# Epigenetic regulation of plastin 3 expression by the macrosatellite *DXZ4* and the transcriptional regulator CHD4


## Authors

Eike A. Strathmann, Irmgard Hölker, Nikolai Tschernoster, ..., Cecile Martinat, Janine Altmüller, Brunhilde Wirth

## Correspondence

brunhilde.wirth@uk-koeln.de



**Dysregulated PLS3 levels associate with various genetic diseases, PLS3 upregulation protects against spinal muscular atrophy and is a biomarker of cancer. A high copy number of *DXZ4* macrosatellite facilitates the escape of *PLS3* from X-inactivation in women. Moreover, independent of sex, *PLS3* expression is epigenetically regulated by the CHD4/NuRD complex.**



Strathmann et al., 2023, The American Journal of Human Genetics *110*, 442–459
March 2, 2023



https://doi.org/10.1016/j.ajhg.2023.02.004

CellPress

# Epigenetic regulation of plastin 3 expression by the macrosatellite *DXZ4* and the transcriptional regulator CHD4

Eike A. Strathmann,[1,2,3] Irmgard Hölker,[1,2,3] Nikolai Tschernoster,[1,4] Seyyedmohsen Hosseinibarkooie,[1,2,3,7] Julien Come,[5] Cecile Martinat,[5] Janine Altmüller,[4,8] and Brunhilde Wirth[1,2,3,6,*]

## Summary

Dysregulated Plastin 3 (*PLS3*) levels associate with a wide range of skeletal and neuromuscular disorders and the most common types of solid and hematopoietic cancer. Most importantly, *PLS3* overexpression protects against spinal muscular atrophy. Despite its crucial role in F-actin dynamics in healthy cells and its involvement in many diseases, the mechanisms that regulate *PLS3* expression are unknown. Interestingly, *PLS3* is an X-linked gene and all asymptomatic *SMN1*-deleted individuals in SMA-discordant families who exhibit *PLS3* upregulation are female, suggesting that *PLS3* may escape X chromosome inactivation. To elucidate mechanisms contributing to *PLS3* regulation, we performed a multi-omics analysis in two SMA-discordant families using lymphoblastoid cell lines and iPSC-derived spinal motor neurons originated from fibroblasts. We show that *PLS3* tissue-specifically escapes X-inactivation. *PLS3* is located ~500 kb proximal to the *DXZ4* macrosatellite, which is essential for X chromosome inactivation. By applying molecular combing in a total of 25 lymphoblastoid cell lines (asymptomatic individuals, individuals with SMA, control subjects) with variable *PLS3* expression, we found a significant correlation between the copy number of *DXZ4* monomers and *PLS3* levels.

Additionally, we identified chromodomain helicase DNA binding protein 4 (CHD4) as an epigenetic transcriptional regulator of *PLS3* and validated co-regulation of the two genes by siRNA-mediated knock-down and overexpression of *CHD4*. We show that CHD4 binds the *PLS3* promoter by performing chromatin immunoprecipitation and that CHD4/NuRD activates the transcription of *PLS3* by dual-luciferase promoter assays.

Thus, we provide evidence for a multilevel epigenetic regulation of *PLS3* that may help to understand the protective or disease-associated *PLS3* dysregulation.

## Introduction

Plastin 3 (PLS3 [MIM: 300131]) is a $Ca^{2+}$-dependent F-actin binding and bundling protein, associated with pathologies of the musculoskeletal system, nephrological disorders, malignancies of the solid and hematopoietic system, and numerous neuromuscular disorders.[1] Knockout or genetic variants of *PLS3* are causative for osteoporosis (MIM: 300910), while overexpression is associated with osteoarthritis.[1–4] Furthermore, PLS3 abundance is positively associated with the severity of congenital anomalies of the kidney and urinary tract (CAKUT).[5] Increased *PLS3* levels are a hallmark of the most common solid tissue malignancies and are found in circulating tumor cells of individuals with colorectal cancer and breast cancer.[1,6–9] In Sézary syndrome (mycosis fungoides [MIM: 254400]), an aggressive form of cutaneous T cell lymphomas, circulating $CD4^+$ T cells show increased expression of *PLS3* compared to normal $CD4^+$ T cells.[10,11] Finally, PLS3 is associated with a wide range of neuromuscular disorders, such as spinal muscular atrophy (SMA [MIM: 253400]), amyotrophic lateral sclerosis (ALS), and *CHP1*-associated ataxia (spastic ataxia 9, autosomal recessive [MIM: 618438]), where it acts as genetic protective modifier.[12–14]

SMA is a devastating motor neuron disorder, caused by homozygous loss of survival of motor neuron 1 (*SMN1* [MIM: 600354]).[15] SMN is an essential housekeeping protein and its complete loss is embryonically lethal. All individuals with SMA carry 1 to 6 copies of the survival of motor neuron 2 (*SMN2* [MIM: 601627]) gene, which differs from *SMN1* in five nucleotides. Due to a splicing defect, *SMN2* produces mainly nonfunctional *SMN* transcripts lacking exon 7. Of the remaining mRNA, only about 10% are correctly spliced.[16–19] By this, the *SMN2* copy number inversely correlates with the severity of SMA.[16]

Rarely, SMA-discordant families include fully asymptomatic individuals that share the causative *SMN1* deletion and the same *SMN2* copy number as their affected siblings. Overexpression of *PLS3* protects from intermediate and mild forms of SMA and rescues axonal growth defects

[1]Institute of Human Genetics, University Hospital of Cologne, University Cologne, Kerpener Str. 34, 50931 Cologne, Germany; [2]Center for Molecular Medicine Cologne, University of Cologne, 50931 Cologne, Germany; [3]Institute for Genetics, University of Cologne, 50674 Cologne, Germany; [4]Cologne Center for Genomics and West German Genome Center, University of Cologne, 50931 Cologne, Germany; [5]INSERM/ UEVE UMR 861, Université Paris Saclay, I-STEM, 91100 Corbeil-Essonnes, France; [6]Center for Rare Diseases, University Hospital of Cologne, 50931 Cologne, Germany
[7]Present address: Department of Neuroscience, University of Virginia, Charlottesville, VA, USA
[8]Present address: Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Core Facility Genomics, Charitéplatz 1, 10117 Berlin, Germany
*Correspondence: brunhilde.wirth@uk-koeln.de
https://doi.org/10.1016/j.ajhg.2023.02.004.

and motor neuron function in humans and various animal models.[12,20,21] The mechanisms that regulate the expression of this genetic modifier are largely unknown. In healthy individuals, expression of *PLS3* was thought to be limited to solid tissues.[22,23] However, in about 5% of the general population, we have found increased *PLS3* levels in blood.[12] More importantly, we have found up to 40-fold upregulation of PLS3 in blood and Epstein Barr virus-transformed lymphoblastoid (EBV) cell lines derived from white blood cells in asymptomatic individuals compared to their affected siblings in seven SMA-discordant families.[12,24] In fibroblasts (FBs) from the same siblings, no differences in *PLS3* expression have been found. Instead, iPSC-derived spinal motor neurons (MNs) differentiated from these FBs show a differential *PLS3* expression pattern similar to lymphoblastoid cells.[24] All fully asymptomatic individuals in SMA-discordant families with *PLS3* upregulation are female, indicating a sex-specific mechanism of upregulation, such as escape from X chromosome inactivation (XCI). *PLS3* is indeed X-linked and a known facultative escape gene.[25–28]

*PLS3* is located on Xq23 in close proximity to the unique macrosatellite *DXZ4*.[29,30] This macrosatellite is essential for XCI, as its deletion leads to a de-condensation of the inactivated X chromosome ($X_i$) in mice.[31] In humans, the copy number of *DXZ4* was estimated to range between 50 and 100 tandem repeats of a 3 kb repeat monomer.[32] In the recently published telomere-to-telomere assembly of the human X chromosome, the copy number of *DXZ4* consists of 55 repeats.[33] In females on the $X_i$, *DXZ4* is hypomethylated and binds the architectural protein CCCTC-binding factor (CTCF [MIM: 604167]), both features of active chromatin.[32] In hemizygous males and on the active X chromosome ($X_a$) of females, *DXZ4* has heterochromatic epigenetic features.[32,34] Given the extreme differences in the copy number of *DXZ4* and the hypomethylated *DXZ4* locus on the $X_i$, we hypothesize that the macrosatellite influences the escape from XCI of genes in its nuclear neighborhood, such as *PLS3*. Here, we performed a comprehensive multi-omics analysis to uncover the multilevel regulatory mechanism of *PLS3* expression in various tissues in both sexes. We validated that *PLS3* is able to escape XCI in iPSC-derived spinal MNs from asymptomatic females. Furthermore, we measured the copy number of *DXZ4* in 25 EBV cell lines using molecular combing and found a significant linear correlation with the expression levels of *PLS3* in females.

Independent of the X-inactivation status of a gene or the general grade of chromatin activity, a transcription factor or transcriptional regulator must be available in the target tissue in order to achieve gene expression. The transcription factors that drive the expression of *PLS3* are unknown. By analysis of transcriptome data of male EBV cells expressing *PLS3*, we identified chromodomain helicase DNA binding protein 4 (CHD4 [MIM: 603277]), localized on chromosomal region 12p13.31 as an epigenetic transcriptional regulator of *PLS3*. CHD4 is one of multiple motor proteins of the nucleosome remodeling deacetylase (NuRD) complex.[35] CHD4/NuRD binds DNA and histones and performs nucleosome sliding and positioning.[36] We validated CHD4/NuRD as transcriptional regulator of *PLS3* by application of siRNA-mediated knock-down of *CHD4*, overexpression of CHD4 from a plasmid, chromatin immunoprecipitation, and dual-luciferase promoter assays. Thereby, we were able to show that CHD4 interacts with the *PLS3* promoter and positively regulates the expression of the genetic modifier in EBV and HEK293T cells. Thus, we provide evidence for a multilevel epigenetic regulation of *PLS3* that may help to understand the protective or disease-associated *PLS3* dysregulation. Our findings may help to understand and to predict dysregulation of *PLS3* in pathologic conditions, such as cancer, SMA, and osteoarthritis.

## Material and methods

### Individuals' material

Individual-derived skin FBs have been established from three SMA3-affected and three asymptomatic siblings belonging to two unrelated SMA-discordant families.[12,24] iPSCs from family 1 and 2 were previously generated and reported.[24] EBV cells were generated from blood samples, including six asymptomatic females, 32 SMA-affected individuals with high or low expression levels of *PLS3* of both sexes, and six healthy control subjects as previously reported[12] (Table S1). Informed written consent was obtained from individuals with SMA, caregivers, and family members according to the Declaration of Helsinki, and the study was approved by the ethics committee of the University Hospital of Cologne under the approval numbers 04–138 and 13–022.

### Cell lines and maintenance

HEK293T cells and FBs were maintained in DMEM (ThermoFisher Scientific) with 10% FBS (Sigma), 1% penicillin and streptomycin (ThermoFisher Scientific), and 6.25 μg/mL Amphotericin B (ThermoFisher Scientific). EBV cells were maintained with RPMI 1640 medium with 20% FBS, 1% penicillin and streptomycin (ThermoFisher Scientific), and 6.25 μg/mL Amphotericin B (ThermoFisher Scientific). The cells were cultured at 37°C and 5% $CO_2$.

### Motor neuron differentiation from iPSCs

The transcriptomes of spinal MNs of two siblings from family 1 were sequenced in triplicates. In family 2, two iPSC clones were sequenced in triplicates. The iPSCs were grown on Matrigel-coated cell culture plates in mTeSR1 medium (STEMCELL Technologies). The medium was changed on a daily basis. The cells were split by treatment with Alfazyme (PAA) at a confluence of about 70%–75%. The medium was supplemented with 10 μM ROCK inhibitor Y-27632 (Tocris Bioscience) after plating to single-cell survival.

iPSCs were differentiated into spinal MNs in 384-well plates as previously described.[37] Briefly, hiPSC were dissociated enzymatically using Stem Pro Accutase (ThermoFisher) for 5 min and plated in 25 $cm^2$ flask (Dutscher) or spinner flask (Corning) at a density of $0.2 \times 10^6$ cell/mL in a total volume of (respectively) 10 mL and 70 mL. Cells were incubated in suspension in a neural induction medium containing of DMEM/F12 and Neurobasal medium supplemented with B27 (Life Technologies), N2 Supplement (Life Technologies), Pen-Strep 0.1%, β-mercaptoethanol 0.1% (Life Technologies),

ascorbic acid (0.5 μM, Sigma-Aldrich), Chir99021 (3 μM, Stemgent), SB431542 (20 μM, Tocris), LDN 193189 (0.2 μM, Stemgent), and Y-27632 (10 μM, Stemcell). After two days, caudalization of the neural progenitors was obtained by addition of retinoic acid (RA 0.1 μM) and smoothened agonist (SAG, 0.5 μM). After one week in suspension, brain-derived neurotrophic factor (BDNF, 10 ng/mL), glial-derived neurotrophic factor (GDNF, 10 ng/mL), and N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-S-phenylglycine *t*-butyl ester (DAPT, 10 nM) were added in the medium without SB431542 and LDN193189. Between day 10 and 14, MN progenitors were converted into MNs. At this step (day 10), embryoid bodies were dissociated into single cells with Trypsin-EDTA and plated either in 384-well plates (Corning) or cell culture dish coated with poly-ornithin (20 μg/mL, Sigma-Aldrich) and 5 g/mL laminin (Invitrogen) into the same medium with Y-27632. MNs were obtained after 14 days of differentiation and the differentiation efficiency was assessed by immunolabeling for Islet1 (ISL1, from Neuromics Ref GT15051) and Hb9 (from DHSB ref. 81.5C10).

### Extraction of RNA

RNA for sequencing and RT-qPCR experiments was extracted from densely populated EBV, FB, or spinal MN cultures using the RNeasy Blood & Tissue Kit (Qiagen) or the Monarch Total RNA Miniprep Kit according to the instruction manuals. DNA was digested using DNaseI and the RNA was solved in 50 μL nuclease-free $H_2O$.

