2p.(Gln310-Ala316del) fibroblasts versus healthy control fibroblasts

Gene Set	Description	Size	Expect	Ratio	P Value	FDR
GO:0008283	cell proliferation	1986	383.28	1.4532	0	0
GO:0006928	movement of cell or subcellular component	1967	379.61	1.5700	0	0
GO:0009888	tissue development	1925	371.51	1.5854	0	0
GO:2000026	regulation of multicellular organismal development	1908	368.23	1.6077	0	0
GO:0046903	secretion	1605	309.75	1.4237	0	0
GO:0048585	negative regulation of response to stimulus	1545	298.17	1.4723	0	0
GO:0022610	biological adhesion	1377	265.75	1.7611	0	0
GO:1902531	regulation of intracellular signal transduction	1824	352.02	1.3721	1.1102e-15	1.2616e-13
GO:0006468	protein phosphorylation	1860	358.96	1.3567	7.5495e-15	7.7115e-13
GO:0006811	ion transport	1608	310.33	1.3856	1.1102e-14	1.0853e-12

#### NIPBL p.L1365\* fibroblasts versus healthy control fibroblasts

Gene Set	Description	Size	Expect	Ratio	P Value	FDR
GO:0006928	movement of cell or subcellular component	1967	165.37	1.6327	0	0
GO:0009888	tissue development	1925	161.84	1.7363	0	0
GO:2000026	regulation of multicellular organismal development	1908	160.41	1.8328	0	0
GO:0022610	biological adhesion	1377	115.77	1.9954	0	0
GO:0007267	cell-cell signaling	1575	132.42	1.6916	3.3307e-16	1.4419e-13
GO:0008283	cell proliferation	1986	166.97	1.5991	4.4409e-16	1.8351e-13
GO:0010647	positive regulation of cell communication	1732	145.62	1.6276	3.7748e-15	1.2256e-12
GO:0046903	secretion	1605	134.94	1.6007	4.6496e-13	8.4539e-11
GO:0051049	regulation of transport	1765	148.39	1.5567	1.1726e-12	1.9382e-10
GO:1901700	response to oxygen-containing compound	1556	130.82	1.5747	8.6525e-12	1.1918e-9

**Table S2 - related to Figure S5**

**qPCR primer for the analysis of ChIP samples**

<b>primer</b>	<b>forward</b>	<b>reverse</b>	<b>genomic coordinates (hg19)</b>
150	GGCTCAGGACAGAAGTGACC	AGGTCGGCAGAGGCTCC	chr22:32925080-32925266
204	TTCAGCCGGTTCAAGGGACG	CTAGGGAGGAGGACAGAGGCAAGAG	chr11:2058315-2058501
234	AATGACGGAGCTGCGAGATAC	TTGCTAGGCAACCTGACAGAC	chr5:43313810-43313992
331	AGGTGGCACTGTTGCATAG	CCTGCCAAAGGAGAGCTTTATC	chr11:1630677-1630849
449	AGATGAAGCACAGGTGGTGAAC	CAACACCGACTGCTTGAAATCC	chr11:20359017-20359262
451	CCTGGCAATTTACGGTGTAGG	ATAAGCCGCACTAGAAGGACCC	chr22:37252291-37252540
452	TTCGGAAGAGGCAGGGAAG	GTGGCAGCACCCAAGTTAG	chr22:38259504-38259680
454	GAGTCTTTCCTTGCCTCTTC	GGGAGCCCATAAATGTCAC	chr5:36935893-36936139
456	TTCCCAGAGCCCAGAATACAAG	GTAGCGCGAAATAAGGATTCCG	chr19:4246755-4246932
457	GACAGGCCCACGTTATTGAC	ACCCTCCAGCAAGACAGAAG	chr19:7,694,524-7,694,677
458	CCGACCACAAGGACGGATTC	CCGCTCTCAGCTTCCTCTTC	chr19:8008922-8009140
461	CCGCCCACCTAACCTTTATTC	AGTTGCGAGGCAGAGTTTG	chr9:130829458-130830053
462	AAAGCAGAGGGCGACTTAGG	TCAGGAAGCTGTAGCGAGAG	chr9:139780556-139781018
463	GCCGAGCAGCATTGACCAAC	TCTCGGCTGGGCTCTACTTC	chr3:88107986-88108569

**Table S5:** GO-term analysis of differentially expressed genes in NIPBLΔN HEK293, related to Figure 6 and Figure S7

Biological pathways from DEGs in HEK (common DEGs,  $|FC| > 1$ ,  $p < 0.05$ ), reduction of redundancy by weighted set cover (<http://www.webgestalt.org/>).

