nature genetics

NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome

Emma T Tonkin[1,3], Tzu-Jou Wang[1,3], Steven Lisgo[1], Michael J Bamshad[2] & Tom Strachan[1]

Cornelia de Lange syndrome (CdLS) is a multiple malformation disorder characterized by dysmorphic facial features, mental retardation, growth delay and limb reduction defects[1,2]. We indentified and characterized a new gene, *NIPBL*, that is mutated in individuals with CdLS and determined its structure and the structures of mouse, rat and zebrafish homologs. We named its protein product delangin. Vertebrate delangins have substantial homology to orthologs in flies, worms, plants and fungi, including Scc2-type sister chromatid cohesion proteins, and *D. melanogaster* Nipped-B. We propose that perturbed delangin function may inappropriately activate *DLX* genes, thereby contributing to the proximodistal limb patterning defects in CdLS. Genome analyses typically identify individual delangin or Nipped-B-like orthologs in diploid animal and plant genomes. The evolution of an ancestral sister chromatid cohesion protein to acquire an additional role in developmental gene regulation suggests that there are parallels between CdLS and Roberts syndrome.

The multisystem nature of the CdLS phenotype suggests that it is caused by a microdeletion or microduplication affecting several genes or by a single gene that regulates various target genes. A high-density BAC microarray comparative genome hybridization screen found no evidence for a consistent pattern of microdeletion or microduplication[3]. Because CdLS is rare and most cases are sporadic, genome-wide linkage screens are problematic. As an alternative, we analyzed chromosomal breakpoints associated with CdLS, focusing first on three classical cases with *de novo* balanced translocations, including the previously described translocations t(3;17)(q26.3;q23.1)[4] and t(14;21)(q32;q11)[5]. We first analyzed the 3q26.3 breakpoint because of

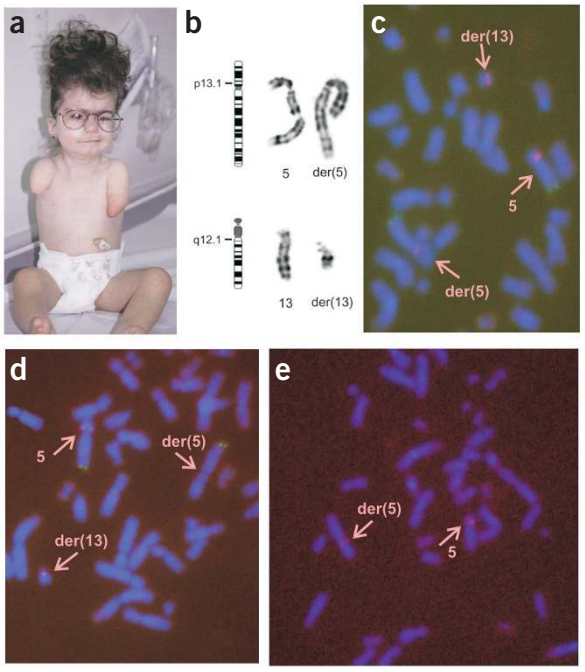

Figure 1 FISH mapping of a 5p13 translocation breakpoint in an individual with classical CdLS. (**a**) Individual with classical CdLS with characteristic limb and facial abnormalities (including an upturned triangular nose, long philtrum, thin upper lip, downturned corners of the mouth; see fuller description for individual P46 in **Table 1**). (**b**) Giemsa chromosome banding showing a balanced *de novo* t(5;13)(p13.1;q12.1) translocation. (**c–e**) Metaphase chromosome FISH with the breakpoint-spanning BAC clone CTD-2653m23 (**c**) and overlapping fosmid clones G248P84262B4 (**d**) and G248P8840C10 (**e**), all labeled with Spectrum Red. Labeled in green is a chromosome 5q telomere-specific probe. Arrows indicate the normal chromosome 5 and the der(5) t(5;13)(p13.1;q12.1) and der(13) t(5;13)(p13.1;q12.1) chromosomes. In occasional metaphases a weak G248P84262B4 signal can be detected on the der(5) chromosome as well as a strong signal on the der(13). The combined data suggest that the most likely location for the breakpoint is close to the proximal end of the region of overlap for inserts of G248P84262B4 and G248P8840C10 (**Fig. 2a**).

[1]Institute of Human Genetics, University of Newcastle, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK. [2]Department of Pediatrics and Department of Human Genetics University of Utah Salt Lake City, Utah 84112, USA. [3]These authors contributed equally to this work. Correspondence should be addressed to T.S. (tom.strachan@ncl.ac.uk).