### Transcriptome analysis

Approximately 1,200 ng RNA (20 ng/μL in 60 μL) was sent for RNA sequencing to deCODE genetics, Iceland. For each sibling of family 1, RNA-seq of iPSC-derived spinal MNs was performed in three replicates. For family 2, two independent iPSC clones with each three replicates were generated. In addition to that, RNA sequencing was performed on RNA from 40 EBV cell lines from six asymptomatic females, 32 SMA-affected individuals with various expression of *PLS3* of both sexes, and six healthy control subjects, including all siblings from families 1 and 2. Finally, RNA from six FB cell lines was sent to RNA sequencing including all siblings from families 1 and 2. The library preparation was done using the TruSeq RNA library Prep Kit (Illumina). RNA was sent for whole-exome sequencing to deCODE genetics, Iceland. The sequencing was performed on a HiSeq2500 (Illumina) machine. For each sample, about 128,000,000 paired reads were generated. The transcriptome data were aligned to the human reference genome assembly GRCh37/hg19 using HISAT2.[38] In each sample we annotated about 35,400 transcripts. Variants were called using BCFTOOLS.[39,40] X chromosome bi-allelic SNVs were identified from vcf-files of spinal MNs using the following filter criteria: read depth $\geq$ 100, Phred-scaled quality score $\geq$ 30, and SNV ratio between 0.1 and 0.9.

### Differential expression (Kallisto)

The differential expression analysis was performed using KALLISTO and DESeq2.[41,42] Sample identity was validated using NGSCheckMate.[43] A principal component analysis (built-in function of DESeq2) was performed to validate that samples with the same phenotype (e.g., asymptomatic) or biological replicates cluster together separately. The list of differentially expressed genes was filtered for p values $\leq$ 0.001, a base mean of $\geq$500, and an absolute $\log_2$ fold-change of $\geq$0.5.

### RT-qPCR

Approximately 800 ng RNA from EBV cells or HEK293T cells was reverse transcribed into cDNA using the QuantiTect Reverse Transcription Kit (QIAGEN) in accordance with the instruction manual. Real-time qPCR was performed using the SybrGreen Master Mix (Applied Biosciences) in accordance with the manufacturer's instruction manual with 20 ng cDNA per reaction. All experiments were performed using a StepOnePlus Real-Time PCR system (Applied Biosciences). *HPRT* was used as housekeeping gene (Table S2). An internal reference sample was included to all 96-well plates to compare the expression levels of all measured samples. All experiments were conducted in triplicates. The expression levels were analyzed using the ΔΔCt-method.[44]

### *CHD4* siRNA knock-down

siRNA-mediated knock-down of *CHD4* was conducted in three male and three female EBV cell lines. 750,000 EBV cells were seeded into 24-well plates and transfected with 1 μM *CHD4* siRNA (ThermoFisher Scientific, siRNA ID 121307) or a mock siRNA (AllStars negative Control siRNA, Qiagen, Cat. No./ID 1027281). Lipofectamine 3000 (ThermoFisher Scientific) was used as transfection reagent according to the instruction manual and the transfection was carried out in six replicates. On the next day, we harvested the cells, extracted RNA, performed reverse transcription, and measured the expression of *PLS3* and *CHD4* by RT-qPCR. *HPRT* was used as housekeeping gene.

### CHD4 overexpression assay

CHD4 was overexpressed in HEK293T cells using a *CHD4* expressing plasmid. In the first experiment, 125,000 HEK293T cells were transfected with 25 ng, 50 ng, 75 ng, or 100 ng of a CHD4 expression vector under a *CMV* promoter (pLV[Exp]-CMV>hCHD4 [GenBank: NM_001273.5], VectorBuilder). In a second experiment, 450,000 HEK293T cells were transfected with 250 ng, 500 ng, 1,000 ng, or 1,500 ng of the CHD4 expression vector, respectively. Lipofectamine 3000 (ThermoFisher Scientific) was used as transfection reagent according to the instruction manual and the transfection was carried out in three to four replicates. At the next day, the cells were harvested and RNA was extracted and reverse transcribed to perform RT-qPCR. The expression of *CHD4* and *PLS3* was measured. *HPRT* was used as housekeeping gene.

### Dual-luciferase promoter assay

A fragment of 1,101 bp of the *PLS3* promoter (GRCh37/hg19 ChrX: 114,794,485–114,795,585) including the transcription start site was cloned into a luciferase vector (pRP[Exp]-PLS3(1.0 kb)>Luc2, VectorBuilder). The coding sequence of *CHD4* (GenBank: NM_001273.5) was cloned into a vector under a *CMV* promoter (pLV[Exp]-CMV>hCHD4, VectorBuilder). Next, 125,000 HEK293T cells were seeded into 24-well plates in 500 μL medium. On the next day, the cells were co-transfected with 20 ng of the Firefly-luciferase vector under the *PLS3* promoter, 10 ng of the Renilla-luciferase vector (pRL-TK, Promega) as internal reference and several concentrations (0 ng, 6.25 ng, 12 ng, 25 ng) of the *CHD4* overexpression vector using 0.75 μL Lipofectamine 3000 in accordance with the instruction manual. In addition to that, cells were co-transfected with 50 ng of a Firefly-luciferase vector under a *CMV* promoter (pRP[Exp]-CMV>Luc2, Promega) and 10 ng of the Renilla-luciferase vector (pRL-TK, Promega) as positive control. A negative control was co-transfected with 10 ng of a Firefly-luciferase vector without promoter (pGL-4.10>Luc2, Promega) and 10 ng of the Renilla-luciferase vector

(pRL-TK, Promega). After 24 h, cells were harvested and lysed for 30 min with 1xlysis buffer and the luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) on a GloMax 96 luminometer. Three independent experiments with each four technical replicates per treatment group were performed.

### DNA extraction for whole-genome sequencing/read depth analysis of *DXZ4*

Genomic DNA (gDNA) was extracted from FBs using the DNeasy Blood & Tissue Kit (Qiagen) in accordance with manufacturer's instructions. Approximately 1,200 ng DNA (20 ng/μL in 60 μL) was sent for DNA sequencing to deCODE genetics, Iceland. The library preparation was done using the TruSeq DNA library Prep Kit (Illumina). Paired-end sequencing by synthesis (SBS) using 2×150 cycles of incorporation and imaging was performed on Illmina HiSeqX sequencers. The whole-genome datasets were aligned to the human reference genome assembly GRCh37/hg19 and the read depth for each position of the chromosomes 22 and X was counted by Samtools.[45] The average read depth for chromosome 22 and the *DXZ4* locus was calculated. The read depth of the variable region (ChrX: 114,959,000–115,006,000) of *DXZ4* was normalized by the read depth of chromosome 22.

### Molecular combing

This method uses multi-color DNA fiber probes to mark macrosatellite repeat regions on linear stretched genomic DNA molecules.[46] Custom-made fiber probes were designed by Genomic Vision. Each individual repeat monomer was covered by two differently labeled (red and green) 1.1 kb fiberprobes separated by a 0.4 kb gap. Additionally, a BAC (RP11-761E20, 179 kb) covered the region distally of *DXZ4* as telomeric probe visualized as a blue signal. A centromeric red probe was covered by the fosmid G2487005D12 and is located 35 kb from the *DXZ4* sequence. To perform molecular combing, high-molecular-weight genomic DNA was extracted from 750,000 EBV cells using the Fiber Prep DNA Extraction Kit (Genomic Vision). The DNA was embedded in agarose plugs and treated with proteinase K overnight at 50°C, melted at 68°C, and digested with β-agarose at 42°C overnight. The DNA was gently mixed with 1.2 μL Combing buffer. Vinyl-silane coated coverslips were slowly plunged into the DNA solution at constant speed of 300 μm/s using the FiberComb Molecular Combing system (Genomic Vision). The coverslips were dried at 65°C for 2 h. To hybridize the combed DNA with the fiber probes, the coverslips were de-hydrated by washing with increasing concentrations of 70%, 90%, and 90% ethanol for 1 min, respectively. After that, the coverslips were air-dried at room temperature for 10 min protected from light. Fiber probes were mixed 1:1 with de-ionized formamide and incubated at 37°C for 30 min. Next, 20 μL of the fiber probe/formamide mix were transferred to a microscope slide and the combed coverslip was placed upside-down on the microscope slide avoiding air bubbles. The combed DNA and the fiber probes were co-denaturized for 5 min at 90°C in a humidified chamber and incubated 16 to 20 h at 37°C. Finally, the coverslip was removed from the slide and washed three times with pre-warmed hybridization buffer at 60°C. Next, the coverslip was placed upside-down on a microscope slide with 20 μL detection solution and incubated at 37°C in a humidified chamber. The coverslip was removed and washed three times with detection washing buffer for 3 min each time. Then, the coverslip was washed with PBS for 3 min and de-hydrated by washing with increasing concentrations of 70%, 90%, and 90% ethanol for 1 min, respectively. The air-dried cover slips were scanned using the FiberVision automated scanner (Genomic Vision). The images were analyzed using the FiberStudio software. The lengths of the *DXZ4* locus and the upstream and downstream regions were marked manually and the copy number was calculated as the length of the DNA in kilobases divided by three, as the *DXZ4* repeat monomer is 3 kb in size. Only signals that contained both upstream and downstream regions were counted as complete signals and further analyzed.

### Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed using the iDeal ChIP-qPCR kit (Diagenode) in accordance with the instruction manual. In short, for one immunoprecipitation, $2.5 \times 10^6$ EBV or HEK293T cells were fixed in PBS by adding 1/10 fixation buffer for 13 min at room temperature. Glycine was added in a proportion of 1/10 to stop the fixation for 5 min on a shaker. The cells were centrifuged at 500 × *g* for 10 min at 4°C. Next, the cell pellet was washed twice with ice-cold PBS. After centrifugation, 1 mL per million cells ice-cold lysis buffer (il1b) was added and the samples were incubated for 20 min at 4°C on a rotator. After centrifugation at 500 × *g* at 4°C, the supernatant was discarded. The cells were re-suspended in 1 mL per million cells ice-cold lysis buffer il2, incubated for 10 min at 4°C, and centrifuged at 500 × *g*. Protease inhibitor cocktail was added to the shearing buffer iS1b in a 1:200 ratio and finally mixed with 0.1% SDS on ice. The cell pellets were mixed with 150 μL shearing buffer per 1.8 million cells. The DNA was sheared for three to six cycles using a Bioruptor Pico Sonification (Diagenode) device. To check the fragment size, an aliquot of the sheared DNA was reverse crosslinked (incubation for 4 h at 65°C), purified with a MicroChIP Diapure Column, and analyzed on a 1.5% agarose gel. An aliquot of 1% (2.5 μL) of the sheared DNA was kept aside as input. For one immunoprecipitation, 30 μL of Protein A magnetic beads were washed three times with 1 mL ChIP buffer iC1b according to the manual. A ChIP reaction mix was prepared containing for one reaction 6 μL BSA, 1.8 μL 200× protease inhibitor cocktail, 20 μL 5× iC1b buffer, and 5 μg of the required ChIP-grade Antibody (anti-CHD4 [rabbit], Abcam ab72418, anti-CTCF [rabbit], Diagenode, C15410210, anti-IgG [rabbit], Diagenode, C15410206). 70 μL of the ChIP reaction mix was added to each sample and incubated at 4°C for 4 to 5 h. The sheared DNA (250 μL) was added to the prepared beads and incubated on a rotator overnight at 4°C. The next day, the samples were centrifuged shortly and placed into a magnetic rack for 1 min to allow the beads to be captured by the magnetic field, then the supernatant was removed and the beads were washed with buffer iW1. The samples were incubated for 5 min on a rotator at 4°C. The washing steps were repeated as described for the wash buffers iW2 to iW4. DIB buffer was completed by adding 1 μL proteinase K to 100 μL. The IP samples were re-suspended in 100 μL of DIB buffer, while 97.5 μL were added to the input samples. The IP and input samples were incubated at 55°C for 15 min and at 100°C for another 15 min. Then the tubes were briefly spun down and placed in the magnetic rack for 1 min. The supernatant, which contained the DNA, was purified using a MicroChIP DiaPure column. RT-qPCR was performed with input and immunoprecipitated DNA using primers directed to the promoters of *PLS3* and *TCEAL4* as well as *MB* and *H19* as controls (Table S2). The recovery was calculated as:

$$\%\ recovery = 2^{(Ct\ input - 6.64 - Ct\ sample)} * 100\% \qquad \text{(Equation 1)}$$

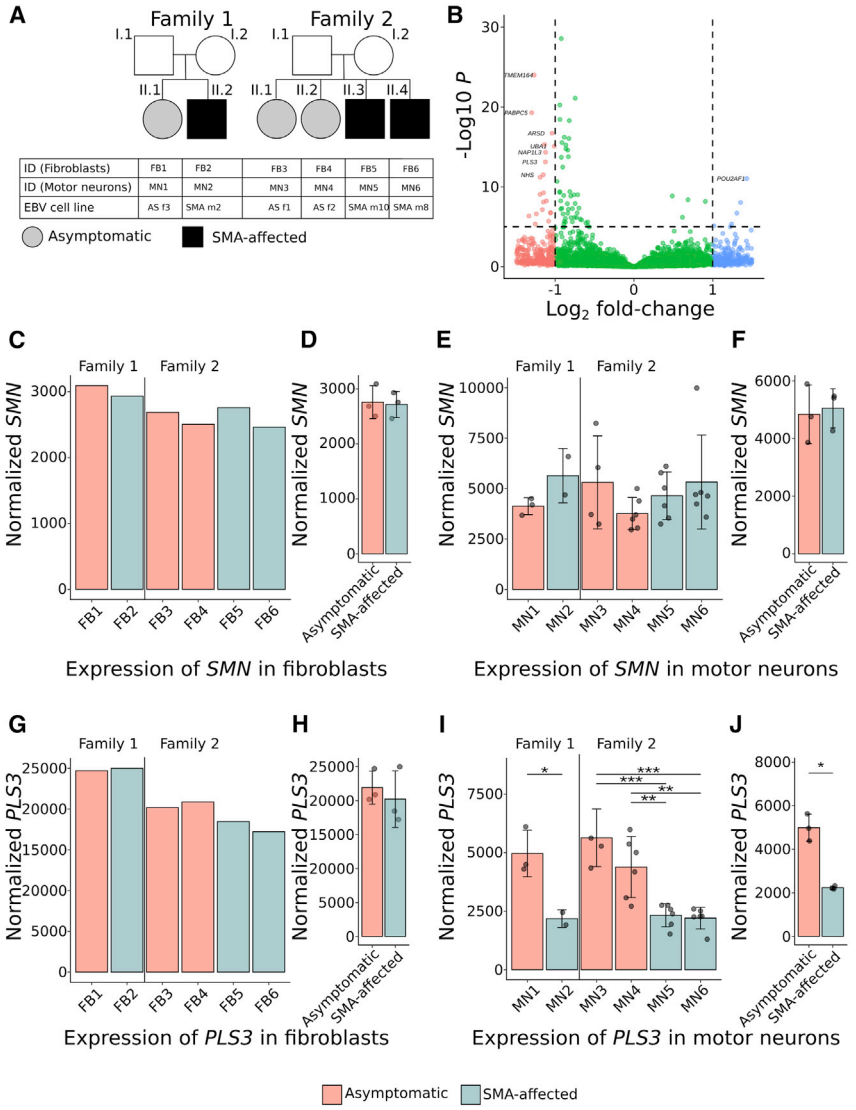


**Figure 1. Expression of *SMN* and *PLS3* in fibroblasts and iPSC-derived spinal motor neurons**

(A) Pedigrees of two SMA-discordant families comprising both affected (black) and asymptomatic (gray) individuals, all carrying homozygous *SMN1* deletions and three *SMN2* copies.

