Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin

Kosuke Izumi1,2, Ryuichiro Nakato2, Zhe Zhang3, Andrew C Edmondson1, Sarah Noon1, Matthew C Dulik4, Ramakrishnan Rajagopalan4, Charles P Venditti5, Karen Gripp6, Joy Samanich7, Elaine H Zackai1,8, Matthew A Deardorff1,8, Dinah Clark1, Julian I Allen8,9, Dale Dorsett10, Ziva Misulovin10, Makiko Komata2, Masashige Bando3, Maninder Kaur1, Yuki Katou2, Katsuhiko Shirahige2,11 & Ian D Krantz1,8

Transcriptional elongation is critical for gene expression regulation during embryogenesis. The super elongation complex (SEC) governs this process by mobilizing paused RNA polymerase II (RNAP2). Using exome sequencing, we discovered missense mutations in AFF4, a core component of the SEC, in three unrelated probands with a new syndrome that phenotypically overlaps Cornelia de Lange syndrome (CdLS) that we have named CHOPS syndrome (C for cognitive impairment and coarse facies, H for heart defects, O for obesity, P for pulmonary involvement and S for short stature and skeletal dysplasia). Transcriptome and chromatin immunoprecipitation sequencing (ChIP-seq) analyses demonstrated similar alterations of genome-wide binding of AFF4, cohesin and RNAP2 in CdLS and CHOPS syndrome. Direct molecular interaction of the SEC, cohesin and RNAP2 was demonstrated. These data support a common molecular pathogenesis for CHOPS syndrome and CdLS caused by disturbance of transcriptional elongation due to alterations in genome-wide binding of AFF4 and cohesin.

Transcriptional regulation during embryogenesis is of paramount importance in determining the precise temporospatial gene expression pattern. Recent evidence suggests that such transcriptional control is mainly achieved by regulation of transcriptional elongation. Approximately 50 bp after transcriptional initiation, RNAP2 pauses in genes whose expression is mainly controlled at the elongation phase. For this transcriptional pausing, NELF (negative elongation factor) and DSIF (DRB sensitivity–inducing factor, a heterodimer of SPT4 and SPT5) are required. Serine residues located within the C-terminal heptapeptide repeats of RNAP2 undergo phosphorylation modifications during gene transcription. Ser5-phosphorylated RNAP2 (RNAP2 Ser5ph) is enriched at the promoter-proximal region, and Ser2 phosphorylation occurs after transcriptional elongation starts. Therefore, Ser2-phosphorylated RNAP2 (RNAP2 Ser2ph) is regarded as actively elongating RNAP2 (ref. 3). Mobilization of the paused RNAP2 machinery is governed by the SEC, which comprises multiple proteins including AFF4, ELL2 and positive transcription elongation factor b (P-TEFb), which is a heterodimer of CDK9 and cyclin T1 (ref. 4). CDK9 phosphorylates the Ser2 residue of RNAP2 to initiate transcriptional elongation.

As many multiple-congenital-anomaly genetic conditions are due to disruption of proper transcriptional processes during embryogenesis, disruption of the transcriptional elongation process would be a likely pathogenic mechanism for developmental disorders. However, no causal links have been identified thus far. Through the characterization of a new genetic disorder caused by gain-of-function mutations in the AFF4 gene, encoding a key component of the SEC, we have established a link between the dysregulated transcription observed among disorders of cohesin function (collectively termed ‘cohesinopathies’) and disruption of RNAP2 elongation, caused by altered genome-wide binding of AFF4 and cohesin.

RESULTS
A new genetic disorder caused by AFF4 missense mutations
Germline mutations of cohesin complex structural and regulatory components cause CdLS (MIM 122470, 300590, 610759, 614701)

1Division of Human Genetics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 2Research Center for Epigenetic Disease, Institute for Molecular and Cellular Biosciences, The University of Tokyo, Tokyo, Japan. 3Center for Biomedical Informatics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 4Department of Pathology and Laboratory Medicine, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 5National Human Genome Research Institute (NHGRI), US National Institutes of Health, Bethesda, Maryland, USA. 6Division of Medical Genetics, A.I. duPont Hospital for Children, Wilmington, Delaware, USA. 7Department of Pediatrics, Division of Genetics, Montefiore Medical Center, Bronx, New York, USA. 8The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. 9Division of Pulmonary Medicine, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 10Department of Biochemistry and Molecular Biology, St. Louis University School of Medicine, St. Louis, Missouri, USA. 11Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Kawaguchi, Japan. Correspondence should be addressed to I.D.K. (ian2@mail.med.upenn.edu) or K.S. (kshirahi@iam.u-tokyo.ac.jp).

Received 14 October 2014; accepted 30 January 2015; published online 2 March 2015; doi:10.1038/ng.3229
and CdLS is a multisystem developmental disorder characterized by craniofacial dysmorphism, intellectual disabilities, growth retardation, limb anomalies and several other systemic abnorma-

lities. Mutations in NIPBL have been identified in nearly 60% of CdLS probands, and HDAC8, SMC1A, SMC3 and RAD21 mutations account for an additional small portion of patients with CdLS6–9. The probands in the current study (CHOPS T254A and CHOPS R258W) were originally suspected of having CdLS, owing to the presence of intellectual disability, short stature and craniofacial dysmorphism (Fig. 1a). However, their physical features are clinically distinctive from those of typical CdLS probands, which allowed us to classify this as a new clinical entity (Supplementary Table 1). We propose ‘CHOPS syndrome’ as an acronym to describe this new genetic disorder: C for cognitive impairment and coarse facies, H for heart defects, O for obesity, P for pulmonary involvement and S for short stature and skeletal dysplasia.