Published online 16 May 2004; doi:10.1038/ng1363

Table 1 Phenotypes of selected individuals with CdLS with identified mutations

Individual	Category[a]	Nature of mutation	Location of mutation[b]	Inherited or *de novo*	Amino acid change	Phenotype
P2	Classical	1-bp insertion	7306_7307insG, exon 43	*De novo*	S2462X	Growth retardation, microbrachycephaly, flat facial profile, long philtrum, thin lips, crescent-shaped mouth, low-set ears, synophrys, bushy eyebrows, general hirsutism, hearing impairment, myopia, micromelia, clindactyly, proximally placed thumbs, fixed flexion of the elbows, syndactyly of the feet, recurrent pulmonary infection, hypoplastic umbilicus, bilateral inguinal hernias, undescended testes
P3	Classical	Missense	7289A→G, exon 43	*De novo*	Y2430C[c]	Severe growth retardation, lobster limb defect, characteristic face, feeding difficulty, gastroesophageal reflux
P11	Classical	Splice acceptor AG dinucleotide	5575–2A→G, intron 29	*De novo*	Undetermined	Growth retardation, characteristic facial features including long philtrum, anteverted nostrils, thin lips, crescent-shaped mouth, left postaxial polydactyly, bilateral hearing impairment, heart murmur
P13	Mild	3-bp deletion	3616–3618 delATA, exon 14	Change absent in maternal DNA; no paternal DNA available	I1206del[c]	Growth retardation, small hands, microcephaly, speech delay, inguinal hernia
P16	Classical	Missense	4043T→G, exon 17	*De novo*	L1348R[c]	Growth retardation, synophrys, long philtrum, thin upper lip, speech delay, high arched palate, congenital heart defect, short digits, feeding difficulty with gastroesophageal reflux
P18	Mild	Missense	3931T→C, exon 17	*De novo*	C1311R[c]	Small for gestational age, growth retardation, microcephaly, mild synophrys, neat arched eyebrows, depressed nasal bridge, smooth philtrum, thin lips, micromelia with short first and fifth digits of hands
P27	Classical	Nonsense	7903G→T, exon46	No parental DNA samples available	E2635X	Severe growth retardation, microcephaly, flat facial profile, synophrys, bushy eyelashes, short nose, anteverted nostrils, thin lips, downturned corners of mouth, micrognathia, small hands, hypoplastic and overlapping toes, bilateral syndactyly of second and third toes, delayed speech
P29	Classical	1-bp insertion	347_348insA, exon 4	No parental DNA samples available	Y116X	Severe growth retardation, microcephaly, micromelia of the hands, feeding difficulty, atrioseptal defect type II, cryptochidism
P37	Classical	Splice acceptor AG dinucleotide	7686–2A→G, intron 44	No parental DNA samples available	Undetermined	Growth retardation, characteristic facial features including long philtrum, anteverted nostrils, thin lips, crescent-shaped mouth, bilateral upper limb reduction defect
P46	Classical	t(5;13)	intron 1 (see **Fig. 2a**)	*De novo*	N/A	Severe growth retardation, characteristic facial abnormalities and bilateral transverse upper limb reduction defects (see **Fig. 1a**), ventricular septal defect, cleft palate, sacral myelomeningocele

[a]Classical and mild CdLS are primarily distinguished according to the severity of growth retardation and the extent of developmental delay[29,30]. [b]Nucleotide numbering refers to the cDNA sequence specifying the 2,804–amino acid isoform and commencing at the +1 position of the initiation codon. [c]Known amino acid conservation data are as follows: Ile1206 is conserved in mouse, rat and zebrafish, changed to methionine in fly and valine in *C. elegans*; Cys1311 is conserved in mouse, rat, zebrafish, fly and *C. elegans*; Leu1348 is conserved in mouse, rat and zebrafish, changed to isoleucine in fly and *C. elegans*; Tyr2430 is conserved in mouse, rat, zebrafish and fly, changed to phenylalanine in *C. elegans*.

perceived phenotypic overlap between duplication 3q syndrome and mild CdLS[6,7]. The 3q breakpoint disrupts a large gene undergoing unusual alternative splicing, but we found no additional mutations specific to any individuals with CdLS[3]. Molecular analyses of regions spanning the 17q23, 14q32 and 21q11 breakpoint regions also did not identify a gene likely to underlie CdLS (data not shown).

We localized the breakpoints in a third translocation case to 5p13.1 and 13q12.1 (**Fig. 1a**,**b** and **Table 1**). Fluorescent *in situ* hybridization (FISH) mapping identified BACs crossing the 5p breakpoint (**Fig. 1c**) and the 13q breakpoint. CdLS was recently reported to be associated with a 5p13.1-5p14.2 deletion (D. Viskochil, personal communication), and so we focused on the 5p breakpoint. We continued FISH mapping with fosmids until we identified two clones with overlapping inserts that mapped to either side of the 5p13 breakpoint. G248P84262B4 gave a clear hybridization signal on the normal chromosome 5 and the der(13), indicating preferential binding to the region telomeric to the 5p breakpoint (**Fig. 1d**). In contrast, G248P88840C10 hybridized clearly to the der(5) chromosome but was not visible on the der(13) chromosome (**Fig. 1e**).