(B) Volcano plot of 80 differentially expressed genes, including *PLS3*, identified in transcriptomes of iPSC-derived spinal motor neurons.

(C–F) Gene expression levels of *SMN* in FB- and spinal MN-transcriptome data measured as normalized counts indicate no differences between asymptomatic and SMA-affected individuals in both cell types.

(G–J) Gene expression levels of *PLS3* in FB- and spinal MN-transcriptome data. A $\log_2$-fold difference of 1.12 was found between asymptomatic and affected siblings in MNs. Data are represented as mean ± SD (*p < 0.05; **p < 0.01; ***p < 0.001).

three asymptomatic and three SMA-affected individuals from two SMA-discordant families (Figure 1A). All siblings of both families share the homozygous deletion of *SMN1* and have each three copies of *SMN2*.[12,24] Overall, we identified 80 differentially expressed genes comparing asymptomatic and symptomatic individuals, including 61 X-linked genes (Figure 1B and Table S3). Both cell types show low indifferent expression of *SMN* in all six siblings (Figures 1C–1F). *PLS3* was strongly expressed in FBs of both asymptomatic and affected individuals without differences between the phenotypes (Figures 1G and 1H). However, we validated differential expression of *PLS3* in MNs and found a 1.12 $\log_2$-fold upregulation in asymptomatic female compared to their affected male siblings, confirming our previous results (Figures 1I and 1J).[12,24]

As *PLS3* is a known facultative escape gene, the doubling in the expression in female siblings raised the question of whether the gene indeed escaped XCI in the spinal MNs.[25–28] To investigate this, we further analyzed the spinal MN transcriptomes. The sequenced populations of iPSCs derived from single cell clones. As previously reported for other iPSCs, all cells derived from a single clone maintain the same XI-status as the original derived cell.[48] We tested whether this is indeed the case and measured the average read depth of *XIST* in the spinal MN transcriptomes. The asymptomatic sibling of family 1 showed an average read depth of *XIST* of about 27, while the two female siblings of family 2 showed average read depths of about 10 (MN3.1) and 8 (MN4.1). In the male sample, no *XIST* expression has been found (Figures S1A–S1G). Our

## Statistics

All statistical analyses were performed using R (v.4.1.2 (2021-11-01)) and RStudio (v.2021.09.2 Build 382). All plots were generated using the ggplot2 package. Wilcoxon rank sum tests were performed to analyze differences in expression levels. Multiple comparisons were conducted by ANOVA, followed by post-hoc Dunnett's tests. Linear correlation was calculated as Spearman's rank correlation coefficient ρ. A p value of less than 0.05 was considered as significant (*p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001).

## Results

### *PLS3* escapes from XCI in iPSC-derived spinal motor neurons

To validate previously described *PLS3* and *SMN* expression patterns, we performed differential expression analysis in transcriptomes of FBs and iPSC-derived spinal MNs, which derived from the same FBs.[47] The transcriptomes belong to

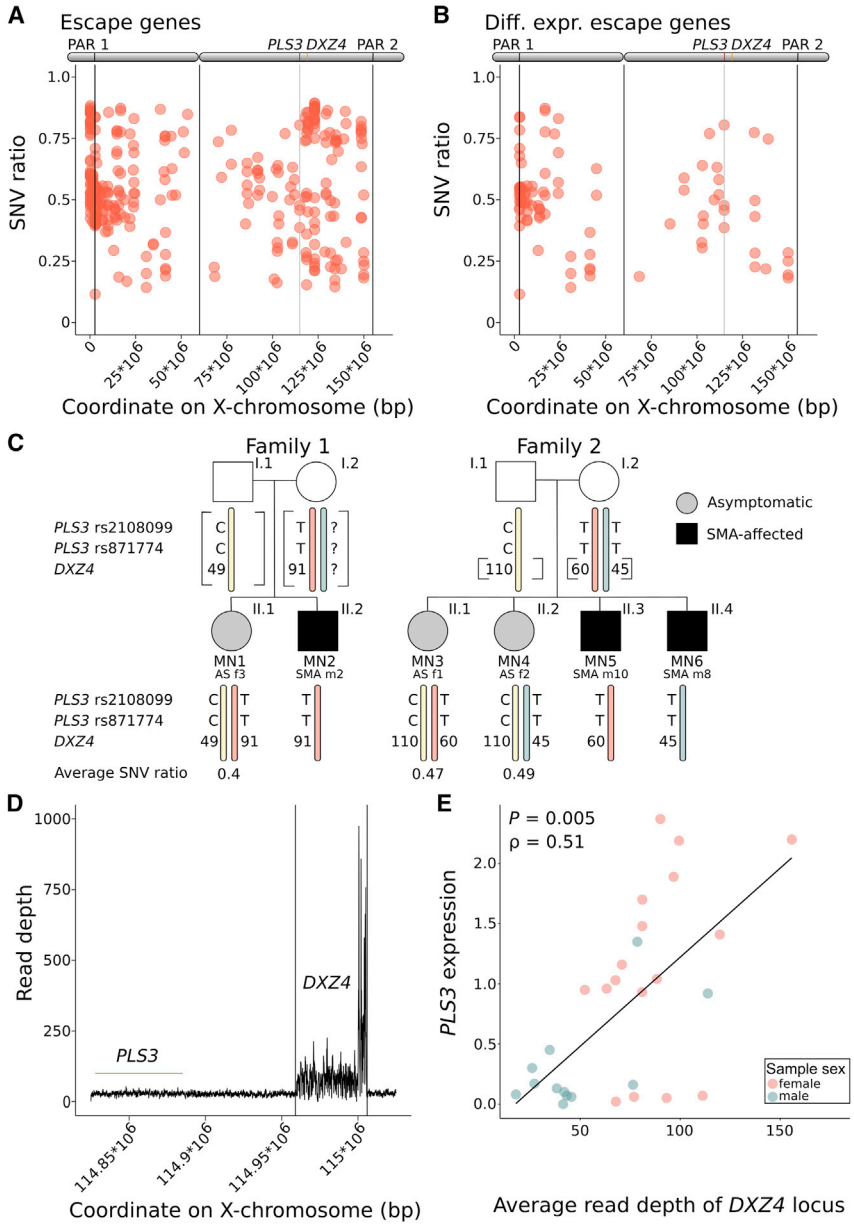


**Figure 2. Bi-allelic expression of *PLS3***

(A) Escape genes (including *PLS3*) were identified in transcriptomes of iPSC-derived spinal MNs by filtering exonic bi-allelic SNVs with an SNV ratio of 0.1–0.9. The escape genes were distributed along the X chromosome.

(B) Location and SNV ratio of escape genes that were significantly differentially expressed, including *PLS3*.

(C) Two exonic *PLS3* SNVs were identified in transcriptome data of spinal MNs of both families. The female siblings were heterozygous (C/T) for both SNVs, while the male siblings carry the major (T) allele. The *DXZ4* copy numbers measured by molecular combing are given for each sibling.

(D) The read depth of *DXZ4* in 32 whole-genome datasets was measured and normalized. The *PLS3* locus shows a constant read depth, while the *DXZ4* locus shows a highly variable region.

(E) The average normalized read depth of *DXZ4* compared to expression levels of *PLS3* measured by RT-qPCR. A linear relationship was determined by Spearman's correlation coefficient test.

findings indicate that the XCI in both families stayed intact, but the female sibling of family 1 showed stronger *XIST* levels than the two sisters from family 2.

Next, we identified potential escape genes, by calling X-linked heterozygous bi-allelic SNVs from the MN transcriptomes. We filtered for an SNV ratio between 0.1 and 0.9, which is a commonly used cut-off for the definition of escape genes.[26–28,49–51] Next, we discarded genes that belong to the pseudoautosomal regions (PAR1 ChrX: 1–2,699,520, PAR2 ChrX: 154,931,044–160,000,000, GRCh37/hg19) and after that, we identified 147 bi-allelic SNVs referring to 107 bi-allelically expressed X-linked genes including *PLS3* in both families (Figures 2A and S2, Table S4). Since the X chromosome contains approximately 900 coding genes, we found that about 10% of X-linked genes escaped the XCI. Taking into account that we were only able to detect escape genes that contain an exonic bi-allelic SNV, this number seemed plausible. Not all bi-allelically expressed genes are necessarily differentially expressed, as the expression of the genes may be regulated by other mechanisms. Therefore, we compared our list of bi-allelically expressed genes with our list of significant differentially expressed genes and found 35 differentially expressed escape genes, including *PLS3* (Figure 2B, Table S4).

In male samples, bi-allelic SNVs belonging to 15 genes were found outside of the PARs (Figures S2A–S2C, Table S5). The majority of them are either pseudogenes or multicopy genes and were therefore discarded. In the female samples, the SNVs were distributed along the chromosome (Figures S2D–S2F).

We identified two exonic bi-allelic SNVs in the exons 11 and 12 of *PLS3*. Interestingly, both SNVs were found in all three asymptomatic siblings of both families. Both SNVs were validated by Sanger sequencing (Figure S3A). All female siblings in both families are indeed heterozygous (C/T) for both SNVs, while the male individuals show the major allele (T) (Figures S3A and S3B). The average SNV ratio of *PLS3* in the MNs of the female siblings was about 0.4–0.6, indicating that both alleles were transcribed in equal proportions (Figures 2A, 2B, and S3C). Both SNVs were previously listed in the gnomAD database as rs871774 (ChrX: 114,880,423-C-T [GRCh37/hg19]; GenBank: NM_005032.7, c.1294T>C; GenBank: NP_005023.2, p.Leu432=) and

rs2108099 (ChrX: 114,879,399-C-T [GRCh37/hg19]; GenBank: NM_005032.7, c.1242T>C; GenBank: NP_005023.2, p.Pro414=). The total allele frequencies of the minor alleles of both SNVs is about 0.06, while the allele frequency in the European population is about 0.03. Interestingly, the allele frequency was about 10 times higher in the African populations; 0.26 for rs871774 and 0.32 for rs2108099, respectively.

Hereby, we were able to show that *PLS3* is differentially expressed in spinal MNs, but not in FBs of asymptomatic individuals compared to their SMA-affected siblings of two SMA-discordant families. The differential expression of *PLS3* can be explained by its escape from XCI in spinal MNs as both alleles were transcribed in similar proportions.

### *DXZ4* copy number correlates with the expression of *PLS3*

Escape from XCI is an epigenetic form of transcriptional regulation. Genes that escape XCI show hypermethylation within the gene bodies, while promoter regions are hypomethylated.[52–59] Generally, escape genes are enriched with histone marks that resemble those of active chromatin. *PLS3* is located directly adjacent to *DXZ4*, a macrosatellite with a highly variable copy number, which is essential for XCI.[31,32] The genetic locus of *DXZ4* on the $X_i$ is characterized by euchromatin, while the $X_a$ consists of heterochromatin.[34] We hypothesized that the unusual chromatin state of the *DXZ4* locus on the $X_i$ has an influence on the expression and escape of neighboring genes, such as *PLS3*. To investigate this, we analyzed whole-genome sequencing (WGS) datasets from 22 asymptomatic or SMA-affected individuals as well as 10 unrelated control subjects with various *PLS3* expression levels of both sexes (Table S1). For each WGS dataset, we calculated the average read depth of *PLS3* (ChrX: 114,827,819–114,885,179, GRCh37/hg19) and *DXZ4* (ChrX: 114,959,000–115,006,000). The *PLS3* locus showed a constant read depth (Figure 2D). The locus of *DXZ4* showed strong variations in the read-depth in each sample, reflecting the shotgun approach of next-generation sequencing (Figure 2D). We calculated the average read-depth of the *DXZ4* locus for each WGS dataset and normalized this number with the average read depth of chromosome 22, as this is the smallest human autosome. We found strong differences in the normalized average *DXZ4* read-depth between the samples, which indicates strong differences in the *DXZ4* copy numbers in the study population. Next, we compared the average read depth of *DXZ4* with *PLS3* levels previously measured by RT-qPCR and found a significant linear correlation indicating that the copy number of *DXZ4* indeed influences the expression levels of *PLS3* ($p \approx 0.005$, $\rho \approx 0.51$) (Figure 2E).