The striking phenotypic similarities among the three unrelated probands, all born to unaffected, non-consanguineous parents, led us to hypothesize that their clinical features were likely the result of a detrimental germline de novo mutation in the same gene. To test this hypothesis, we performed exome sequencing on these three probands. Each proband had 414 (CHOPS T254A), 720 (CHOPS R258W) and 725 (CHOPS T254S) rare deleterious mutations (nonsynonymous, protein-truncating, deletion, duplication or splice-site mutations). Exome statistics are listed in Supplementary Table 2. Among these, we identified variants in 16 genes that were common to all probands. However, the variants in 14 genes were also present in in-house control samples, arguing against causality. We further scrutinized variants in the remaining two genes, AFF4 and ZFHX3. We confirmed the genetic variants of AFF4 and ZFHX3 by Sanger sequencing. All of the missense mutations found in AFF4 (c.761C>G (p.Thr254Ser) and c.772C>T (p.Arg258Trp)) were de novo (not present in the six biological parents of the three probands) (Fig. 1b and Supplementary Fig. 1). All three missense mutations mapped to the ALF (AF4-LAF4-FMR2) homology domain of AFF4, and these missense mutations altered highly evolutionarily conserved amino acids (Fig. 1b). All of the identified ZFHX3 variants were inherited from one of the parents of each proband (all of whom were unaffected). We confirmed paternity using STS markers in all

Figure 1 Identification of a new genetic disorder and its causative gene. (a) CHOPS syndrome probands: CHOPS T254S (female with short stature, intellectual disability, chronic lung disease, obesity, brachydactyly, vertebral abnormalities, patent ductus arteriosus, horseshoe kidney and dysmorphic facial features), CHOPS T254A (male with short stature, intellectual disability, tracheomalacia, subglottic and tracheal stenosis, obesity, brachydactyly, cervical vertebral abnormalities, ventricular septal defect and patent ductus arteriosus, cryptorchidism, hearing loss and dysmorphic facial features) and CHOPS R258W (female with short stature, intellectual disability, laryngomalacia, narrow oropharynx, brachydactyly, kyphoscoliosis, patent ductus arteriosus, ventricular septal defect, cataracts and dysmorphic facial features). Written permission to publish the photographs was obtained from the parents of the CdLS probands, which allowed us to hypothesize that their clinical features were likely the result of a detrimental germline de novo mutation in the same gene. To test this hypothesis, we performed exome sequencing on these three probands. Each proband had 414 (CHOPS T254A), 720 (CHOPS R258W) and 725 (CHOPS T254S) rare deleterious mutations (nonsynonymous, protein-truncating, deletion, duplication or splice-site mutations). Exome statistics are listed in Supplementary Table 2. Among these, we identified variants in 16 genes that were common to all probands. However, the variants in 14 genes were also present in in-house control samples, arguing against causality. We further scrutinized variants in the remaining two genes, AFF4 and ZFHX3. We confirmed the genetic variants of AFF4 and ZFHX3 by Sanger sequencing. All of the missense mutations found in AFF4 (c.761C>G (p.Thr254Ser) and c.772C>T (p.Arg258Trp)) were de novo (not present in the six biological parents of the three probands) (Fig. 1b and Supplementary Fig. 1). All three missense mutations mapped to the ALF (AF4-LAF4-FMR2) homology domain of AFF4, and these missense mutations altered highly evolutionarily conserved amino acids (Fig. 1b). All of the identified ZFHX3 variants were inherited from one of the parents of each proband (all of whom were unaffected). We confirmed paternity using STS markers in all
three probands. Screening of an additional 25 probands with atypical features of CdLS failed to identify mutations in the AFF4 gene. However, none of these probands exactly fit the CHOPS syndrome phenotype, indicating that mutations in this gene are highly correlated with the specific phenotype as characterized by the three patients described here. Collectively, these data support the hypothesis that missense mutations affecting the ALF homology domain of AFF4 cause CHOPS syndrome.

Mechanism of AFF4 mutations leading to CHOPS syndrome

A missense mutation affecting the ALF homology domain of Aff1 (also known as Af4) was reported in the robotic mouse, an ataxia mouse model created by N-ethyl-N-nitrosourea (ENU) mutagenesis. The pathogenetic mechanism of this missense mutation is a gain-of-function effect due to decreased clearance of the protein by SIAH1 ubiquitin E3 ligase. Given the vicinity of the location of the missense mutations in our probands to that described in the robotic mouse, we hypothesized that the missense mutations found in the three probands would likewise disrupt ubiquitination-dependent proteasomal degradation of the AFF4 protein. To test this hypothesis, we created an AFF4 and SIAH1 overexpression model using HEK293T cells. When an expression vector for wild-type AFF4 was transfected with an expression vector for SIAH1, the amount of AFF4 protein was markedly decreased; however, constructs for AFF4 containing the missense substitutions found in the three CHOPS syndrome probands showed hindered degradation of AFF4 with SIAH1 overexpression (Supplementary Fig. 2a). The addition of MG132, which is a proteasomal degradation inhibitor, resulted in the recovery of AFF4 bands.

Several genes including MYC and JUN have been identified as direct transcriptional targets of AFF4 (ref. 12). We confirmed upregulation of MYC and JUN expression in patient-derived skin fibroblast samples by quantitative RT-PCR, although one proband with an AFF4 p.Arg258Trp alteration had a normal level of JUN expression (Fig. 2b). We also evaluated the expression levels of MYC and JUN in HEK293T cells transfected with overexpression vectors for AFF4. With wild-type AFF4 overexpression, the expression levels of MYC and JUN increased, confirming that these genes are downstream targets of the AFF4 protein. Overexpression of the mutant AFF4 constructs resulted in stronger activation of MYC expression. Without coexpression of SIAH1, the difference between wild-type and mutant AFF4 overexpression was minimal; however, with the addition of SIAH1 overexpression, the expression difference between the wild-type and mutant AFF4 constructs increased both for MYC and JUN (Fig. 2c).