The 5p breakpoint mapped close to a previously predicted gene-like sequence of unknown function, called *IDN3*. Using *in silico* analyses and in-house cDNA sequencing, we determined that *IDN3* was a gene fragment and that it comprises 90 kb of a new 190-kb gene, which we named *NIPBL* (*Nipped-B*-like; **Fig. 2a**,**b**). *NIPBL* contains 47 exons and is predicted to generate isoforms of 2,804 or 2,697 amino acids (**Fig. 2c**). Northern-blot analysis confirmed the predicted 9.8-kb transcript size and showed that *NIPBL* was strongly expressed in fetal and adult kidney, fetal liver, adult placenta, heart, skeletal muscle and thymus, but weakly or almost undetectably expressed in fetal and adult brain and lung and in adult liver, colon, small intestine and leukocytes (**Supplementary Fig. 1** online).

We screened other individuals with CdLS for mutations in *NIPBL* and identified nine plausible point mutations, at least five of which arose *de novo* (**Table 1** and **Supplementary Fig. 2** online). As we found *NIPBL* mutations in individuals with severe and mild CdLS, phenotype variation can be explained, at least in part, by allelic heterogeneity. The spectrum and distribution of mutations imply that pathogenesis arises from loss or altered function of a single *NIPBL* allele. Our mutation detection rate was ~50%. Locus heterogeneity could be a factor, but limitations of the screening methods is a plausible explanation for the comparatively low mutation detection rate. Considerable intrafamilial variation in phenotype, even between siblings with CdLS[8], has been reported, suggesting that additional factors may be important.

Using BLAST and BLAT analyses we determined the full-length sequence of the mouse, rat and zebrafish *NIPBL* homologs (data not shown). The exon structure is very well conserved in vertebrates and specifies a protein of ~2,800 amino acids. Sequence identities between human delangin and vertebrate orthologs are 96% (for mouse and rat) and 63% (for zebrafish; data not shown). TBLASTN searches against expressed-sequence tag databases showed that the two C-terminal isoforms are conserved in cow, pig, mouse, rat and chick (data not shown).

We also identified delangin homologs in flies (*Drosophila melanogaster* Nipped-B, *Anopheles gambiae* XM_320088), worms (*Caenorhabditis elegans* PQN-85, *C. briggsae* CBG0727), plants (*Arabidopsis thaliana* NM_121558, *Oryza sativa* NM_186173) and fungi (Scc2 family of sister chromatid cohesion proteins). In each case, homology is largely confined to a segment of ~1,500 amino acids spanning most of the delangin C-terminal half (**Fig. 2c**). Discounting small polyglutamine- and lysine-rich segments, most of the homologs do not share significant homology with any other protein sequence predicted from the relevant genome sequence. Because they are expected to be essential and most of their sequence shows homology to delangins, they may be viewed as orthologs. The pufferfish may be an exception: BLAT analyses suggest that there are two related *NIPBL*-like gene sequences.

Many of the fungal homologs have crucial chromosomal roles: *Saccharomyces cerevisiae* Scc2 and *Schizosaccharomyces pombe* Mis4 in sister chromatid cohesion[9,10] and *Coprinus cinereus* Rad9 in meiotic chromosome pairing and DNA repair[11]. Some metazoan orthologs, however, are known or likely to be developmental regulators. By facilitating activation of remote enhancers, the *D. melanogaster* Nipped-B protein regulates the Ultrabithorax (*Ubx*) and *Cut* homeobox genes[12,13]. RNA-interference knock-down of the gene encoding *C. elegans* PQN-85 results in a high level of embryonic lethality; survivors have a paralyzed uncoordinated phenotype, body morphology defects and sometimes a vulval defect (J. Ahringer, personal communication).