A limitation of our bioinformatics approach was the inability to discriminate between the two *DXZ4* alleles in females, which could be of different size. To measure the exact *DXZ4* copy number allele-specifically, we performed a method called molecular combing in 25 EBV cell lines.[60] EBV cells from 6 asymptomatic females, 11 SMA-affected females, as well as 8 SMA-affected males were cultured (Table S1). The phenotype classification of the EBV cells (asymptomatic, SMA-affected) is based on the phenotype of the individuals from which the material derived. Note that most high *PLS3*-expressing SMA women show much milder SMA phenotype than expected from only the *SMN2* copies, but they were still classified as SMA since they showed some SMA symptomes.[12] We extracted high-molecular-weight genomic DNA from each sample. The DNA was linearized on vinyl-silane-coated coverslips and hybridized with custom-made fiber probes. Each *DXZ4* repeat monomer of 3 kb was covered by two differently labeled fiber probes of 1.1 kb length, allowing the determination of the exact number of *DXZ4* copies for each allele. The regions up- and downstream of *DXZ4* were covered by two additional fiber probes, allowing the selection of only intact DNA stretches covering the entire *DXZ4* region (Figure 3A). Only complete signals, which consisted of all three regions, were further analyzed. We detected *DXZ4* copy numbers between 27 and 110 repeats (Figures 3B and S4). Interestingly, of the 17 female samples there was only one sample (SMA f9) with two *DXZ4* alleles of similar size. All other samples showed two clearly distinguishable *DXZ4* alleles. Our findings underline the high variability of *DXZ4* copy numbers in the human population. In parallel to that, RNA was extracted from each EBV cell line, and relative normalized expression levels of *PLS3* were measured by RT-qPCR (Figure 3C). Each cell line that showed only 10% or less *PLS3* expression levels compared to the sample with the strongest *PLS3* levels was defined as *PLS3* low expresser. Finally, we compared the measured *DXZ4* copy numbers measured by molecular combing with the *PLS3* levels measured by RT-qPCR. First, we compared the copy numbers of the larger *DXZ4* allele between *PLS3* high- and low-expressing females and found a significant difference ($p \approx 0.008$) with an average copy number of 90.5 in high expressers compared to only 53.1 in low expressers (Figure 4A). Furthermore, we found a significant linear correlation between the *DXZ4* copy number (larger allele) and the *PLS3* levels ($p \approx 0.01$, $\rho \approx 0.6$) (Figure 4B). The smaller *DXZ4* allele in female cells did not differ between the two groups; in average, high expressers had 67 and low expressers 51.5 repeats and no correlation (Figures 4C and 4D). In males, there was no difference in the average *DXZ4* copy number between high (79.7) and low (70) expressers (Figure 4E) and no correlation between the expression of *PLS3* and the copy number of *DXZ4* ($p \approx 0.89$, $\rho \approx 0.05$) (Figure 4F).

To compare our results obtained by molecular combing with the previous bioinformatics approach, we calculated the average *DXZ4* copy number in female cells and found a significant difference between the two groups ($p \approx 0.02$), showing an average of 71.8 *DXZ4* copy numbers

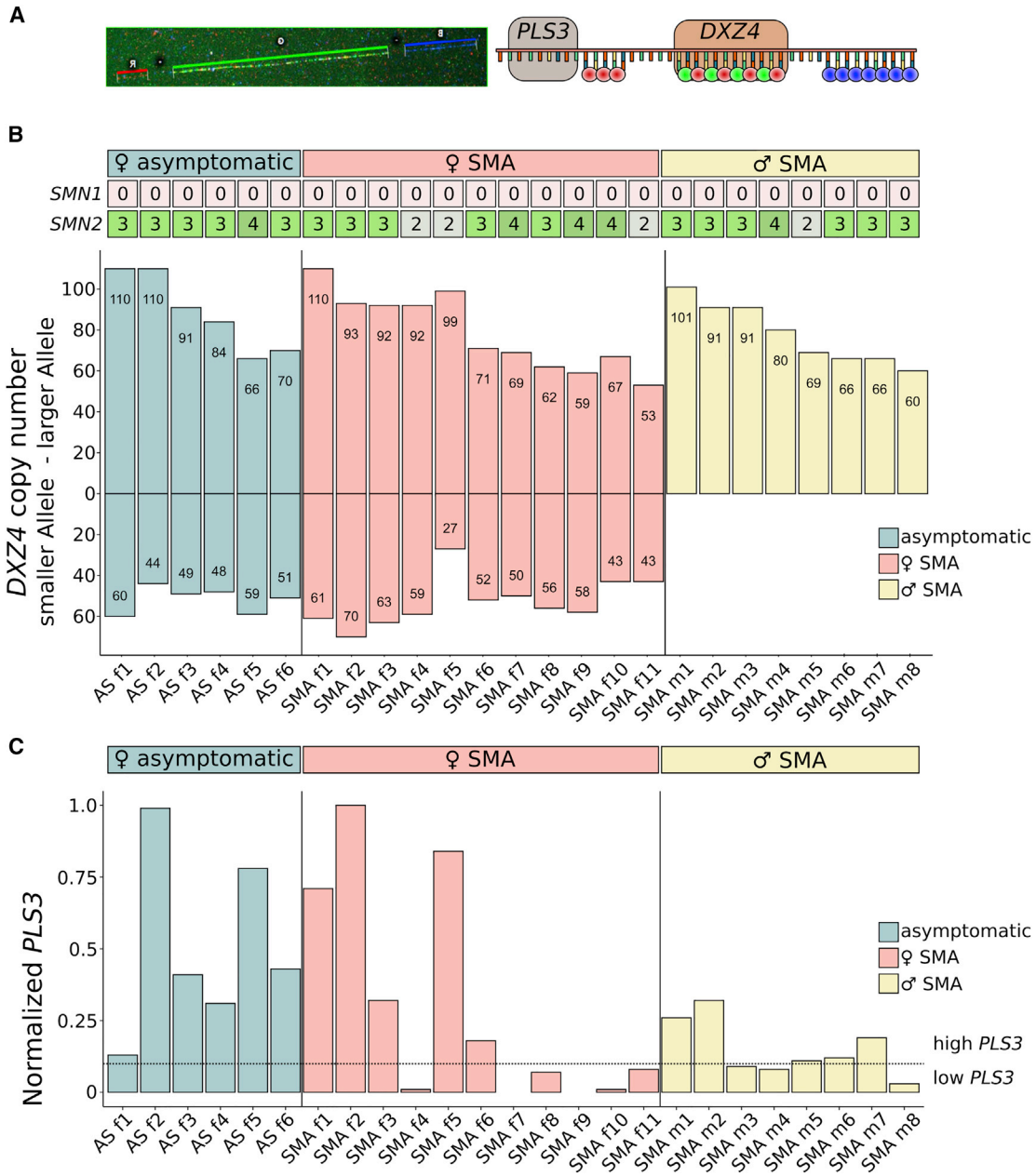


**Figure 3. The copy number of *DXZ4* determined by molecular combing**

(A) The fiber probes used for molecular combing consist of one red and one green 1.1 kb fiber probe covering each *DXZ4* repeat from each site and two fiber probes covering the up- and downstream regions.

(B) *DXZ4* copy numbers measured by molecular combing. The larger allele in females is shown in the upper half of the figure and the smaller allele in the lower half of the figure. The *SMN2* copy number is given for each sample.

(C) *PLS3* expression levels of 17 female and 8 male EBV cell lines measured by RT-qPCR. Samples are ordered by phenotype as female asymptomatic individuals (AS f), SMA-affected females (SMA f), and SMA-affected males (SMA m). Samples that express 10% or less compared to the strongest expresser were defined as *PLS3* low expressers for further analysis.

in high expressers and only 59.3 in low expressers (Figure 4G). There was a significant difference between the average *DXZ4* copy number of each sample and the *PLS3* levels ($p \approx 0.02$, $\rho \approx 0.58$) (Figure 4H).

Overall, we found a significant correlation of the *DXZ4* copy number and the expression of *PLS3* in females, but not in males. Females with an increased *PLS3* expression harbor at least one *DXZ4* allele with increased copy number (>70 repeats), while the absolute copy number difference between the two alleles differs significantly between high expressers and low expressers.

## Segregation analysis

Next, we performed a segregation analysis in the families 1 and 2 to verify whether there is a linkage disequilibrium between the minor allele of *PLS3* SNVs (rs871774 and rs2108099), which we have identified in the transcriptome data, and the larger *DXZ4* allele. Since both siblings in

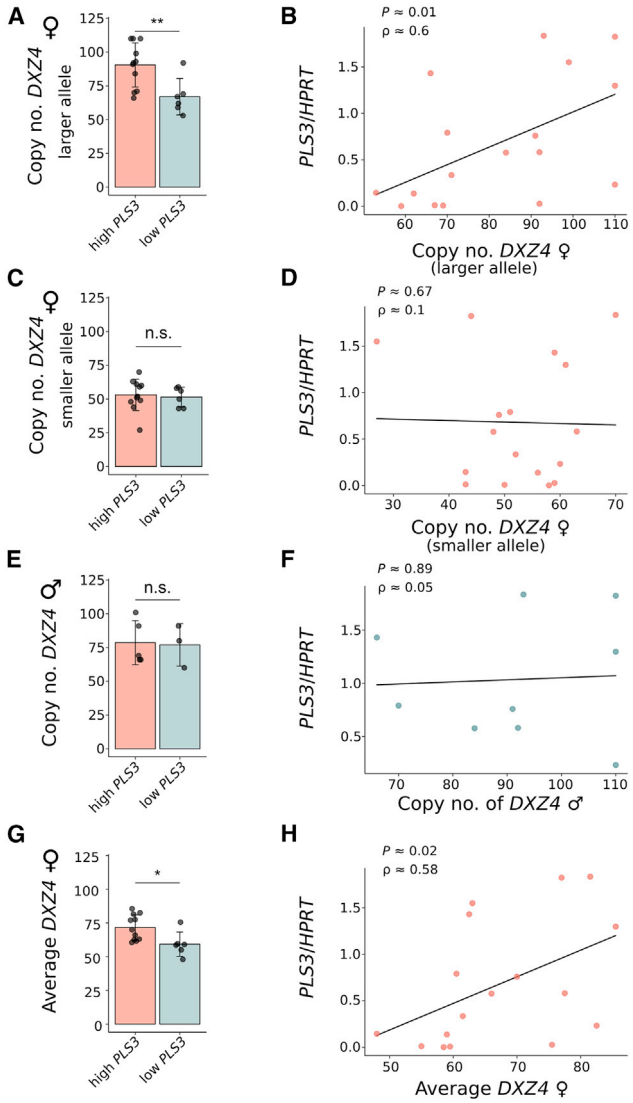


**Figure 4. Linear relationship between the copy number of *DXZ4* and *PLS3* expression in EBV cells**

(A) The copy number of the larger *DXZ4* allele was compared between females with low (n = 6) and high (n = 11) *PLS3* expression.

(B) A significant linear correlation between the *DXZ4* copy number and the *PLS3* levels was found.

(C and D) Comparison of the *DXZ4* copy number (smaller allele) between females with low (n = 6) and high (n = 11) *PLS3* expression showed no significant differences (C) and no linear correlation (D).

(E and F) Male EBV cells showed no difference in the *DXZ4* copy number between *PLS3* high- and low-expressers (E) and no linear correlation was found (F).

(G) The average DXZ4 copy number in females differed significantly between *PLS3* high- and low-expressers.

(F) There was a linear correlation between the average *DXZ4* copy number and the expression of *PLS3*. Data are represented as mean ± SD. The difference in the expression were analyzed by Wilcoxon rank-sum test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). Linear relationships are determined by Spearman's correlation coefficient test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

family 1, the affected male and the asymptomatic sister, inherited the large *DXZ4* allele with 91 repeats from the mother together with the major (T) allele of both SNVs, the minor allele (C) of both SNVs together with the smaller *DXZ4* allele with 49 repeats had to have been inherited from the father (Figure 2C). In family 2, both affected male siblings inherited different smaller *DXZ4* alleles of 60 and 45 repeats, respectively, and the major (T) allele of both SNVs from their mother. Consequently, the two asymptomatic sisters inherited the large *DXZ4* allele with 110 repeats together with the minor (C) allele of both *PLS3* SNVs from the father (Figures 1A and 2C). Since the asymptomatic sister in family 1, in addition to all other asymptomatic females of our cohort, failed to show the minor (C) allele of the *PLS3* SNVs but carried a large *DXZ4* allele, a linkage disequilibrium between the minor PLS3 alleles and a large *DXZ4* allele can be excluded. Also, the direct use of the *PLS3* SNVs as marker for high *PLS3* expression can be excluded.

### CHD4/NuRD is a transcriptional regulator of *PLS3*

Independent of the X-inactivation status of an X-linked gene or the general activity of the chromatin, a gene can be expressed only if a consecutive transcription factor or transcriptional regulator is available. To identify transcription factors that contribute to the regulation of *PLS3*, we performed differential expression analysis of EBV transcriptomes. To reduce biological variability caused by the *DXZ4* copy number and escape of *PLS3* from XCI, we first included only male samples in the analysis (Figure 5A). We compared the four cell lines with the strongest expression of *PLS3* against 11 cell lines with lower or without *PLS3* expression and identified 19 differentially expressed genes, including *PLS3* ($log_2$-fold change > 2; p value < 0.01) and the transcription factor zinc finger homeobox 4 (*ZFHX4* [MIM: 606940]) (Table S6). ZFHX4 is a 397 kDa transcription factor associated with several malignancies.[61] This protein interacts with CHD4, a core member of the NuRD complex, an important epigenetic transcriptional regulator. In the glioblastoma tumor-initiating cell state, ZFHX4 and CHD4/NuRD co-regulate the expression of various genes and ChIP-seq revealed a direct binding of CHD4 to the promoter of *PLS3*.[61]

We decided to further investigate whether ZFHX4 and CHD4/NuRD regulate the expression of *PLS3* in EBV cells and measured the expression of both *ZFXH4* and *CHD4* as well as *PLS3* in seven EBV cell lines by RT-qPCR (Figure 5B). We found a significant linear correlation between *CHD4* and *PLS3* ($p \approx 0.009$, $\rho \approx 0.68$) (Figure 5C) but not between *ZFHX4* and *PLS3* (Figure 5D). *ZFHX4* showed only low expression levels in EBV cells. For these reasons, we decided to concentrate solely on *CHD4* as candidate and performed various experiments to validate CHD4/NuRD as transcriptional regulator of *PLS3*.