Because the SEC has an important role in the precise regulation of the immediate gene responses to various stimuli such as heat shock, retinoic acid or serum as well as in developmental regulation, we evaluated the effects of the AFF4 missense mutations on the immediate gene response using patient-derived skin fibroblast cell lines. These cell lines demonstrated upregulation of the FOS and EGR1 genes, which was comparable to that in control cell lines, upon serum stimulation (Supplementary Fig. 2a). Therefore, the proband-derived cell lines did not demonstrate an impaired response to serum stimulation with regard to the gene expression levels of FOS and EGR1, and mutant cells still maintained the capacity to mount a response to various stimuli.

Transcriptome analysis of CHOPS syndrome

Because of the critical role of the SEC in transcriptional regulation, we looked for abnormalities in global gene expression. We performed genome-wide expression profiling using Affymetrix U133 Plus 2.0 arrays on patient-derived skin fibroblast cell lines. We used two patient-derived samples (CHOPS T254A and CHOPS T254S) and three age- and sex-matched control samples, with all samples run in duplicate. We did not use the CHOPS R258W cell line owing to the lack of an appropriate ancestry-matched control cell line. Using a false discovery rate (FDR) cutoff of 0.2, we identified 288 genes downregulated and 445 genes upregulated in patient-derived samples relative to control samples (Supplementary Fig. 3 and Supplementary Table 3). Gene ontology term analysis using DAVID showed that the upregulated genes were enriched for homeobox proteins, skeletal system development/morphogenesis and anterior/posterior pattern formation, as well as embryonic organ development/morphogenesis, and that the downregulated genes were enriched for actin-binding proteins and extracellular matrix components (Supplementary Table 4). Previously, Luo et al. reported the results of AFF4 ChIP-seq and RNA sequencing (RNA-seq) after knockdown by small interfering RNA (siRNA) of AFF4 (ref. 12). Combining the data reported by Luo et al., we selected seven genes as probable direct targets of AFF4, owing to the presence of an AFF4 peak near their promoter regions and downregulation of their expression levels upon AFF4 knockdown. All seven target genes were upregulated by 9 to 127% (mean = 48.9%) in skin fibroblasts from patients with AFF4 mutations in comparison to control samples (Table 1 and Supplementary Fig. 4). The upregulation of five of these genes was statistically significant: MYC, JUN, TMEM100, ZNF711 and FAM13C. These observations further support the notion that AFF4 mutations render gain-of-function effects.

Transcriptome similarities between CHOPS syndrome and CdLS

Given the phenotypic similarities between the three CHOPS syndrome probands with AFF4 mutations and individuals with CdLS, we compared the gene expression pattern of skin fibroblast samples with AFF4 mutations to those from CdLS probands with NIPBL frameshift mutations using the Affymetrix Human Gene 2.0 ST array. We observed a general positive correlation (Pearson’s correlation coefficient = 0.34, P < 1 × 10$^{-300}$) for the gene expression patterns of CHOPS syndrome and CdLS (Fig. 3a). We identified the top 250 most dysregulated (overexpressed and underexpressed) genes for these diagnoses. Expression levels of up- and downregulated genes were similar for the CdLS probands and the probands with AFF4 mutations (Fig. 3b,c). Likewise, genes whose expression was downregulated in CdLS were similarly underexpressed in the probands with

| Gene       | Encoded product                                                                 | Mean, control | Mean, patient | Change in CHOPS (%) | P value |
|------------|---------------------------------------------------------------------------------|---------------|---------------|---------------------|---------|
| ADAMTS1    | ADAM metallopeptidase with thrombospondin type 1 motif, 1                       | 8.822         | 8.946         | +9.0                | 0.778   |
| FAM13C     | Family with sequence similarity 13, member C                                    | 2.720         | 3.063         | +26.8               | 0.013   |
| IMP2       | IMP2 inner-mitochondrial-membrane peptidase-like (S. cerevisiae)               | 6.850         | 7.008         | +11.6               | 0.095   |
| JUN        | Jun proto-oncogene                                                              | 8.794         | 9.257         | +37.9               | 0.030   |
| MYC        | v-Myc myelocytomatosis viral oncogene homolog (avian)                           | 7.667         | 8.571         | +87.2               | 0.030   |
| TMEM100    | Transmembrane protein 100                                                       | 3.179         | 4.364         | +27.4               | 0.001   |
| ZNF711     | Zinc-finger protein 711                                                         | 2.723         | 3.565         | +79.3               | 0.001   |

Table 1 Expression levels of AFF4 target genes

Group means are the average expression levels of genes measured by microarray, after normalization and log2 transformation. The processed measurements were compared between groups by the SAM method to generate the P values, and non-log-transformed measurements were used to calculate the percentage expression change between groups. Details are given in the Online Methods.
AFF4 mutations. Dysregulated genes in CHOPS syndrome demonstrated a similar expression pattern as observed in the samples from CdLS probands (Fig. 3b,c). There were 29 common genes within the top 250 most upregulated genes in CdLS and CHOPS syndrome (odds ratio (OR) = 13.05, P = 4.9 × 10⁻²¹, Fisher’s test). The top 250 most downregulated genes in CdLS and CHOPS syndrome included 27 genes in common (OR = 11.04, P = 4.4 × 10⁻¹⁸, Fisher’s test). For further analyses, we established hTERT-immortalized cell lines from fibroblasts derived from patients with CHOPS syndrome as well as CdLS and performed total RNA-seq analyses (Supplementary Table 5).