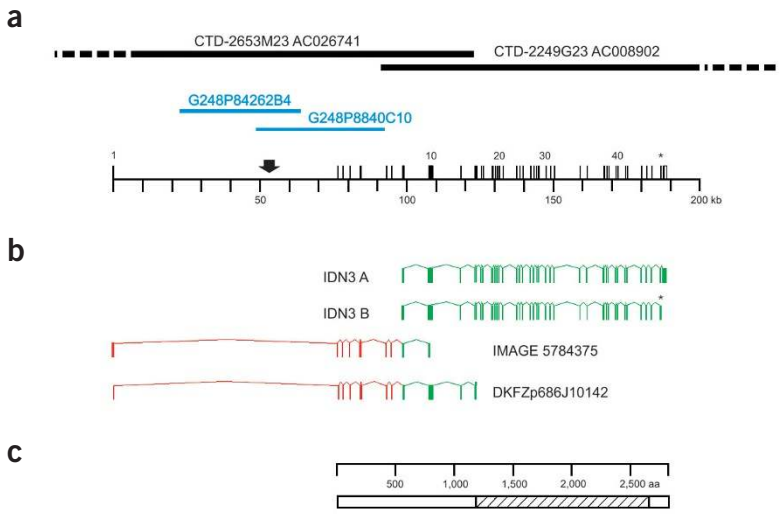

Figure 2 Normal exon-intron organization and expressed products of the gene *NIPBL* severed by the 5p13 breakpoint. (**a**) Genomic organization and FISH clones. A scale diagram of the exon-intron organization of *NIPBL* is shown at the bottom. Thick black horizontal bars at top show relevant BAC clones spanning and extending beyond the gene (solid and dotted black lines, respectively). Two fosmid clones that immediately flank the breakpoint (marked by the vertical arrow; see also **Fig. 1d**,**e**) are shown in blue. (**b**) Cognate cDNAs. Linked vertical bars illustrate exon selection in representative cDNAs used to determine the full exon complement, including the original IDN3 A and IDN3 B clones and overlapping clones that defined new 5′ exons (in red). The 5′ end of the IMAGE 5784375 clone sequence defines the start of exon 1. The asterisk denotes alternative 3′ end truncation distinguishing the short isoform. (**c**) Protein product. The coding sequence commences in exon 2 and continues either to exon 47, generating a long isoform, or to an expanded variant of exon 46, generating a slightly shorter isoform. The open bar represents the long (2,804 amino acids) isoform, with the internal striped box representing a region that is highly conserved during evolution and significantly related to fungal Scc2-type proteins. The short isoform contains 2,697 amino acids, of which residues 1–2,683 are identical to those of the long isoform, leaving a 14–amino acid C-terminal end that is unrelated to the 121–amino acid C-terminal end of the long isoform.

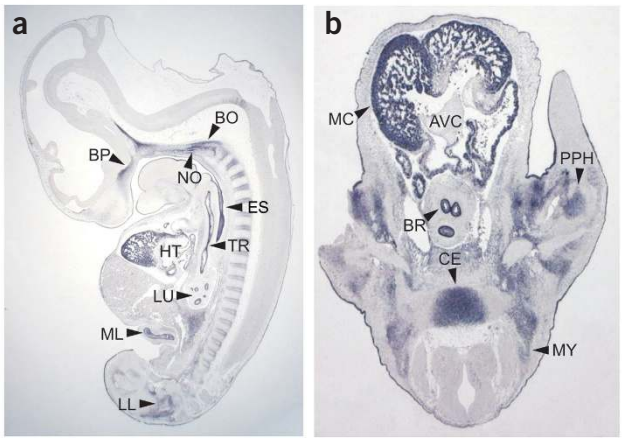

Figure 3 Specific embryonic expression of *NIPBL* transcripts shown by *in situ* hybridization. Bright-field microscopy images after hybridization to embryonic sections of a 424-bp antisense riboprobe corresponding to sequences in exons 10–12 of *NIPBL*. Images represent a midline section through a sagittally sectioned Carnegie stage 18 embryo (**a**) and a transverse section of a Carnegie stage 17 embryo (**b**). BP, cartilage primordium of the basisphenoid bone; BO, cartilage primordium of the basioccipital bone; NO, notochord; ES, esophagus; HT, heart; TR, trachea; LU, lung; ML, midgut loop; LL, lower limb; MC, heart myocardium; AVC, A-V cushion; BR, bronchus; PPH, precartilage primordium of the humerus; CE, centrum (vertebral body); MY, migrating myoblasts.

Because embryonic expression can differ substantially between some human-mouse orthologs[14], we carried out *NIPBL in situ* hybridization analyses on human embryonic tissue sections. The observed expression pattern was largely consistent with the CdLS phenotype (**Figs. 3** and **4**). *NIPBL* was expressed in developing limbs (**Fig. 3a**,**b**) and later in cartilage primordia of the ulna and of various hand bones (**Fig. 4c**). Sites of craniofacial expression included the cartilage primordium of the basioccipital and basisphenoid skull bones (**Figs. 3a** and **4f**) and elsewhere in the head and face, including a region encompassing the mesenchyme adjacent to the cochlear canal (**Fig. 4e**,**f**).