Next, we performed siRNA-mediated knock-down of *CHD4* in three male and three female EBV cell lines (AS f1, AS f2, AS f5, SMA m2, SMA m6, SMA m11), which were transfected with either a siRNA against *CHD4* or a mock control. The expression levels of *CHD4* and *PLS3* were measured by RT-qPCR. In addition to that, we

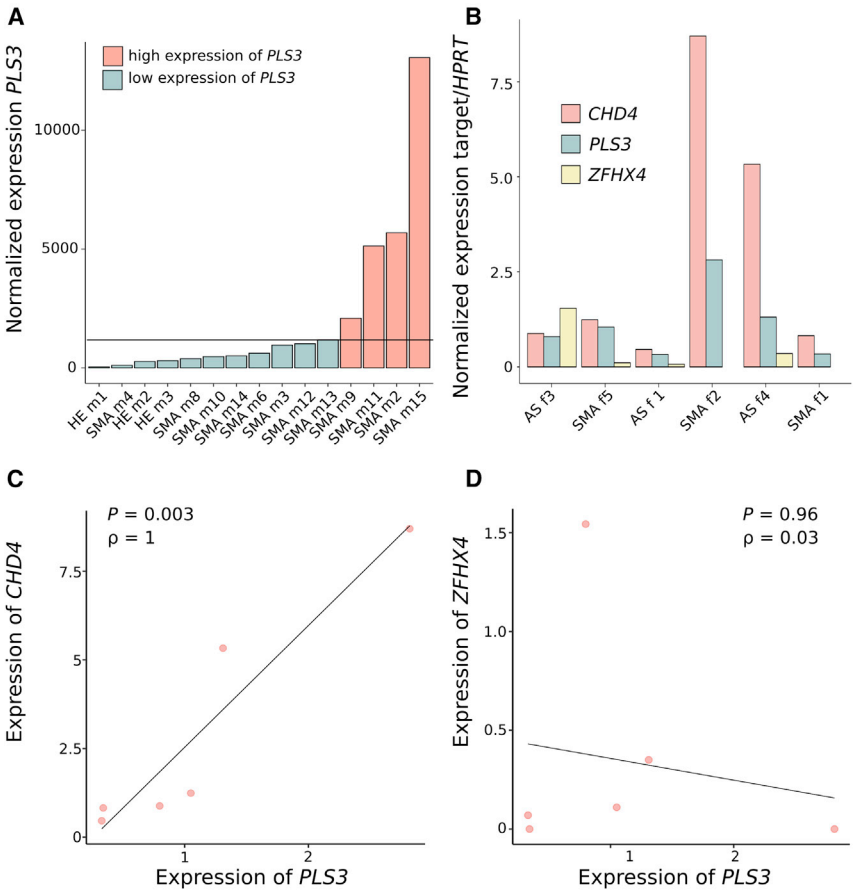


**Figure 5. Identification of CHD4 as transcriptional regulator of *PLS3***

(A) Differential expression analysis of 15 male EBV cell lines with different expression of *PLS3* identified ZFHX4 as putative transcription factor of *PLS3*.

(B) RT-qPCR of the genes *PLS3*, *ZFHX4*, and *CHD4* in seven EBV cell lines.

(C) Significant linear correlation between *CHD4* and *PLS3* expression.

(D) *ZFHX4* levels compared to the expression of *PLS3* show no significant linear correlation. Linear relationships are determined by Spearman's correlation coefficient test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

measured the expression of *TCEAL4*, a facultative escape gene, which was differentially expressed in the spinal MNs (Table S3) and is a known target of CHD4/NuRD.[62] *CHD4* levels were reduced by approximately 50% in all treated cells ($p < 0.01$), while the *PLS3* and *TCEAL4* levels were reduced by 40% ($p < 0.01$) (Figure 6A). We found a significant linear correlation between the expression levels of *CHD4* and *PLS3* ($p \approx 5.905e{-}08$, $\rho \approx 0.978$) (Figure 6B). By this, we were able to show that the downregulation of *CHD4* has a direct impact on the expression of *PLS3*.

Next, we studied the influence of CHD4 overexpression on *PLS3* levels in HEK293T cells. We transfected 125,000 HEK293T cells with increasing concentrations of a *CHD4* expressing plasmid DNA (25, 50, 75, and 100 ng), which was verified by RT-qPCR (Figure 6C). We found a significant increase of *PLS3* levels after application of 75 ng ($p \approx 0.03$) and 100 ng ($p \approx 0.03$) of the overexpression vector (Figure 6D). Overall, we found a significant linear correlation between the *CHD4* and *PLS3* expression levels in HEK293T cells ($p \approx 0.001$, $\rho = 0.65$) (Figure 6E). However, we found a large variability in the expression of *PLS3* at these low concentrations of the overexpression vector. Therefore, we repeated the experiment with higher *CHD4* plasmid concentrations and transfected 450,000 HEK293T cells with 250 ng, 500 ng, 1,000 ng, and 1,500 ng DNA. We found significant differences in the expression of *CHD4* between the lowest vector concentration and the cells that were treated with 1,000 ng ($p \approx 0.004$) as well as 1,500 ng of the vector ($p \approx 0.0009$) (Figure 6F). Furthermore, we found significant differences in the expression of *PLS3* between the lowest concentration and the cells that were treated with 1,000 ng ($p \approx 0.005$) as well as 1,500 ng of the vector ($p \approx 0.0003$) (Figure 6G). A comparison of the *CHD4* and the *PLS3* expression revealed a strong significant linear correlation ($p \approx 0.001$, $\rho = 0.8$) (Figure 6H). Our data clearly show that the expression levels of *CHD4* correlate with *PLS3* levels in EBV and HEK293T cells.

### CHD4 is able to bind DNA outside of the NuRD complex

CHD4/NuRD either activates or represses transcription, depending on the cellular context.[62] However, there is some disagreement in the literature whether CHD4 has DNA-binding capacity outside of the NuRD complex.[63] To investigate this, we performed ChIP-qPCR in three EBV cell lines (SMA f5, AS f2, SMA f2) and in HEK293T cells. The sheared DNA was hybridized with ChIP-grade antibodies. A CTCF antibody was used as a positive control for the chromatin immunoprecipitation and the precipitated DNA was amplified by RT-PCR using primers targeting imprinted maternally expressed noncoding transcript (*H19* [MIM: 103280]) DNA (positive control), and myoglobin (*MB* [MIM:160000]) as a negative control (Figures 7A–7D). The CHD4-immunoprecipitated DNA was amplified using primers targeting the promoters of *PLS3* and *TCEAL4* as well as *MB*. The CHD4 antibody preferentially bound to both *PLS3* and *TCEAL4*, but not to the *MB* control (Figures 7A–7D). Furthermore, we hybridized DNA with an antibody against immunoglobulin heavy constant gamma 1 (IgG, IGHG1 [MIM: 147100]) to control for specificity of the immunoprecipitation. The IgG antibody bound the *PLS3* and *TCEAL4* promoter regions to a low degree, validating the specificity of the CHD4 antibody. Our

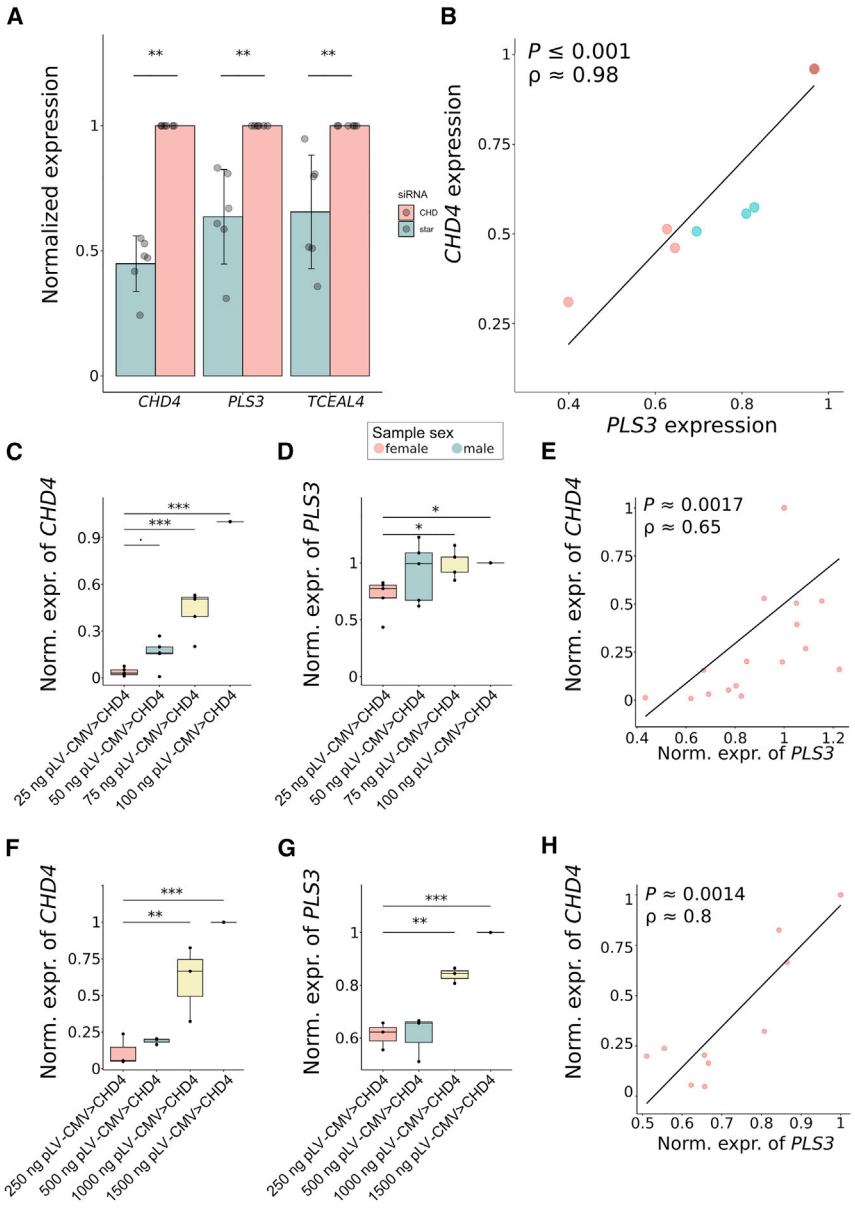


**Figure 6. Validation of CHD4 as transcriptional regulator of *PLS3***

(A) siRNA-mediated knock-down of *CHD4* in EBV cells against a mock control. The differences in the expression were compared by Wilcoxon rank sum tests. Data are represented as mean ± SD.

(B) Comparison of the *CHD4* and *PLS3* expression levels after siRNA-mediated knock-down of *CHD4*. A linear relationship was determined by Spearman's correlation coefficient test.

(C and D) RT-qPCR of 125,000 HEK293T cells treated with different amounts of a CHD4 overexpression vector.

(E) Comparison of the *CHD4* and *PLS3* expression levels in cells transfected with different amounts of a CHD4 expression vector. Linear relationships are determined by Spearman's correlation coefficient test.

(F and G) RT-qPCR of 450,000 HEK293T cells treated with different amounts of a CHD4 expression vector.

(H) Comparison of the *CHD4* and *PLS3* expression levels in cells transfected with different amounts of a CHD4 expression vector.

(C, D, F, and G) The variances were analyzed by an ANOVA followed by a post-hoc Dunnett's test for comparing several treatments with one control. Data are represented as mean ± SD (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

(E and H) Linear relationships are determined by Spearman's correlation coefficient test.

dataset supports the hypothesis that CHD4 is indeed able to bind DNA specifically outside of the NuRD complex, while the promoters of both *PLS3* and *TCEAL4* are targets of CHD4.