We observed positive correlation between transcriptional changes in CHOPS syndrome and CdLS (Supplementary Fig. 5), as we did for primary cultured cells. We scored 302 upregulated genes and 216 downregulated genes in CHOPS syndrome (Supplementary Fig. 5b,c and Supplementary Table 6). Of the 78 upregulated and 220 downregulated genes in CdLS, 30 and 72 genes, respectively, showed significantly similar expression tendencies to those observed in CHOPS syndrome. Although the CHOPS R258W sample did not show overexpression of JUN in quantitative RT-PCR (Fig. 2b), the profile of dysregulated genes was very similar for CHOPS R258W in comparison to CHOPS T254A and CHOPS T254S (Supplementary Fig. 5d).

Chromatin accumulation of AFF4 in CHOPS syndrome

With the confirmation that the missense mutations found in the three CHOPS syndrome probands caused decreased clearance of mutated AFF4 protein due to acquired resistance to SIAH1-mediated proteasomal degradation, we analyzed the amount of chromatin-associated SEC component in the CHOPS syndrome cell lines. The amount of AFF4 protein in patient-derived skin fibroblasts was elevated, supporting the suggestion that the missense mutations found in these probands result in a more stable AFF4 protein (Fig. 4a). The mRNA levels of AFF4 were comparable between proband and control samples (Supplementary Fig. 6a). In the samples from probands, the wild-type and mutant alleles were expressed in almost equal amounts (Supplementary Fig. 6b). We further analyzed the amount of chromatin-associated AFF4, ELL2 and CDK9 in CHOPS syndrome cell lines and found that accumulation of AFF4 mainly occurred on the chromatin fraction in the CHOPS syndrome cell lines (Fig. 4b). The increase in AFF4 accumulation in CHOPS syndrome proband CHOPS T254A was due to the mutant AFF4 allele, as the amount of

**Figure 3** Similar transcriptional profile for CHOPS syndrome and CdLS. (a) General comparison of expression changes in CHOPS syndrome and CdLS demonstrated positive correlation between the two syndromes. (b) Comparison of the top 250 most dysregulated genes in each syndrome. Left, upregulated and downregulated genes in CdLS showed similar changes in CHOPS syndrome samples. Right, upregulated and downregulated genes in CHOPS syndrome demonstrated similar changes in CdLS samples. The bottom of each box represents the first quartile, and the top of the each box represents the third quartile. The line in the middle of each box represents the median. The notches/whiskers demonstrate the confidence interval. (c) Volcano plots. Top left, volcano plot defining the top 250 genes whose expression levels were higher (purple dots) and lower (green dots) in the CHOPS syndrome samples. These plots represent the top 250 genes that were either up- or downregulated in the probands with the highest magnitude of difference and P < 0.05. Bottom left, distribution of these 250 genes, whose expression was up- or downregulated in CHOPS syndrome, in the CdLS samples. Top right, distribution of the top 250 genes whose expression was upregulated (purple dots) or downregulated (green dots) in the CdLS syndrome samples. Bottom right, distribution of these 250 genes, whose expression was up- or downregulated in CdLS, in the CHOPS syndrome samples. The genes from the plots above are labeled with the same colors in the bottom plots.
AFF4 protein was clearly decreased in cell lines where the mutated AFF4 allele was deleted (Fig. 4b and Supplementary Fig. 7). The amount of other SEC components such as ELL2 and CDK9 did not show major alterations in the chromatin fraction of the CHOPS syndrome cell lines relative to control cells. The CdLS cell line with an NIPBL frameshift mutation (CDL006) did not show elevation in the amounts of any of these proteins (Fig. 4b).

**DISCUSSION**

Through the identification of gain-of-function mutations in AFF4 as the cause of a new syndrome consisting of a highly conserved pattern of features that we have termed CHOPS syndrome, we demonstrate that an altered RNA2 distribution and resultant transcriptional elongation abnormalities underlie the transcriptional dysregulation of CHOPS syndrome as well as CdLS.

CHOPS syndrome represents the first human developmental disorder caused by germline mutations in an SEC component, although synonymous mutations were previously suggested to correlate with autism in two individuals.13,14 AFF4 is a scaffold protein comprising the core component of the SEC. In the probands reported here, the missense mutations create a resistance to ubiquitination-dependent proteasomal degradation, resulting in excessive amounts of accumulation of mutant AFF4 protein. This persistence of AFF4 protein in turn results in an altered genomic distribution of AFF4 and leads to the unique transcriptome pattern of CHOPS syndrome. The transcriptomic effects of the AFF4 gain-of-function mutations seen in CHOPS syndrome were opposite to those observed in AFF4 knockout experiments, confirming that the disease mechanism of these mutant AFF4 proteins is a gain of function rather than haploinsufficiency or a dominant-negative effect. In addition, the phenotype of the CHOPS syndrome probands is different from that of the AFF4 knockout mouse as well as human subjects with genomic deletions encompassing the AFF4 gene, lending further support to a gain-of-function effect.15,16 Comparing the AFF4 binding patterns among the upregulated and downregulated genes, an alteration was seen only in the upregulated genes, further supporting the notion that the primary effect of the AFF4 mutations found in CHOPS syndrome is gain of function, leading to transcriptional activation (Supplementary Fig. 11). The fundamental role of transcriptional elongation in the control of gene expression during embryogenesis has recently begun to be better understood. Therefore, CHOPS syndrome can be regarded as a multiple-congenital-anomaly syndrome caused by disturbance of the transcriptional elongation process.