NIPBL was also expressed in the spinal column, notochord and surface ectoderm (**Figs. 3a**,**b** and **4b**), sclerotome and what seem to be migrating myoblasts (**Fig. 3b**). Expression in the developing heart was pronounced in the atrial and ventricular myocardium and in the ventricular tubeculae but absent in the endocardial cushions (**Fig. 3a**,**b**). *NIPBL* was also expressed in the developing esophagus, trachea and midgut loops (**Fig. 3a**), in the bronchi of the lung (**Figs. 3b** and **4d**) and in the tubules of the metanephros (**Fig. 4a**).

Expression in organs and tissues not typically affected in CdLS (*e.g.*, the developing trachea, bronchi, esophagus, heart and kidney) may reflect a bias towards underreporting of more subtle aspects of the phenotype or problems that typically present later in life. Respiratory and feeding difficulties and gastroesophageal reflux are known CdLS complications[1,15], individuals with CdLS have a greater incidence of congenital heart abnormalities[1,16] and renal abnormalities can be found in >50% of classical cases (A. Selicorni, personal communication). Expression in the mesenchyme surrounding the cochlear canal (**Fig. 4e**) may be related to the hearing impairment commonly found in CdLS[17]. Expression of *NIPBL* in embryonic brain was not evident, but the main neurodevelopmental deficits in CdLS are thought to occur during late gestation[18,19].

The involvement of Nipped-B in activating the *Ubx* and *Cut* homeobox genes[12] may provide insights into the molecular basis of CdLS pathogenesis. *Ubx* suppresses limb formation in the fly abdomen by repressing Distalless (*Dll*), a gene required for distal limb development[20]. The Dlx family of mammalian *Dll* homologs are involved in multiple developmental processes, including limb and branchial arch patterning, neurogenesis and hematopoiesis[21]. They are expressed in the apical ectodermal ridge of the developing limb bud, which partly coordinates limb outgrowth, and also in facial primordia. Therefore, the proximodistal limb patterning defect that underlies limb reduction in CdLS and possibly the associated facial abnormalities could largely be explained by inappropriate activation of *DLX* genes. Mutations in fly *Cut* cause leg and wing abnormalities. The mouse homolog *Cutl2* (*Cux2*) is dynamically expressed in branchial arch and limb bud progress zones[22], and so reduced expression of a human homolog in CdLS could also contribute to facial and limb abnormalities. The other mouse homolog, *Cutl1*, is widely expressed but important in lung development[23].

A dual role for Nipped-B in sister chromatid cohesion and developmental regulation was recently confirmed[24]. Similar dual roles

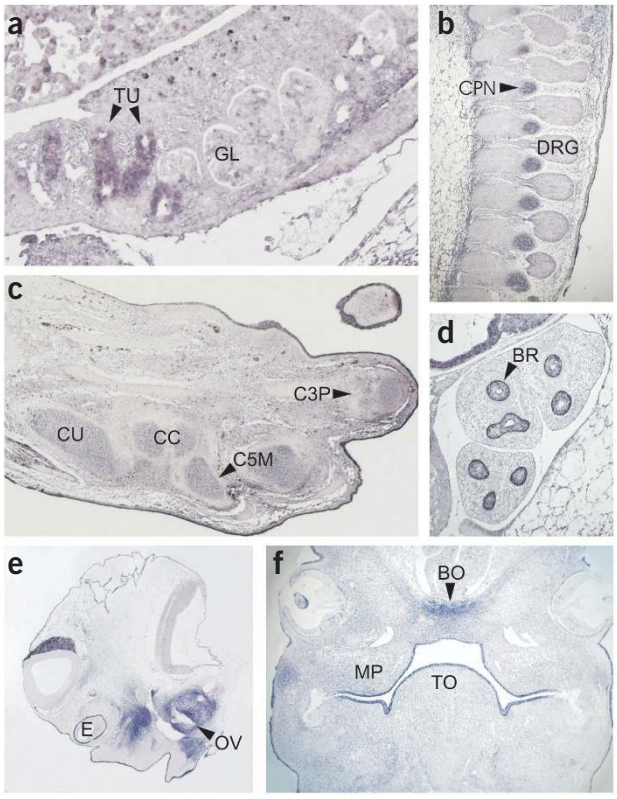

Figure 4 Detail of *NIPBL* expression in human embryonic development. Expression of *NIPBL* transcripts in specific structures of the metanephros (**a**), spinal column (**b**) and lung (**d**) of a single Carnegie stage 18 embryo, and in the hand plate of a Carnegie stage 21 embryo (**c**). Lower panels illustrate head and craniofacial structures in sagittal (**e**) and transverse (**f**) sections of a Carnegie stage 18 embryo. TU, metanephric tubule; GL, metanephric glomerulus; CPN, cartilage primordium of the neural arch; DRG, dorsal root ganglia; CU, cartilage primordium of the ulna; CC, cartilage primordium of the carpal bones; CM5, cartilage primordium of the metacarpal of the fifth digit; CP3, cartilage primordium of the phalangeal of the third digit; BR, bronchi of the lung; E, eye; OV, otic vesicle; BO, cartilage primordium of the basioccipital bone; MP, maxillary process; TO, tongue.