### CHD4/NuRD is an activator of *PLS3*

We were able to show that the expression of CHD4 influences the expression of *PLS3* in EBV and HEK293T cells and showed that CHD4 binds the promoter of *PLS3*. However, in the literature it is under debate by which mechanisms CHD4/NuRD regulates gene expression, as the complex is most prominently known as transcriptional repressor.[64] One mechanism that was suggested by several groups is promoter occupancy.[65] By this mechanism, CHD4/NuRD blocks the promoter region so that interaction with other genetic elements is excluded, leading to a repression of the target gene. Activation of a gene would be achieved only by alternative promoters. However, other groups hypothesized that CHD4/NuRD positively regulates expression of several target genes by direct interaction with specific transcription factors. Hereby, we analyzed whether the interaction of CHD4 with the promoter of *PLS3* positively regulates the expression to exclude the possibility that the regulation is due to promoter occupancy. We performed dual-luciferase-promoter assays in 125,000 HEK293T cells. The cells in all sample groups were co-transfected with 0.25 ng of a Renilla-luciferase reporter construct as internal reference. All treatment groups were co-transfected with 20 ng of a Firefly-luciferase vector under a *PLS3* promoter (1,022 bp) and sequential concentrations (0 ng, 6.25 ng, 12.5 ng, and 25 ng) of the *CHD4* overexpression vector under a *CMV* promoter, which has been previously established in the overexpression assays. The positive control was co-transfected with 50 ng of a Firefly-luciferase vector under a *CMV* promoter. One negative control was co-transfected with 10 ng of a Firefly-luciferase vector without a promoter sequence (empty control). A second negative control was transfected only with 0.25 ng of the Renilla-luciferase vector as internal reference. We compared the endogenous

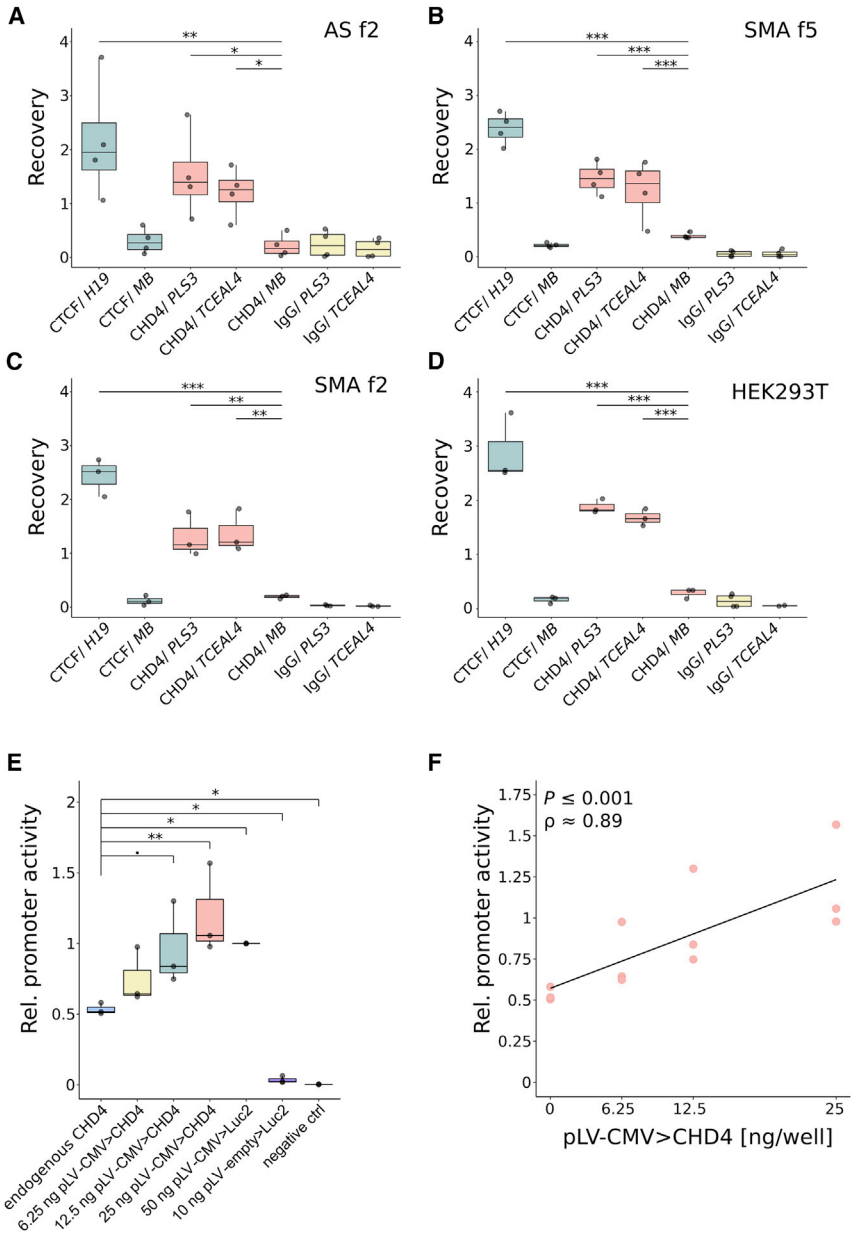


**Figure 7. CHD4 interacts with the promoter of *PLS3* and activates transcription**

(A–D) Chromatin immunoprecipitation of CHD4 in three EBV cell lines and HEK293T cells. The CHD4 antibody preferentially bound to both *PLS3* and *TCEAL4* but not to *MB* (negative control). The mean ± SD for three independent replicates is given. The variances were analyzed by an ANOVA followed by a post-hoc Dunnett's test for comparing several treatments with one control. Data are represented as mean ± SD (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

(E) Activity of the *PLS3* promoter in a dual-luciferase-promoter assay. HEK293T cells were co-transfected with a Firefly-luciferase vector under a *PLS3* promoter and various concentrations of a *CHD4* overexpression vector. The mean ± SD for three independent replicates is given. The variances were analyzed by an ANOVA followed by a post-hoc Dunnett's test for comparing several treatments with one control (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). (D) A direct comparison of the relative promoter activity of *PLS3* with the amount of CHD4 overexpression vector. Linear relationships are determined by Spearman's correlation coefficient test.

CHD4 expression (0 ng CHD4 overexpression vector) to the three treatment groups. The relative promoter activity was significantly increased by about 50% after transfection with 25 ng of the CHD4 overexpression vector (Figure 7E). Generally, we obtained an increase of the relative *PLS3* promoter activity with rising concentrations of CHD4 and found a significant linear correlation ($p \approx 0.001$, $\rho \approx 0.89$) (Figure 7F).

In conclusion, our data indicate that CHD4 is directly interacting with the *PLS3* promoter. This interaction positively regulates the expression of *PLS3*.

## Discussion

Overall, we show here that *PLS3* is able to escape X chromosome inactivation in spinal MNs. Using molecular combing on stretched DNA fibers, we determined the copy number of the macrosatellite *DXZ4*, which is essential for X chromosome inactivation and localized 500 kb apart from *PLS3*. We show that females with increased expression of *PLS3,* including *SMN1*-deleted asymptomatic females, carry at least one large *DXZ4* allele (>70 repeats), while low *PLS3* expressers carry two small *DXZ4* alleles. The expression of *PLS3* correlates with the *DXZ4* copy number in females, but not in males. In addition to that, we identified CHD4/NuRD as an epigenetic transcriptional regulator of *PLS3*. The motor protein CHD4 directly interacts with the *PLS3* promotor to regulate gene expression.

### *PLS3* escapes from XCI

Approximately 15% of X-linked genes escape XCI and are bi-allelically expressed. Another 15% of genes variably escape in a tissue-specific manner. One of those facultative escape genes is *PLS3*.[25–28] Indeed, we show here that *PLS3* escapes XCI in spinal MNs. The expression levels of *PLS3* in MNs are approximately doubled in asymptomatic females in comparison to affected male siblings, which could be explained by the escape of *PLS3* from XCI and is underlined by the average *PLS3* SNV ratio of 0.4–0.6 (Figure S3C), showing that both alleles are transcribed at similar proportions. It is not entirely clear why *PLS3* shows strong

expression levels in FBs, without differential expression between asymptomatic and SMA-affected individuals. One explanation might be that *PLS3* escapes XCI in FBs, but that this escape is masked by the strong expression levels and limited by the availability of a specific transcription factor that interacts with CHD4/NuRD. In that case, the transcription of *PLS3* mRNA reached its maximum capacity and, therefore, a difference in the expression of the gene would not be measurable. On the other hand, it is known that the NuRD complex also interacts with other CHD isoforms, such as CHD3.[62] This interaction is tissue specific and it could be that there is a stronger interaction between CHD3/NuRD and the *PLS3* promoter. Another possibility would be that CHD3/NuRD interacts with other specific transcription factors that do not interact with the escaped *PLS3* promoter. It is also likely that the chromosomal compartmentalization in FBs differs from MNs and no escape from XCI takes place.

In recent years it became clear that our genome is partitioned into megabase-scaled highly self-interacting regions called topologically associating domains (TADs).[66] TADs are separated from each other by TAD boundaries, which suppress interactions between different TADs and are highly conserved among species, cell types, and tissues.[66,67] However, the $X_i$ differs from the $X_a$ and is partitioned into two massive megadomains (0–115 Mb and 115–155.3 Mb) of high self-interaction. *DXZ4* and *PLS3* are located at Xq23, near the boundary between these two megadomains (ChrX: 114,867,433–114,919,088).[68] We found a significant correlation between the copy number of the macrosatellite and the expression of *PLS3* in EBV cells in females (Figures 4A–4D). Of note, in 95% of individuals, *PLS3* is not expressed at all or at an extremely low level in the hematopoietic system. Even in those 5% of individuals who express *PLS3* in blood, this is very low in comparison to the *PLS3* expression in fibroblasts or iPSC-derived spinal MNs.[12,24] *DXZ4* consists of hypomethylated open chromatin marks on the $X_i$.[32,34] This chromosomal conformation could explain the correlation between the *DXZ4* copy number and the expression levels of *PLS3*, as the sheer size of the macrosatellite would influence the localization of the genomic locus in the nucleus. Thereby, the copy number of the macrosatellite may influence the expression of neighboring genes, such as *PLS3*. Furthermore, it is known that open chromatin locates on the surface of the Barr body, while heterochromatin locates to the center of the $X_i$.[26] If this would indeed be the case, it would mean that the escape of *PLS3* is favored, if the $X_i$ harbors a *DXZ4* allele with a high copy number. This mechanism would also explain why there is no correlation between the expression of *PLS3* and the *DXZ4* copy number in males, as males are hemizygous and harbor only one (active) X chromosome.

We found increased expression of *PLS3* in EBV cells, which had at least one *DXZ4* allele of increased copy number. Furthermore, there was a significant difference in the absolute copy number between high and low expressers. Low expressers harbor usually two *DXZ4* alleles with similar copy numbers or smaller than about 70 repeats in our datasets. Recently, a bioinformatics tool, DeepLoop, was published, which is able to enhance chromatin interaction mapping by applying bias correction and deep-learning-based signal enhancement. This tool was used to analyze a dataset from GM12878 cells. The paternal X chromosome is inactivated in this dataset. Their data support that the *DXZ4* locus on the $X_i$ escaped XCI. They conclude that the escape regions near *DXZ4* are mechanistically coupled to the formation of the megadomains.[69] This finding supports our hypothesis that the escape of *PLS3* and other neighboring genes is indeed associated with the macrosatellite *DXZ4*.

### CHD4/NuRD as epigenetic regulator of *PLS3*

The fact that we identified rare cases of *PLS3* overexpression in EBV cell lines of males was enigmatic and pointed toward additional layers of *PLS3* regulation. We identified CHD4 as epigenetic transcriptional regulator of *PLS3* and found co-regulation of the two genes in EBV and HEK293T cells (Figures 6 and 7). Our data indicate that, in contrast to the *DXZ4* copy number, the regulation of *PLS3* by CHD4/NuRD is not sex-specific. ZFHX4, a coregulator of CHD4 and the CHD4/NuRD complex, was identified as transcriptional regulator of *PLS3* by the analysis of only male EBV transcriptomes (Table S6). This was done to exclude the influence of bi-allelic expression by escape of *PLS3* from XCI. However, subsequently, we investigated the *ZFHX4* and *CHD4* expression in relation to *PLS3* by RT-PCR analysis in EBV cells of both sexes. In addition, we knocked-down *CHD4* in three males and three females to verify the regulatory effect of CHD4. In both sexes, the decrease in the expression of *CHD4* led to a decrease in the expression of *PLS3*. Overall, we found a significant linear correlation. It is of course likely that the escape from XCI in females increases the amount of *PLS3* mRNA in addition to CHD4/NuRD. However, the regulation of *PLS3* expression by CHD4/NuRD is not *per se* sex specific as the knock-down worked in both sexes in a similar proportion (Figure 6B). One important aspect is that male SMA-affected siblings of asymptomatic females also show half-dose *PLS3* expression in iPSC-derived MNs and even very low in EBV cells, indicating that other transcription and/or epigenetic factors influence cell type, specifically *PLS3* expression.

CHD4 most likely acts as part of the NuRD complex.[35,70–73] The core proteins of NuRD contain several subunits including the metastasis-associated proteins MTA1/2/3 (*MTA1* [MIM: 603526], *MTA2* [MIM: 603497], and *MTA3* [MIM: 609050]), the methyl-CpG binding domain proteins MBD2/3 (*MBD2* [MIM: 603547] and *MBD3* [MIM: 603573]), the histone deacetylases HDAC1/2 (*HDAC1* [MIM: 601241] and *HDAC2* [MIM: 605164]), the retinoblastoma binding proteins RBBP4/7 (*RBBP4* [MIM: 602923] and *RBBP7* [MIM: 300825]), the GATA zinc finger domain containing proteins GATA2a/2b (*GATA2* [MIM: 137295]) as well as the motor subunits CHD3/4/5 (*CHD3* [MIM: 602120] and

*CHD5* [MIM: 610771]). MBD2/3 has the capacity to selectively recognize methylated DNA.[74] Indeed, in a ChIP-seq dataset in human breast cancer cell lines, MBD3 overlapped with the promoter of *PLS3,* indicating that CHD4/NuRD and MBD3 act together to activate or repress genes.[75] CHD3 and CHD4 are responsible for the ATPase activity of the NuRD complex and harbor conserved PHD fingers, chromodomains, and a DNA-binding domain.[76] The CHD4/NuRD complex was originally described as transcriptional silencer.[73,77,78] However, multiple studies have shown that CHD4/NuRD either activates or represses transcription depending on the cellular context.[62,79] It is not clear how CHD4/NuRD regulates transcription and it is sometimes stated that CHD4 misses DNA binding capacity.[63] However, several ChIP-seq and ChIP-qPCR studies have shown that the protein has DNA-binding capacity independent of other components of the NuRD complex or transcription factors.[80] A ChIP-seq in therapy-resistant tumor-initiating glioblastoma cells revealed a direct binding of CHD4 to the *PLS3* promoter.[61] These results are in line with our ChIP-RT-qPCR experiments, in which we showed that the protein is able to directly bind to the promoter region of *PLS3* (Figures 7A–7D). Furthermore, a more recent study indicates that mutations in *CHD4* may disrupt DNA binding activity of the protein underlining the ability to directly interact with DNA *in vivo*.[81] Our dual-luciferase-promoter assays showed that CHD4 is sufficient to activate transcription of a reporter gene (Figures 7E and 7F). One mechanistic explanation would be that a change in the CHD4 level changes the stoichiometry between CHD4/NuRD and endogenous CHD3/NuRD. It is known that the subunit composition of NuRD is tissue specific, while the isoforms containing either *CHD3* or *CHD4* have a distinct nuclear localization and a distinct set of target genes.[62,77] Most likely, CHD4/NuRD activates transcription of target genes by recruitment-specific transcription factors.[82] Promoter hypomethylation and overexpression of *PLS3* are hallmarks of several cancers.[1] In Sézary syndrome, circulating $CD4^+$ T cells show increased expression of PLS3, TWIST1 (MIM: 601622), and GATA6 (MIM: 601656) compared to normal $CD4^+$ T cells.[10,11] The promoter regions of all three genes have been found to be hypomethylated in Sézary syndrome $CD4^+$ T cells, indicating an epigenetic mechanism of gene regulation.[83] Interestingly, TWIST1 and GATA6 are known to be regulated by CHD4/NuRD.[84,85]

One limitation of our study is that we do not know which specific transcription factors interact with CHD4/NuRD. In addition to that, we do not understand which mechanisms regulated the expression of *CHD4* in the tested cell types. One challenge is that multiple mechanisms of *PLS3* regulation work simultaneously. To reduce the number of parameters and to examine the influence of CHD4 without the effect of changes of the X-inactivation status, we performed knock-down experiments in EBV cells of both sexes. We assumed that the X-inactivation status would not change between *CHD4* and mock siRNA-transfected cells. The X-inactivation status would only influence the *PLS3* expression in females, as males are hemizygous for *PLS3* and *DXZ4*. siRNA-mediated knockdown of *CHD4* in EBV cells confirmed our finding that CHD4 is an epigenetic regulator of *PLS3*.