Although accumulation of AFF4 was observed for the transcriptionally activated genes in CHOPS syndrome and CdLS, there was no correlation between AFF4 ChIP-seq and RNA-seq or RNA2 Ser2ph ChIP-seq in either syndrome (Supplementary Fig. 10). This finding suggests that AFF4 controls gene expression for only a subset of AFF4-bound genes, which is consistent with previously reported data by Luo et al., demonstrating that the knockdown of AFF4 leads...
to downregulation of expression for only a very small subset of genes (seven genes) among those that harbor an AFF4 ChIP-seq peak around the TSS in the HEK293T cell line22. In CHOPS syndrome and CdLS, transcriptional alterations are more likely seen for the genes where expression is mainly regulated by the transcriptional elongation step. Similar findings were reported for other transcription elongation controller molecules such as DSIF and NELF22,27,28.

In the CHOPS syndrome fibroblast cell lines, even with the massive accumulation of AFF4 in the chromatin fraction, the amount of other SEC components (ELL2 and CDK9) remained unchanged. However, the amount of these proteins forming the SEC is likely to be increased, given that the primary effect seen in CHOPS syndrome is activation of AFF4 target genes. CDK9 is known to form a complex with BRD4 or 7SK RNA and HEXIM outside of the SEC, as well as existing as a free CDK9 molecule18–21. Therefore, the amount of CDK9 forming the SEC is likely to be increased with the concurrent decrease in the amount of CDK9 forming complexes with BRD4 or 7SK RNA and HEXIM or free CDK9 in the nucleus. This speculation is supported by the fact that transfection with construct for AFF4 alone and HEXIM or free CDK9 in the nucleus. This speculation is supported by the fact that transfection with construct for AFF4 alone exerted transcriptional activation effects on SEC target genes such as MYC and JUN, even without concurrent overexpression of other SEC components (Fig. 2c).

AFF4 is expressed in fetal brain, lung and heart and in adult brain, heart, skeletal muscle and pancreas22. It is very intriguing that all CHOPS syndrome probands manifested with chronic lung disease, given the strong expression of AFF4 in fetal lung. In addition, AFF4 is implicated in regulation of appetite, as hypothalamic AFF4 expression is upregulated upon fasting23. Therefore, it is possible that gain-of-function mutations in AFF4 lead to increased appetite, resulting in the obesity seen in CHOPS syndrome probands.

Although CHOPS syndrome and CdLS are clinically recognizable as distinct entities, there is some phenotypic overlap between these two diagnoses suggesting a common underlying pathogenesis. Given the apparently divergent molecular etiologies of each, we were somewhat surprised to discover striking similarities in the transcriptomic profiles of CHOPS syndrome and CdLS.

Effects on gene transcription by mutated cohesin components or regulators have been proposed as the pathogenic mechanism of CdLS, establishing a critical role for cohesin in transcriptional regulation24. One mechanism by which cohesin controls transcription is through the creation of looping DNA structures that allow for physical interactions between distal regulatory elements and promoter regions25,26. However, the exact mechanism by which cohesin regulates transcription remains poorly understood.

We discovered that both CHOPS syndrome and CdLS display a similar disruption in the behavior of AFF4 and RAD21. Interestingly, the genes whose expression is significantly increased in CHOPS syndrome and CdLS demonstrate elevated AFF4 binding. In combination with the finding that similar sets of genes are upregulated and downregulated in CdLS and CHOPS syndrome, we can conclude that the altered genome-wide distribution of AFF4 binding affects cohesin binding in CHOPS syndrome, with the inverse relationship being true in CdLS (where the altered genome-wide distribution of cohesin affects AFF4 binding).

Furthermore, as a molecular link between CHOPS syndrome and CdLS, we demonstrated a physical interaction between the SEC, the STAG1-cohesin complex and RNAP2. This represents the first demonstration, to our knowledge, of such a physical interaction not only between the SEC and cohesin but also between cohesin and RNAP2. Recently, functional differences between the STAG1 and STAG2 components of cohesin have started to emerge. Previously, the Stag1 knockout mouse model demonstrated the importance of STAG1 in transcriptional regulation, as well as a critical function in chromosomal segregation27,28. Our results point to an important role for STAG1 in the pathogenesis of CdLS.

A role for cohesin in transcription elongation has previously been proposed. Cohesin selectively binds genes in which RNAP2 pauses just downstream of the TSS, and it was hypothesized that cohesin facilitates transition of paused RNAP2 to elongation24,29. This role of cohesin in transcriptional elongation is likely facilitated by NIPBL-MAU2, with AFF4 and the SEC mediating its effect. Recently, cohesin loaders (NIPBL-MAU2) were shown to be important in maintaining nucleosome-free regions in yeast models, and it was hypothesized that derangements of nucleosome positioning cause transcriptional disturbance in cohesinopathies30. Our observation suggests that NIPBL and MAU2 also have a role in the transcriptional elongation process, which requires the presence of paused RNAP2 at gene promoters.

In summary, we report the identification of a new genetic disorder, CHOPS syndrome, caused by mutations in AFF4 and demonstrate that altered genome-wide binding of AFF4 leading to RNAP2 gene body accumulation underlies the transcriptional abnormalities observed. We also demonstrate a molecular link between CHOPS syndrome and CdLS and implicate cohesin in transcriptional elongation. These observations underscore the importance of proper proximal pausing of RNAP2 as a crucial regulator of gene expression during human embryogenesis.