can be expected for vertebrate delangins, suggesting a possible parallel between CdLS and Roberts syndrome (OMIM 268300), which is characterized by growth retardation, limb reduction defects, craniofacial abnormalities and premature centromere separation. We assayed C-banded samples from individuals with CdLS for premature centromere separation but, perhaps unsurprisingly, detected no abnormalities; targeted knock-down of both alleles might be more informative. If delangin does have a dual functional role, the housekeeping role in facilitating *trans* interactions between sequences on sister chromatids could be satisfied with a basal level of expression. An additional role in enabling long-distance *cis* interactions (between promoter plus remote enhancer) for select target genes could require strong expression in tissues and organs where the target genes are active. The Scc2–Nipped-B–delangin family provides a model system for investigating evolutionary diversification of protein function.

METHODS

Chromosome analyses. We used thymidine to synchronize phytohemagglutinin-stimulated blood cultures and carried out G-banding according to standard protocols[25]. For premature centromere separation assays, we carried out standard C-banding[25] on fresh slides of samples from 12 individuals with CdLS. We scored 25 cells from each individual sample for premature centromere separation. We carried out chromosome FISH analysis by nick translation labeling of assorted genomic clones with SpectrumRed (Vysis) according to the manufacturer's instructions. Genomic clones included YAC, BAC and fosmid clones. We hybridized labeled probes along with a prelabeled chromosome 5p telomere-specific probe (Qbiogene) to metaphase chromosomes using standard methodology.

DNA sequencing and mutation screening. We obtained Image cDNA clones from the MRC Geneservice (see URL below). We sequenced all inserts using a combination of vector-specific and insert-specific primers and the MegaBACE ET system (Amersham). We screened mutations by direct sequencing and, when exons were suitably small, by SSCP-heteroduplex analysis using standard protocols. In the latter case, we denatured PCR products, size-fractionated them in 1× MDE gels (BioWhittaker) containing 5% glycerol and 0.6× TBE buffer at 300 V for ~20 h (depending on fragment size) and visualized them by silver staining. Any samples that had band differences relative to an unaffected control were sequenced using the MegaBACE ET system (Amersham). If the primary chromatogram suggested the presence of a deletion or insertion, we cloned the PCR product and sequenced a number of transformants to confirm the change. We reamplified and resequenced all mutations to confirm that the change observed was not the result of base misincorporation by the DNA polymerase. We used a panel of genomic DNA samples from 45 individuals of European descent (mostly from the UK, some from Poland and Ireland) with CdLS. For each mutation, we also screened 200 normal chromosomes. We screened parental DNA (when available) to confirm that the observed mutation had occurred *de novo*.

Because of the very long coding sequence (8,412 nucleotides), our mutation screening protocol surveyed a subset (26) of the 46 coding exons, namely exons 2–8, 13–20, 23, 24, 30, 34–36, 38, 40, 43, 45 and 46. The coding sequence sampled in these 26 exons corresponds to ~31% of the total. This means that approximately one-third of the coding sequence was sampled in 45 individuals and more than one-half of the proximal intronic sequence was also sampled for splicing mutations. On the basis of observed relative frequencies of splice site mutations and other mutations in large multiexon genes, and assuming that all affected individuals are heterozygotes (as expected from the strong evidence for autosomal dominant transmission[26]) and that there are no strong mutational hot spots in the coding sequence, the identification of 9 mutations in a panel of 45 individuals with CdLS (**Table 1**) equates roughly to a detection rate >50%. The relatively low mutation detection rate could reflect limitations of the mutation screening protocol: only coding sequences and proximal intronic sequences were analyzed, and the gene is large and possibly prone to undetected large-scale mutations.

Gene expression analyses. We designed PCR primers to amplify a 424-bp cDNA probe spanning exons 10–12 of *NIPBL*, which should hybridize to transcripts encoding both the long and short isoforms (primer sequences are available on our website; see URL below). For northern-blot analyses, we labeled the probe to high specific activity with [α^{32}P]-dCTP by random priming. After removing unincorporated nucleotides (NICK column, Amersham), we hybridized the probe against blots of human adult and fetal RNA (Clontech) containing ~2 μg of mRNA per lane at 42 °C overnight in ULTRAhyb (Ambion). We washed the blots in 0.1× SSC in 0.1SDS at 65 °C before exposing them to film. After removing the test probe, we rehybridized the blots with random-primed labeled control cDNA for human β-actin.