The transcriptional regulation of *PLS3* seems to be highly complex. Various epigenetic mechanisms, such as escape from XCI, the copy number of *DXZ4*, and the epigenetic regulator CHD4/NuRD, seem to influence the *PLS3* expression independently of each other or in combination. Over the last decade, the importance of PLS3 as genetic modifier was not only shown in neuromuscular disorders, it was also suggested as important biomarker in several malignancies, as well as osteoarthritis and in CAKUT.[5,86,87] To unravel the complex regulation of PLS3 may facilitate our understanding of disease pathologies in this wide range of disorders. Especially in the cancer field, where PLS3 is involved in epithelial-mesenchymal transition, the gene regulation may be crucial to develop novel treatment strategies.

## Data and code availability

The next-generation sequencing datasets and R-scripts supporting the current study have not been deposited in a public repository but are available from the corresponding author upon reasonable request.

## Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.02.004.

## Acknowledgments

We would like to thank all the individuals with SMA and their families that participated in our study. Furthermore, we thank Genomic Vision for their support and the design of the molecular combing fiber probes. We thank deCODE genetics, Iceland for next-generation sequencing the genomes and transcriptomes of all individuals in this study. Finally, we thank DIAGENODE for their incredible help with the optimization of our ChIP experiments.

This work was supported by the German Research Foundation (Wi 945/17-1 [project-ID 398410809]; Wi 945/18-1 [project ID 384170921]; Wi 945/19-1 [project-ID 417989143]; SFB1451 [project-ID 431549029-A01], KY96/1-2 [project ID 269018619] and GRK1960 [project ID 233886668]), the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 2012-305121 "Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)," and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956185 (SMABEYOND), the Center for Molecular Medicine Cologne (project No C18) to B.W.

## Author contributions

E.A.S. designed the experiments, analyzed all data, and wrote the manuscript. I.H. helped with all the experiments and performed

the ChIP experiments. N.T. performed the molecular combing experiments. J.A. established and supervised the molecular combing method, J.C. and C.M. performed the differentiation of the iPSCs into spinal motor neurons, S.H. performed the RNA and protein isolation from MNs, and B.W. conceptualized and supervised the project and wrote and edited the manuscript. All authors read and edited the manuscript.

## Declaration of interests

The authors declare no competing interests.

Received: July 14, 2022
Accepted: February 3, 2023
Published: February 21, 2023

## Web resources

GeneCards, genecards.org
gnomAD, https://gnomad.broadinstitute.org
LDlink, https://ldlink.nci.nih.gov/
Oligocalc, http://biotools.nubic.northwestern.edu/OligoCalc.html
OMIM, https://www.omim.org/
SNPcheck, https://genetools.org/SNPCheck/snpcheck.htm
UCSC genome browser, https://genome.ucsc.edu/

## References

1. Wolff, L., Strathmann, E.A., Müller, I., Mählich, D., Veltman, C., Niehoff, A., and Wirth, B. (2021). Plastin 3 in health and disease: a matter of balance. Cell. Mol. Life Sci. *78*, 5275–5301. https://doi.org/10.1007/s00018-021-03843-5.
2. van Dijk, F.S., Zillikens, M.C., Micha, D., Riessland, M., Marcelis, C.L.M., de Die-Smulders, C.E., Milbradt, J., Franken, A.A., Harsevoort, A.J., Lichtenbelt, K.D., et al. (2013). PLS3 mutations in X-linked osteoporosis with fractures. N. Engl. J. Med. *369*, 1529–1536. https://doi.org/10.1056/NEJMoa1308223.
3. Tsolis, K.C., Bei, E.S., Papathanasiou, I., Kostopoulou, F., Gkretsi, V., Kalantzaki, K., Malizos, K., Zervakis, M., Tsezou, A., and Economou, A. (2015). Comparative proteomic analysis of hypertrophic chondrocytes in osteoarthritis. Clin. Proteomics *12*, 12. https://doi.org/10.1186/s12014-015-9085-6.
4. Mählich, D., Glasmacher, A., Müller, I., Oppermann, J., Grevenstein, D., Eysel, P., Heilig, J., Wirth, B., Zaucke, F., and Niehoff, A. (2021). Expression and localization of thrombospondins, plastin 3, and STIM1 in different cartilage compartments of the osteoarthritic varus knee. Int. J. Mol. Sci. *22*, 1–19. https://doi.org/10.3390/ijms22063073.
5. Fédou, C., Camus, M., Lescat, O., Feuillet, G., Mueller, I., Ross, B., Buléon, M., Neau, E., Alves, M., Goudounéche, D., et al. (2021). Mapping of the amniotic fluid proteome of fetuses with congenital anomalies of the kidney and urinary tract identifies plastin 3 as a protein involved in glomerular integrity. J. Pathol. *254*, 575–588. https://doi.org/10.1002/path.5703.
6. Kujawski, R., Przybyłowska-Sygut, K., Mik, M., Lewandowski, M., Trzciński, R., Berut, M., Dziki, Ł., Majsterek, I., and Dziki, A. (2015). Expression of the PLS3 gene in circulating cells in patients with colorectal cancer. Pol. Przegl. Chir. *87*, 59–64. https://doi.org/10.1515/pjs-2015-0020.
7. Sugimachi, K., Yokobori, T., Iinuma, H., Ueda, M., Ueo, H., Shinden, Y., Eguchi, H., Sudo, T., Suzuki, A., Maehara, Y., et al. (2014). Aberrant expression of plastin-3 via copy number gain induces the epithelial-mesenchymal transition in circulating colorectal cancer cells. Ann. Surg Oncol. *21*, 3680–3690. https://doi.org/10.1245/s10434-013-3366-y.
8. Yokobori, T., Iinuma, H., Shimamura, T., Imoto, S., Sugimachi, K., Ishii, H., Iwatsuki, M., Ota, D., Ohkuma, M., Iwaya, T., et al. (2013). Plastin3 is a novel marker for circulating tumor cells undergoing the epithelial-mesenchymal transition and is associated with colorectal cancer prognosis. Cancer Res. *73*, 2059–2069. https://doi.org/10.1158/0008-5472.CAN-12-0326.
9. Ueo, H., Sugimachi, K., Gorges, T.M., Bartkowiak, K., Yokobori, T., Müller, V., Shinden, Y., Ueda, M., Ueo, H., Mori, M., et al. (2015). Circulating tumour cell-derived plastin3 is a novel marker for predicting long-term prognosis in patients with breast cancer. Br. J. Cancer *112*, 1519–1526. https://doi.org/10.1038/bjc.2015.132.
10. van Doorn, R., Dijkman, R., Vermeer, M.H., Out-Luiting, J.J., van der Raaij-Helmer, E.M.H., Willemze, R., and Tensen, C.P. (2004). Aberrant expression of the tyrosine kinase receptor EphA4 and the transcription factor twist in Sezary syndrome identified by gene expression analysis. Cancer Res. *64*, 5578–5586. https://doi.org/10.1158/0008-5472.CAN-04-1253.
11. Wang, Y., Su, M., Zhou, L.L., Tu, P., Zhang, X., Jiang, X., and Zhou, Y. (2011). Deficiency of SATB1 expression in Sezary cells causes apoptosis resistance by regulating FasL/CD95L transcription. Blood *117*, 3826–3835. https://doi.org/10.1182/blood-2010-07-294819.
12. Oprea, G.E., Kröber, S., McWhorter, M.L., Rossoll, W., Müller, S., Krawczak, M., Bassell, G.J., Beattie, C.E., and Wirth, B. (2008). Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. Science *320*, 524–527. https://doi.org/10.1126/science.1155085.
13. Janzen, E., Wolff, L., Mendoza-Ferreira, N., Hupperich, K., Delle Vedove, A., Hosseinibarkooie, S., Kye, M.J., and Wirth, B. (2019). PLS3 overexpression delays ataxia in chp1 mutant mice. Front. Neurosci. *13*, 993. https://doi.org/10.3389/fnins.2019.00993.
14. Walsh, M.B., Janzen, E., Wingrove, E., Hosseinibarkooie, S., Muela, N.R., Davidow, L., Dimitriadi, M., Norabuena, E.M., Rubin, L.L., Wirth, B., and Hart, A.C. (2020). Genetic modifiers ameliorate endocytic and neuromuscular defects in a model of spinal muscular atrophy. BMC Biol. *18*, 127. https://doi.org/10.1186/s12915-020-00845-w.
15. Wirth, B. (2021). Spinal muscular atrophy: in the challenge lies a solution. Trends Neurosci. *44*, 306–322. https://doi.org/10.1016/j.tins.2020.11.009.
16. Feldkötter, M., Schwarzer, V., Wirth, R., Wienker, T.F., and Wirth, B. (2002). Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. Am. J. Hum. Genet. *70*, 358–368. https://doi.org/10.1086/338627.
17. Cartegni, L., and Krainer, A.R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat. Genet. *30*, 377–384.
18. Helmken, C., Hofmann, Y., Schoenen, F., Oprea, G., Raschke, H., Rudnik-Schöneborn, S., Zerres, K., and Wirth, B. (2003). Evidence for a modifying pathway in SMA discordant families: reduced SMN level decreases the amount of its interacting partners and Htra2-beta1. Hum. Genet. *114*, 11–21. https://doi.org/10.1007/s00439-003-1025-2.

19. Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc. Natl. Acad. Sci. USA *96*, 6307–6311. https://doi.org/10.1073/pnas.96.11.6307.
20. Hosseinibarkooie, S., Peters, M., Torres-Benito, L., Rastetter, R.H., Hupperich, K., Hoffmann, A., Mendoza-Ferreira, N., Kaczmarek, A., Janzen, E., Milbradt, J., et al. (2016). The power of human protective modifiers: PLS3 and CORO1C unravel impaired endocytosis in spinal muscular atrophy and rescue SMA phenotype. Am. J. Hum. Genet. *99*, 647–665. https://doi.org/10.1016/j.ajhg.2016.07.014.
21. Dimitriadi, M., Sleigh, J.N., Walker, A., Chang, H.C., Sen, A., Kalloo, G., Harris, J., Barsby, T., Walsh, M.B., Satterlee, J.S., et al. (2010). Conserved genes act as modifiers of invertebrate SMN loss of function defects. PLoS Genet. *6*, e1001172. https://doi.org/10.1371/journal.pgen.1001172.
22. Lin, C.S., Lau, A., Huynh, T., and Lue, T.F. (1999). Differential regulation of human T-plastin gene in leukocytes and non-leukocytes: identification of the promoter, enhancer, and CpG island. DNA Cell Biol. *18*, 27–37. https://doi.org/10.1089/104454999315592.
23. Chafel, M.M., Shen, W., and Matsudaira, P. (1995). Sequential expression and differential localization of I-L-and T-fimbrin during differentiation of the mouse intestine and yolk sac. Dev. Dyn. *203*, 141–151. https://doi.org/10.1002/aja.1002030203.
24. Heesen, L., Peitz, M., Torres-Benito, L., Hölker, I., Hupperich, K., Dobrindt, K., Jungverdorben, J., Ritzenhofen, S., Weykopf, B., Eckert, D., et al. (2016). Plastin 3 is upregulated in iPSC-derived motoneurons from asymptomatic SMN1-deleted individuals. Cell. Mol. Life Sci. *73*, 2089–2104. https://doi.org/10.1007/s00018-015-2084-y.
25. Balaton, B.P., Cotton, A.M., and Brown, C.J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol. Sex Differ. *6*, 35. https://doi.org/10.1186/s13293-015-0053-7.
26. Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404. https://doi.org/10.1038/nature03479.
27. Cotton, A.M., Ge, B., Light, N., Adoue, V., Pastinen, T., and Brown, C.J. (2013). Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. Genome Biol. *14*, R122. https://doi.org/10.1186/gb-2013-14-11-r122.
28. Yang, F., Babak, T., Shendure, J., and Disteche, C.M. (2010). Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. *20*, 614–622. https://doi.org/10.1101/gr.103200.109.
29. Horakova, A.H., Calabrese, J.M., McLaughlin, C.R., Tremblay, D.C., Magnuson, T., and Chadwick, B.P. (2012). The mouse DXZ4 homolog retains Ctcf binding and proximity to Pls3 despite substantial organizational differences compared to the primate macrosatellite. Genome Biol. *13*, R70. https://doi.org/10.1186/gb-2012-13-8-r70.
30. Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C., and Chadwick, B.P. (2012). The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. Hum. Mol. Genet. *21*, 4367–4377. https://doi.org/10.1093/hmg/dds270.
31. Bonora, G., Deng, X., Fang, H., Ramani, V., Qiu, R., Berletch, J.B., Filippova, G.N., Duan, Z., Shendure, J., Noble, W.S., and Disteche, C.M. (2018). Orientation-dependent Dxz4 contacts shape the 3D structure of the inactive X chromosome. Nat. Commun. *9*, 1445. https://doi.org/10.1038/s41467-018-03694-y.
32. Giacalone, J., Friedes, J., and Francke, U. (1992). A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. Nat. Genet. *1*, 137–143. https://doi.org/10.1038/ng0592-137.
33. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature *585*, 79–84. https://doi.org/10.1038/s41586-020-2547-7.
34. Chadwick, B.P. (2008). DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. Genome Res. *18*, 1259–1269. https://doi.org/10.1101/gr.075713.107.
35. Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell *95*, 279–289. https://doi.org/10.1016/s0092-8674(00)81758-4.
36. Zhong, Y., Paudel, B.P., Ryan, D.P., Low, J.K.K., Franck, C., Patel, K., Bedward, M.J., Torrado, M., Payne, R.J., van Oijen, A.M., and Mackay, J.P. (2020). CHD4 slides nucleosomes by decoupling entry- and exit-side DNA translocation. Nat. Commun. *11*, 1519. https://doi.org/10.1038/s41467-020-15183-2.
37. Maury, Y., Côme, J., Piskorowski, R.A., Salah-Mohellibi, N., Chevaleyre, V., Peschanski, M., Martinat, C., and Nedelec, S. (2015). Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. Nat. Biotechnol. *33*, 89–96. https://doi.org/10.1038/nbt.3049.
38. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.
39. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
40. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
41. 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. https://doi.org/10.1186/s13059-014-0550-8.
42. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.
43. Lee, S., Lee, S., Ouellette, S., Park, W.Y., Lee, E.A., and Park, P.J. (2017). NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. Nucleic Acids Res. *45*, e103. https://doi.org/10.1093/nar/gkx193.
44. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45. https://doi.org/10.1093/nar/29.9.e45.
45. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome

Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

46. Nguyen, K., Walrafen, P., Bernard, R., Attarian, S., Chaix, C., Vovan, C., Renard, E., Dufrane, N., Pouget, J., Vannier, A., et al. (2011). Molecular combing reveals allelic combinations in facioscapulohumeral dystrophy. Ann. Neurol. *70*, 627–633. https://doi.org/10.1002/ana.22513.
47. Heesen, L., Peitz, M., Dimos, J.T., Brustle, O., and Wirth, B. (2010). Generation and application of iPS cells & motoneurons for tissue-specific analysis of SMN-specific changes in SMA patients. Transgenic Res. *19*, 332.
48. Geens, M., and Chuva De Sousa Lopes, S.M. (2017). X chromosome inactivation in human pluripotent stem cells as a model for human development: back to the drawing board? Hum. Reprod. Update *23*, 520–532. https://doi.org/10.1093/humupd/dmx015.
49. Balaton, B.P., and Brown, C.J. (2016). Escape artists of the X chromosome. Trends Genet. *32*, 348–359. https://doi.org/10.1016/j.tig.2016.03.007.
50. Calabrese, J.M., Sun, W., Song, L., Mugford, J.W., Williams, L., Yee, D., Starmer, J., Mieczkowski, P., Crawford, G.E., and Magnuson, T. (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell *151*, 951–963. https://doi.org/10.1016/j.cell.2012.10.037.
51. Zhang, Y., Castillo-Morales, A., Jiang, M., Zhu, Y., Hu, L., Urrutia, A.O., Kong, X., and Hurst, L.D. (2013). Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. Mol. Biol. Evol. *30*, 2588–2601. https://doi.org/10.1093/molbev/mst148.
52. Berletch, J.B., Ma, W., Yang, F., Shendure, J., Noble, W.S., Disteche, C.M., and Deng, X. (2015). Escape from X inactivation varies in mouse tissues. PLoS Genet. *11*, e1005079. https://doi.org/10.1371/journal.pgen.1005079.
53. Goto, Y., and Kimura, H. (2009). Inactive X chromosome-specific histone H3 modifications and CpG hypomethylation flank a chromatin boundary between an X-inactivated and an escape gene. Nucleic Acids Res. *37*, 7416–7428. https://doi.org/10.1093/nar/gkp860.
54. Kucera, K.S., Reddy, T.E., Pauli, F., Gertz, J., Logan, J.E., Myers, R.M., and Willard, H.F. (2011). Allele-specific distribution of RNA polymerase II on female X chromosomes. Hum. Mol. Genet. *20*, 3964–3973. https://doi.org/10.1093/hmg/ddr315.
55. Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905. https://doi.org/10.1126/science.1237905.
56. Murakami, K., Ohhira, T., Oshiro, E., Qi, D., Oshimura, M., and Kugoh, H. (2009). Identification of the chromatin regions coated by non-coding Xist RNA. Cytogenet. Genome Res. *125*, 19–25. https://doi.org/10.1159/000207514.
57. Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature *523*, 212–216. https://doi.org/10.1038/nature14465.
58. Sharp, A.J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S.B., Dupre, Y., and Antonarakis, S.E. (2011). DNA methylation profiles of human active and inactive X chromosomes. Genome Res. *21*, 1592–1600. https://doi.org/10.1101/gr.112680.110.
59. Kakuda, K., Ikenaka, K., Araki, K., So, M., Aguirre, C., Kajiyama, Y., Konaka, K., Noi, K., Baba, K., Tsuda, H., et al. (2019). Ultrasonication-based rapid amplification of alpha-synuclein aggregates in cerebrospinal fluid. Sci. Rep. *9*, 6001. https://doi.org/10.1038/s41598-019-42399-0.
60. Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J.S., and Bensimon, A. (1997). Dynamic molecular combing: stretching the whole human genome for high-resolution studies. Science *277*, 1518–1523. https://doi.org/10.1126/science.277.5331.1518.
61. Chudnovsky, Y., Kim, D., Zheng, S., Whyte, W.A., Bansal, M., Bray, M.A., Gopal, S., Theisen, M.A., Bilodeau, S., Thiru, P., et al. (2014). ZFHX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. Cell Rep. *6*, 313–324. https://doi.org/10.1016/j.celrep.2013.12.032.
62. Hoffmeister, H., Fuchs, A., Erdel, F., Pinz, S., Gröbner-Ferreira, R., Bruckmann, A., Deutzmann, R., Schwartz, U., Maldonado, R., Huber, C., et al. (2017). CHD3 and CHD4 form distinct NuRD complexes with different yet overlapping functionality. Nucleic Acids Res. *45*, 10534–10554. https://doi.org/10.1093/nar/gkx711.
63. Arends, T., Dege, C., Bortnick, A., Danhorn, T., Knapp, J.R., Jia, H., Harmacek, L., Fleenor, C.J., Straign, D., Walton, K., et al. (2019). CHD4 is essential for transcriptional repression and lineage progression in B lymphopoiesis. Proc. Natl. Acad. Sci. USA *116*, 10927–10936. https://doi.org/10.1073/pnas.1821301116.
64. Bowen, N.J., Fujita, N., Kajita, M., and Wade, P.A. (2004). Mi-2/NuRD: multiple complexes for many purposes. Biochim. Biophys. Acta *1677*, 52–57. https://doi.org/10.1016/j.bbaexp.2003.10.010.
65. Yamada, T., Yang, Y., Hemberg, M., Yoshida, T., Cho, H.Y., Murphy, J.P., Fioravante, D., Regehr, W.G., Gygi, S.P., Georgopoulos, K., and Bonni, A. (2014). Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. Neuron *83*, 122–134. https://doi.org/10.1016/j.neuron.2014.05.039.
66. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. https://doi.org/10.1038/nature11082.
67. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385. https://doi.org/10.1038/nature11049.
68. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.
69. Zhang, S., Plummer, D., Lu, L., Cui, J., Xu, W., Wang, M., Liu, X., Prabhakar, N., Shrinet, J., Srinivasan, D., et al. (2022). DeepLoop robustly maps chromatin interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution. Nat. Genet. *54*, 1013–1025. https://doi.org/10.1038/s41588-022-01116-w.

70. Tong, J.K., Hassig, C.A., Schnitzler, G.R., Kingston, R.E., and Schreiber, S.L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. Nature *395*, 917–921. https://doi.org/10.1038/27699.
71. Wade, P.A., Jones, P.L., Vermaak, D., and Wolffe, A.P. (1998). A multiple subunit Mi-2 histone deacetylase from Xenopus laevis cofractionates with an associated Snf2 superfamily ATPase. Curr. Biol. *8*, 843–846. https://doi.org/10.1016/s0960-9822(98)70328-8.
72. Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Côté, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol. Cell *2*, 851–861. https://doi.org/10.1016/s1097-2765(00)80299-3.
73. Marfella, C.G.A., and Imbalzano, A.N. (2007). The Chd family of chromatin remodelers. Mutat. Res. *618*, 30–40. https://doi.org/10.1016/j.mrfmmm.2006.07.012.
74. Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell Biol. *18*, 6538–6547. https://doi.org/10.1128/MCB.18.11.6538.
75. Shimbo, T., Du, Y., Grimm, S.A., Dhasarathy, A., Mav, D., Shah, R.R., Shi, H., and Wade, P.A. (2013). MBD3 localizes at promoters, gene bodies and enhancers of active genes. PLoS Genet. *9*, e1004028. https://doi.org/10.1371/journal.pgen.1004028.
76. Woodage, T., Basrai, M.A., Baxevanis, A.D., Hieter, P., and Collins, F.S. (1997). Characterization of the CHD family of proteins. Proc. Natl. Acad. Sci. USA *94*, 11472–11477. https://doi.org/10.1073/pnas.94.21.11472.
77. Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433–5438. https://doi.org/10.1038/sj.onc.1210611.
78. Cai, Y., Geutjes, E.J., de Lint, K., Roepman, P., Bruurs, L., Yu, L.R., Wang, W., van Blijswijk, J., Mohammad, H., de Rink, I., et al. (2014). The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. Oncogene *33*, 2157–2168. https://doi.org/10.1038/onc.2013.178.
79. Yang, Y., Yamada, T., Hill, K.K., Hemberg, M., Reddy, N.C., Cho, H.Y., Guthrie, A.N., Oldenborg, A., Heiney, S.A., Ohmae, S., et al. (2016). Chromatin remodeling inactivates activity genes and regulates neural coding. Science *353*, 300–305. https://doi.org/10.1126/science.aad4225.
80. Bornelöv, S., Reynolds, N., Xenophontos, M., Gharbi, S., Johnstone, E., Floyd, R., Ralser, M., Signolet, J., Loos, R., Dietmann, S., et al. (2018). The nucleosome remodeling and deacetylation complex modulates chromatin structure at sites of active transcription to fine-tune gene expression. Mol. Cell *71*, 56–72.e4. https://doi.org/10.1016/j.molcel.2018.06.003.
81. Novillo, A., Fernández-Santander, A., Gaibar, M., Galán, M., Romero-Lorca, A., El Abdellaoui-Soussi, F., and Gómez-Del Arco, P. (2021). Role of chromodomain-helicase-DNA-Binding Protein 4 (CHD4) in Breast Cancer. Front. Oncol. *11*, 633233. https://doi.org/10.3389/fonc.2021.633233.
82. Wilczewski, C.M., Hepperla, A.J., Shimbo, T., Wasson, L., Robbe, Z.L., Davis, I.J., Wade, P.A., and Conlon, F.L. (2018). CHD4 and the NuRD complex directly control cardiac sarcomere formation. Proc. Natl. Acad. Sci. USA *115*, 6727–6732. https://doi.org/10.1073/pnas.1722219115.
83. Wong, H.K., Gibson, H., Hake, T., Geyer, S., Frederickson, J., Marcucci, G., Caligiuri, M.A., Porcu, P., and Mishra, A. (2015). Promoter-specific hypomethylation is associated with overexpression of PLS3, GATA6, and TWIST1 in the sezary syndrome. J. Invest. Dermatol. *135*, 2084–2092. https://doi.org/10.1038/jid.2015.116.
84. Tremblay, M., Sanchez-Ferras, O., and Bouchard, M. (2018). GATA transcription factors in development and disease. Development *145*, dev164384. https://doi.org/10.1242/dev.164384.
85. Mohd-Sarip, A., Teeuwssen, M., Bot, A.G., De Herdt, M.J., Willems, S.M., Baatenburg de Jong, R.J., Looijenga, L.H.J., Zatreanu, D., Bezstarosti, K., van Riet, J., et al. (2017). DOC1-dependent recruitment of NURD reveals antagonism with SWI/SNF during epithelial-mesenchymal transition in oral cancer cells. Cell Rep. *20*, 61–75. https://doi.org/10.1016/j.celrep.2017.06.020.
86. Mäkitie, R.E., Niinimäki, T., Suo-Palosaari, M., Kämpe, A., Costantini, A., Toiviainen-Salo, S., Niinimäki, J., and Mäkitie, O. (2020). PLS3 mutations cause severe age and sex-related spinal pathology. Front. Endocrinol. *11*, 393. https://doi.org/10.3389/fendo.2020.00393.
87. Xin, Z., Li, D., Mao, F., Du, Y., Wang, X., Xu, P., Li, Z., Qian, J., and Yao, J. (2020). PLS3 predicts poor prognosis in pancreatic cancer and promotes cancer cell proliferation via PI3K/AKT signaling. J. Cell. Physiol. *235*, 8416–8423. https://doi.org/10.1002/jcp.29685.