**ACKNOWLEDGMENTS**

We thank K. Nakagawa, S. Watanabe and E. Rappaport for their technical assistance. This work was supported by a Grant-in-Aid for Scientific Research (S) and for innovative science from MEXT (Ministry of Education, Culture, Sports, Science and Technology) (K.S.) and by research grants from the CdLS Foundation, development funds from the Children’s Hospital of Philadelphia and grant P01 HD052860 from the US National Institutes of Health (NIH)/National Institute of Child Health and Development (NICHD) (I.D.K.).

**A U T H O R C O N T R I B U T I O N S**

K.I. and I.D.K. initiated the human studies. K.I., S.N., K.G., J.S., E.H.Z., M.A.D., D.C., J.L.A. and I.D.K. characterized clinical data. K.I., M.C.D., R.R., C.P.V., M. Kaur and I.D.K. performed exome analysis and Sanger sequencing confirmation of the mutation. K.I., A.C.E., D.D., Z.M., M.B. and K.S. designed and performed the biochemical analyses. K.I. performed skin fibroblast expression studies and genome editing experiments. Z.Z. analyzed the skin fibroblast expression array data. R.N., M. Komata, M.B., Y.K. and K.S. performed ChIP-seq and RNA-seq and the accompanying bioinformatics analysis. K.I., K.S. and I.D.K. drafted the manuscript. All authors analyzed the data, discussed the results and commented on the manuscript.

**M E T H O D S**

Methods and any associated references are available in the online version of the paper.

**A c c e s s i o n c o d e s**

Expression array data have been deposited into the Gene Expression Omnibus (GEO) under accessions GSE64031 and GSE64034. ChIP-seq and RNA-seq data have been deposited into the Sequence Read Archive (SRA) under accession SRP050576.

**N O T E**

Any Supplementary Information and Source Data files are available in the online version of the paper.
COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Lin, C. et al. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). Genes Dev. 25, 1486–1498 (2011).
2. Yarnaguchi, Y., Shibata, H. & Handa, H. Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. Biochim. Biophys. Acta 1829, 98–104 (2013).
3. Heidemann, M., Hintermair, C., Voß, K. & Eick, D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim. Biophys. Acta 1829, 55–62 (2013).
4. Lin, C. et al. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. Mol. Cell 37, 429–437 (2010).
5. Liu, J. & Krantz, I.D. Cornelia de Lange syndrome, cohesin, and beyond. Clin. Genet. 76, 303–314 (2009).
6. Gillis, L.A. et al. NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. Am. J. Hum. Genet. 80, 485–494 (2007).
7. Deardorff, M.A. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. Am. J. Hum. Genet. 80, 485–494 (2007).
8. Deardorff, M.A. et al. RAD21 mutations cause a human cohesinopathy. Am. J. Hum. Genet. 90, 1014–1027 (2012).
9. Deardorff, M.A. et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature 489, 313–317 (2012).
10. Isaacs, A.M. et al. A mutation in AF4 is predicted to cause cerebellar ataxia and cataracts in the robotic mouse. J. Neurosci. 23, 1631–1637 (2003).
11. Oliver, P.L., Britou, E., Clark, J., Jones, E.L. & Davies, K.E. Mediation of AF4 protein function in the cerebellum by Siah proteins. Proc. Natl. Acad. Sci. USA 101, 14901–14906 (2004).
12. Luo, Z. et al. The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol. Cell. Biol. 32, 2608–2617 (2012).
13. Sanders, S.J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237–241 (2012).
14. O’Roak, B.J. et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat. Genet. 43, 585–589 (2011).
15. Urano, A. et al. Infertility with defective spermiogenesis in mice lacking AF5q31, the target of chromosomal translocation in human infant leukemia. Mol. Cell. Biol. 25, 6834–6845 (2005).
16. Tschach, A. et al. Molecular cytogenetic analysis of a de novo interstitial deletion of 5q23.3q31.2 and its phenotypic consequences. Am. J. Med. Genet. A. 140, 496–502 (2006).
17. Sun, J. et al. Genetic and genomic analyses of RNA polymerase II–pausing factor in regulation of mammalian transcription and cell growth. J. Biol. Chem. 286, 36248–36257 (2011).
18. Yang, Z., Zhu, Q., Luo, K. & Zhou, G. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature 414, 317–322 (2001).
19. Nguyen, V.T., Kiss, T., Michels, A.A. & Bensaude, O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature 414, 322–325 (2001).
20. Jang, M.K. et al. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II–dependent transcription. Mol. Cell 19, 523–534 (2005).
21. Yang, Z. et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Mol. Cell 19, 535–545 (2005).
22. Taki, T. et al. AF5q31, a newly identified AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia with ins(5;11)(q31;q13q23). Proc. Natl. Acad. Sci. USA 96, 14535–14540 (1999).
23. Komori, T. et al. Regulation of AMP-activated protein kinase signaling by AFF4 protein, member of AF4 (ALL1-fused gene from chromosome 4) family of transcription factors, in hypothalamic neurons. J. Biol. Chem. 287, 19985–19996 (2012).
24. Liu, J. et al. Transcriptional dysregulation in NIPBL and cohesin mutant human cells. PLoS Biol. 7, e1000119 (2009).
25. Wendt, K.S. et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796–801 (2008).
26. Kagey, M.H. et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435 (2010).
27. Remesiero, S. et al. Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. EMBO J. 31, 2076–2089 (2012).
28. Remesiero, S., Cuadrado, A., Gómez-López, G., Pisano, D.G. & Losada, A. A unique role of cohesin-SA1 in gene regulation and development. EMBO J. 31, 2090–2102 (2012).
29. Schaal, C.A. et al. Genome-wide control of RNA polymerase II activity by cohesin. PLoS Genet. 9, e1003382 (2013).
30. Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A. & Uhlmann, F. The Sccl2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. Nat. Genet. 46, 1147–1151 (2014).
ONLINE METHODS

Human subjects. All individuals enrolled in the study were evaluated by clinical geneticists experienced in the diagnosis of CdLS. All patients and family members were enrolled in the study under an institutional review board–approved protocol of informed consent at the Children’s Hospital of Philadelphia.