For tissue *in situ* hybridization, we cloned the 424-bp cDNA fragment into the pGEM-T Easy vector (Promega) and transcribed it with T7 and SP6 RNA polymerases incorporating DIG-11-UTP to generate labeled sense and antisense riboprobes, respectively. We generated additional isoform-specific probes to correspond to the long isoform of 2,804 amino acids and the short isoform of 2,697 amino acids (primer sequences are available on our website; see URL below). We hybridized the probes to sections of human embryonic tissue as described[27]. The isoform-specific probes generated similar expression patterns as the non-isoform-specific probe. We collected and used human embryonic tissue samples with ethical permission from the joint Ethics Committee of the Newcastle Health Authority and with appropriate signed consents. Samples were staged by microscopic examination. We fixed and processed tissue samples as previously described[28]. We selected the material we studied to have normal karyotypes and to be unrelated to disease.

URLs. Sequences of primers used for expression are available at our Newcastle CdLS research website at http://www.ncl.ac.uk/ihg/cdls. Servers used for nucleotide sequence analysis were the US National Center for Biotechnology Information's BLAST server (http://www.ncbi.nih.gov/BLAST/), the University of California at Santa Cruz's BLAT genome search server (http://genome.ucsc.edu/cgi-bin/hgBlat), the Ensembl genome browser (http://www.ensembl.org/), the University of California at Santa Cruz genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway), the NIX suite of nucleotide sequence analysis programs (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/) and Baylor College of Medicine's sequence utilities programs (http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html). Servers for protein sequence analysis included PSORTII (http://psort.nibb.ac.jp) and the DomPred program for predicting protein domains (http://bioinf.cs.ucl.ac.uk/dompred/). The Rep program (http://www.embl-heidelberg.de/~andrade/papers/rep/search.html) allowed us to identify five HEAT repeats in the conserved C-terminal domain. Alignment of multiple orthologous sequences was aided by using the ClustalW program at http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html. IMAGE cDNA clones were obtained from the MRC Geneservice, available at http://www.hgmp.mrc.ac.uk/geneservice/index.shtml.

GenBank accession numbers. Human *NIPBL* mRNA encoding the long delangin isoform, AJ627032; homologous mouse mRNA encoding the long delangin isoform, AJ627033; human *NIPBL* mRNA encoding the short delangin isoform, AJ640137; homologous mouse mRNA encoding the short delangin isoform, AJ640138; Image clone 5784375, AJ627564.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of F. Strachan (1921–2004). We thank many of our current Newcastle colleagues, especially S. Zwolinsky and J. Wolstenholme, for discussions, carrying out chromosome-banding analyses and conducting the premature centromere separation assay; H. Peters and D. Henderson for contributions to analysis of our expression data; S. Humphray for supplying fosmid clones; L. Jackson for facilitating the collaboration between M.B. and the Newcastle group; I. Krantz for discussions; A. Peaford and colleagues for their support; M. Walasek, M. Ireland and many other clinical geneticists for providing blood samples from individuals with CdLS and access to phenotype data; many individuals with CdLS and their families for their generosity; and previous members of the Newcastle CdLS research team, notably M. Smith, P. J. A. Eichhorn and B. Imamwerdi, for their earlier contributions. We thank the UK Community Fund and previously Action Research for providing funding for this project and the MRC-Wellcome Human Developmental Biology Resource for supplying human embryonic tissue samples.

COMPETING INTERESTS STATEMENT
The authors declare that they have no competing financial interests.

Received 19 February; accepted 28 April 2004
Published online at http://www.nature.com/naturegenetics/