Gene Set	Description	Size	Expect	Ratio	P Value	FDR
GO:0051240	positive regulation of multicellular organismal process	1661	19.736	2.2801	9.5321e-8	0.00067473
GO:0040011	locomotion	1738	20.651	2.0822	0.0000022940	0.0013903
GO:0040007	growth	961	11.419	2.5397	0.0000034190	0.0016660
GO:0007610	behavior	582	6.9153	3.0368	0.0000061493	0.0021925
GO:2000026	regulation of multicellular organismal development	1908	22.671	1.9408	0.000010588	0.0029167
GO:0032989	cellular component morphogenesis	1082	12.856	2.3335	0.000012112	0.0029760
GO:0042493	response to drug	994	11.811	2.3707	0.000018391	0.0035138
GO:0043062	extracellular structure organization	400	4.7528	3.3665	0.000024087	0.0042936
GO:0007267	cell-cell signaling	1575	18.714	1.9771	0.000042091	0.0061043
GO:0009611	response to wounding	642	7.6282	2.6219	0.000082512	0.0092607



**Table 6    ChIP-sequencing statistics, related to STAR methods**

Sample ID	Sample	Cell line	sequenced reads	genome
I17-1066-01	SMC3 ChIP	HEK293 wt	26066918	hg19
I17-1066-02	NIPBL ChIP	HEK293 wt	23349918	hg19
I17-1066-04	Input	HEK293 wt	20655989	hg19
I17-1066-05	SMC3 ChIP	HEK293 c.1	20269237	hg19
I17-1066-06	NIPBL ChIP	HEK293 c.1	20505412	hg19
I17-1066-08	Input	HEK293 c.1	24265890	hg19
I17-1284-01	SMC3 ChIP	HEK293 wt	24250592	hg19
I17-1284-02	SMC3 ChIP	HEK293 c.1	25637885	hg19
I17-1310-01	SMC3 ChIP	HEK293 wt	26517943	hg19
I17-1310-02	SMC3 ChIP	HEK293 wt	30302933	hg19
I17-1310-03	Input	HEK293 wt	27249816	hg19
I17-1310-04	SMC3 ChIP	HEK293 c.1	30785668	hg19
I17-1310-05	NIPBL ChIP	HEK293 c.1	32568589	hg19
I17-1310-06	Input	HEK293 c.1	25557021	hg19
I17-1310-07	SMC3 ChIP	HEK293 c.2	30369459	hg19
I17-1310-08	NIPBL ChIP	HEK293 c.2	29778688	hg19
I17-1310-09	Input	HEK293 c.2	23724637	hg19
I17-1066-03	NIPBL ChIP (Beth	HEK293 wt	24539800	hg19
I17-1066-07	NIPBL ChIP (Beth	HEK293 c.1	23793483	hg19

**Table S7 RNA-sequencing statistics, related to STAR methods**

Sample ID	Cell line	sequenced reads	genome	mapped reads		unmapped reads	
				read number	percent	read number	percent
LUEB0120	Control fibroblasts	99439773	hg19	97203044	97,75%	2236729	2,25%
LUEB0129	Control fibroblasts	81304856	hg19	78691729	96,79%	2613127	3,21%
LUEB0130	Control fibroblasts	74240231	hg19	72118457	97,14%	2121774	2,86%
LUEB0123	NIPBL mut fibroblasts	77501979	hg19	75245221	97,09%	2256758	2,91%
LUEB0124	NIPBL mut fibroblasts	81246927	hg19	78853528	97,05%	2393399	2,95%
LUEB0125	NIPBL mut fibroblasts	75030331	hg19	73602333	98,10%	1427998	1,90%
LUEB0126	MAU2 mut fibroblasts	78344410	hg19	76627687	97,81%	1716723	2,19%
LUEB0127	MAU2 mut fibroblasts	97053241	hg19	95405627	98,30%	1647614	1,70%
LUEB0128	MAU2 mut fibroblasts	77196493	hg19	75583417	97,91%	1613076	2,09%
LUEB0157	Control HEK293 cells	111677363	hg19	109161179	97,75%	2516184	2,25%
LUEB0158	Control HEK293 cells	99089929	hg19	95208004	96,08%	3881925	3,92%
LUEB0159	Control HEK293 cells	98830348	hg19	96576515	97,72%	2253833	2,28%
LUEB0160	CRISPR/Cas9 HEK293 c.10 cells	97689690	hg19	93721355	95,94%	3968335	4,06%
LUEB0161	CRISPR/Cas9 HEK293 c.10 cells	93990390	hg19	88485178	94,14%	5505212	5,86%
LUEB0162	CRISPR/Cas9 HEK293 c.10 cells	92127803	hg19	87747608	95,25%	4380195	4,75%
LUEB0163	CRISPR/Cas9 HEK293 c.2.1 cells	100406239	hg19	94761904	94,38%	5644335	5,62%
LUEB0164	CRISPR/Cas9 HEK293 c.2.1 cells	107339878	hg19	98971476	92,20%	8368402	7,80%
LUEB0165	CRISPR/Cas9 HEK293 c.2.1 cells	100408584	hg19	95803500	95,41%	4605084	4,59%