Exome sequencing and Sanger sequencing. Genomic DNA was extracted from peripheral blood cells. Exome capture was performed with Agilent SureSelect v 4 (Agilent Technologies). Captured DNA was sequenced using the Illumina HiSeq 2000 platform with paired-end reads of 100 bp. Sequencing reads obtained in fastq format were subjected to the whole-exome sequence analysis pipeline. Reads were mapped to human genome GRCh37 (1000 Genomes Project version) using Novoalign version 2.08. Base quality recalibration was also performed as part of the mapping in Novoalign. Duplicate reads were marked using Picard version 1.79. The rest of the workflow was based on the Genome Analysis Toolkit (GATK) package 31 best practices through MAU2 (ab46906), STAG1 (ref. 33), STAG2 (ref. 33), SMC1 (ref. 33), STAG1 (ab4457), STAG2 (ab4463), SMC1A (ab21583) and histone H3 (ref. 33) from Abcam. Antibody to RAD21 was custom and used in our previous study9. Antibodies to ELL2 used for immunoblotting (A302-505A) and AFF4 used for ChIP-seq analysis (A302-538A) were purchased from Bethyl Lab. Antibodies to NIPBL (sc-374625) and cyclin T1 (sc-10750) were from Santa Cruz Biotechnology. Antibody to SPT5 was provided by Y. Yamaguchi (Tokyo Institute of Technology). Antibodies to RNAP2 were used in a previous study (GM01864: an 8-year-old European-ancestry female; GM02036: an 11-year-old European-ancestry female; GM08398: an 8-year-old European-ancestry male) for U133P2 array analysis. The samples used for Human Gene 2.0 arrays included two CHOPs syndrome samples (CHOPS T254S (CDL160): a 6-year-old European-ancestry female with a p.Thr254Ser alteration; CHOPS T254A (CDL444)): a 12-year-old European-ancestry male with a p.Thr254Ala alteration) and three age- and sex-matched controls (GM01652: an 11-year-old European-ancestry female; GM02036: an 11-year-old European-ancestry female; GM08398: an 8-year-old European-ancestry male) for U133P2 array analysis. The samples used for Human Gene 2.0 arrays included two CHOPs syndrome samples (CHOPS T254S (CDL160) and CHOPS T254A (CDL444)), two CdLS samples (CDL006: a 7-year-old European-ancestry female with NIPBL mutation (NIPBL c.74_743delCTC; p.Leu24877fs); CDL015: a 10-year-old European-ancestry male with NIPBL mutation (NIPBL c.2969delG;p.Gly990Asfs*2)) and four age- and sex-matched control samples (GM01652, GM01864, GM02036 and GM03348) (GM03348: an 8-year-old European-ancestry female).

RNA was extracted using the RNeasy mini kit (Qiagen). cRNA was purified, fragmented and labeled using the 3’ IVT Express kit (Affymetrix). RNA (100 ng) was used from each sample. First-strand cDNA was synthesized from the extracted RNA, and second-strand cDNA was then synthesized from the first-strand cDNA. Subsequently, in vitro transcription and labeling of the amplified RNA (amplified RNA) was performed. Labeled cRNA was hybridized to Affymetrix GeneChip Human Genome U133 Plus 2.0 arrays.

Purified sense-strand cDNA was generated from extracted RNA using the Ambion WT Expression kit. RNA (100 ng) was used from each sample. Purified sense-strand cDNA samples were fragmented and labeled using the Affymetrix GeneChip WT Terminal Labeling and Controls kit. Labeled cDNA was hybridized to Affymetrix GeneChip Human Gene 2.0 ST arrays for transcriptome analysis. Arrays were washed and stained according to standard Affymetrix protocols using an Affymetrix GeneChip Fluidics Workstation 450 and scanned using an Affymetrix GeneChip Scanner 3000 7G. The DAT files were converted into cell intensities (CEL files) by Affymetrix Command Console software, and these were used for subsequent analyses. The sequences from which these probes were derived were selected from GenBank, dbEST and RefSeq.

Data were processed and analyzed within the R/Bioconductor environment. Original Affymetrix probes on both platforms were mapped to the current version of Entrez Gene using the library files provided by the BRAINARRAY database. Raw data were normalized and summarized with the RMA (robust multichip averaging) method to obtain gene expression levels. Pairwise comparison of sample groups was performed by the SAM (significance analysis of microarrays) method, which evaluated differential gene expression by fold change, P value and FDR. Fold changes and P values were used to create the volcano plots. The top 250 genes with the largest fold change and P < 0.05 were imported into the DAVID online tool for functional categorization. DAVID
reported predefined gene sets or pathways whose members were significantly over-represented in the 250-gene list.