1. Jackson, L., Kline, A.D., Barr, M.A. & Koch, S. de Lange syndrome: a clinical review of 310 individuals. *Am. J. Med. Genet.* **47**, 940–946 (1993).
2. Ireland, M., Donnai, D. & Burn, J. Brachmann-de Lange syndrome Delineation of the clinical phenotype. *Am. J. Med. Genet.* **47**, 959–964 (1993).
3. Tonkin, E.T. *et al.* A giant novel gene undergoing extensive alternative processing is severed by a Cornelia de Lange-associated translocation breakpoint at 3q26.3. *Hum. Genet.* (in the press).
4. Ireland, M., English, C., Cross, I., Houlsby, W.T. & Burn, J. A de novo translocation t(3;17)(q26.3;q23.1) in a child with Cornelia de Lange syndrome. *J. Med. Genet.* **28**, 639–640 (1991).
5. Wilson, W.G., Kennaugh, J.M., Kugler, J.P. & Wyandt, H.E. Reciprocal translocation 14q;21q in a patient with the Brachmann-de Lange syndrome. *J. Med. Genet.* **20**, 469–471 (1983).
6. Steinbach, P. *et al.* The dup(3q) syndrome: report of eight cases and review of the literature. *Am. J. Med. Genet.* **10**, 159–177 (1981).
7. Wilson, G.N., Dasouki, M. & Barr, M. Jr. Further delineation of the dup(3q) syndrome. *Am. J. Med. Genet.* **22**, 117–123 (1985).
8. Krajewska-Walasek, M., Chrzanowska, K., Tylki-Szymanska, A. & Bialecka, M. A further report of Brachmann-de Lange syndrome in two sibs with normal parents. *Clin. Genet.* **47**, 324–327 (1995).
9. Ciosk, R. *et al.* Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* **5**, 243–254 (2000).
10. Furuya, K., Takahashi, K. & Yanagida, M. Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G1 phase. *Genes Dev.* **12**, 3408–3418 (1998).
11. Seitz, L.C., Tang, K., Cummings, W.J. & Zolan, M.E. The rad9 gene of Coprinus cinereus encodes a proline-rich protein required for meiotic chromosome condensation and synapsis. *Genetics* **142**, 1105–1117 (1996).
12. Rollins, R.A., Morcillo, P. & Dorsett, D. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152**, 577–593 (1999).
13. Dorsett, D. Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr. Opin. Genet. Dev.* **9**, 505–514 (1999).
14. Fougerousse, F. *et al.* Human-mouse differences in the embryonic expression patterns of developmental control genes and disease genes. *Hum. Mol. Genet.* **9**, 165–173 (2000).
15. Luzzani, S., Macchini, F., Valade, A., Milani, D. & Selicorni, A. Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical syndromes. *Am. J. Med. Genet.* **119A**, 283–287 (2003).
16. Mehta, A.V. & Ambalavanan, S.K. Occurrence of congenital heart disease in children with Brachmann-de Lange syndrome. *Am. J. Med. Genet.* **71**, 434–435 (1997).
17. Sataloff, R.T., Spiegel, J.R., Hawkshaw, M., Epstein, J.M. & Jackson, L. Cornelia de Lange syndrome. Otolaryngologic manifestations. *Arch. Otolaryngol. Head Neck Surg.* **116**, 1044–1046 (1990).
18. Yamaguchi, K. & Ishitobi, F. Brain dysgenesis in Cornelia de Lange syndrome. *Clin. Neuropathol.* **18**, 99–105 (1999).
19. Vuilleumier, N. *et al.* Neuropathological analysis of an adult case of the Cornelia de Lange syndrome. *Acta Neuropathol. (Berl.)* **104**, 327–332 (2002).
20. Gebelein, B., Culi, J., Ryoo, H.D., Zhang, W. & Mann, R.S. Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev. Cell* **3**, 487–498 (2002).
21. Panganiban, G. & Rubinstein, J.L.R. Developmental functions of the *Distal-less*/Dlx homeobox genes. *Development* **129**, 4371–4386 (2002).
22. Iulianella, A., Vanden Heuvel, G. & Trainor, P. Dynamic expression of murine Cux2 in craniofacial, limb, urogenital and neuronal primordia. *Gene Expr. Patterns* **3**, 571–577 (2003).
23. Ellis, T. *et al.* The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001).
24. Rollins, R.A., Korom, M., Aulner, N., Martens, A. & Dorsett, D. Drosophila Nipped-B protein supports sister chromatid cohesion and opposes the Stromalin/Scc3 Cohesion factor to facilitate long-range activation of the cut gene. *Mol. Cell. Biol.* **24**, 3100–3111 (2004).
25. Rooney D.E. (ed.) *Human Cytogenetics: Constitutional Analysis. A Practical Approach.* 3rd. edn. (Oxford University Press, Oxford, 2001).
26. Russell, K.L. *et al.* Dominant paternal transmission of Cornelia de Lange syndrome: a new case and review of 25 previously reported familial recurrences. *Am. J. Med. Genet.* **104**, 267–276 (2001).
27. Breitschopf, H., Suchanek, G., Gould, R.M., Colman, D.R. & Lassmann, H. *In situ* hybridization with digoxigenin-labeled probes: sensitive and reliable detection method applied to myelinating rat brain. *Acta Neuropathol.*, **84**, 581–587 (1992).
28. Lako, M. *et al.* A novel mammalian Wnt gene, *WNT8B*, shows brain-restricted expression in early development, with sharply delimited expression boundaries in the developing forebrain. *Hum. Mol. Genet.* **7**, 813–822 (1998).
29. Van Allen, M.I. *et al.* Clinical variability within Brachmann-de Lange syndrome: a proposed classification system. *Am. J. Med. Genet.* **47**, 947–958 (1993).
30. Allanson, J.E., Hennekam, R.C. & Ireland, M. De Lange syndrome: subjective and objective comparison of the classical and mild phenotypes. *J. Med. Genet.* **34**, 645–650 (1997).