**Quantitative RT-PCR.** Extracted RNA samples were also used for quantitative RT-PCR. RNA was reverse transcribed with TaqMan Reverse Transcription Reagents (Life Technologies), and synthesized cDNA was used for quantitative PCR analysis by droplet digital PCR system (QX100, Bio-Rad).55-56 Gene expression was analyzed using TaqMan gene expression assay probes, including AFF4 Hs00232683_m1, MYC Hs00153408_m1, JUN Hs01103582_s1, FOS Hs01491486_s1 and EGR1 Hs00152928_m1 (Life Technologies). Internal control probes used were TBP 4326322E and CYC 4326316E (Life Technologies). For allele-specific mRNA detection of AFF4 R258W, CDL006, GM01652, GM02036 and GM03348 (CHOPS R258W) mutation analysis was performed using TopHat version 2.0.8b 37 and Cufflinks version 2.1.1 (ref. 38) with the default parameter set. We applied Ensembl gene annotation (GRCh37) for both TopHat and Cufflinks. We considered insertions, deletions and gene fusions.

**RNA-seq analyses.** For total RNA-seq, we used immortalized skin fibroblast cell lines from CHOPS T254A (CDL444), CHOPS T254S (CDL160), CHOPS R258W, CDL006, GM01652, GM02036 and GM03348 (CHOPS R258W) lines of CHOPS T254A (CDL444), CHOPS T254S (CDL160), CHOPS R258W, CDL006, GM01652, GM02036 and GM03348 (CHOPS R258W) (CDL59): an 8-year-old African-American female with a p.Arg258Trp alteration). All samples were sequenced on the Illumina HiSeq 2500 platform as single-end 50-bp reads and mapped to the human genome (UCSC build hg19). Sequencing and mapping statistics are summarized in **Supplementary Table 5**. Total RNA was extracted using TRizol (Life Technologies) and Nucleopin RNA Reagents (Life Technologies) following the manufacturers’ instructions. The RNA-seq analysis was performed using TopHat version 2.0.8b 37 and Cufflinks version 2.1.1 (ref. 38) with the default parameter set. We applied Ensembl gene annotation (GRCh37) for both TopHat and Cufflinks. We considered protein-coding genes only, and gene-level expression values were estimated by fragments per kilobase of exon per million fragments mapped (FPKM) score. We obtained a total of 519 and 299 differentially expressed genes for CHOPS syndrome and Cdl5, respectively (**Supplementary Table 6**).

**ChIP-seq analyses.** For ChIP-seq, we used immortalized skin fibroblast cell lines of CHOPS T254A (CDL444), CDL006 and GM2036 samples. Chromatin immunoprecipitation was performed as previously described35. Briefly, cells were cross-linked with 1% formaldehyde for 10 min. Soluble cell lysates were prepared by sonication or both MNase treatment and sonication. Lysates were then incubated with protein A or protein G Dynabeads (Life Technologies) conjugated with antibodies. Beads were washed, and bound fractions were eluted. After reversal of cross-linking for elutes and input samples, DNA was purified using the PCR purification kit (Qiagen). DNA from input and immunoprecipitation fractions was further sheared to an average size of approximately 150 bp by ultrasonication (Covaris), end repaired, ligated to sequencing adaptors and amplified according to the manufacturer’s instructions (NEBNext ChiP-seq Library Prep Master Mix Set for Illumina; New England BioLabs). Sequenced reads were mapped to the human genome using Bowtie version 1.1.0 (ref. 40), allowing two mismatches in the first 28 bases per read and outputting only uniquely mapped reads (-n2 –m1 option). Mapping statistics are summarized in **Supplementary Table 7**. ChIP-seq analysis and visualization were performed using DROMPA version 2.5.3 (ref. 41). To facilitate comparison of detected peaks between different ChIP-seq experiments, the number of mapped reads was normalized with the total number of mapped reads. To eliminate uncertain sites, we ignored regions with low mappability (mappability <0.3 for a 1-kb window). Scatter plots for RNAP2, AFF4 and SPT5 are shown in **Supplementary Figures 8 and 10**. We applied CG command to normalize and count the mapped reads within each gene body region. Plotting of averaged reads around TSSs (**Supplementary Fig. 9**) was performed using PROFILE command. RefSeq gene annotation was obtained from the UCSC Genome Browser.

**Genome editing.** For the introduction of AFF4 deletion, the CRISPR/cas9 system was used46. gRNA and hCas9 vectors were purchased from Addgene. AFF4 exon 3 target sequence was introduced into an empty gRNA vector. The gRNA target sequence was ATGGCCGCACATAGGCAG. AFF4 exon 3–targeting gRNA vector and hCas9 plasmids were transfected into skin fibroblast cell lines using the Neon electroporation system (Life Technologies). After electroporation, single-cell cloning was performed. For each clone, the presence or absence of genome editing was confirmed by Sanger sequencing.

31. DePristo, M.A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498 (2011).
32. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92 (2012).
33. Komata, M. et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. Nat. Genet. 45, 1232–1237 (2013).
34. Stasewich, T.J. et al. Regulation of RNA polymerase II activation by histone acetylation in single living cells. Nature 516, 272–275 (2014).
35. Porensky, P.N. et al. A single administration of morpholino antisense oligomer rescues spinal muscular atrophy in mouse. Hum. Mol. Genet. 21, 1625–1638 (2012).
36. Heredia, N.J. et al. Droplet Digital™ PCR quantitation of HER2 expression in FFPE breast cancer samples. Methods 59, S20–S23 (2013).
37. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013).
38. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53 (2013).
39. Komata, M. et al. Chromatin immunoprecipitation protocol for mammalian cells. Methods Mol. Biol. 1164, 33–38 (2014).
40. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
41. Nakato, R., Itoh, T. & Shirahige, K. DROMPA: easy-to-handle peak calling and visualization software for the computational analysis and validation of ChiP-seq data. Genes Cells 18, 589–601 (2013).
42. Mail, P. et al. RNA-guide human genome engineering via Cas9. Science 339, 823–826 (2013).