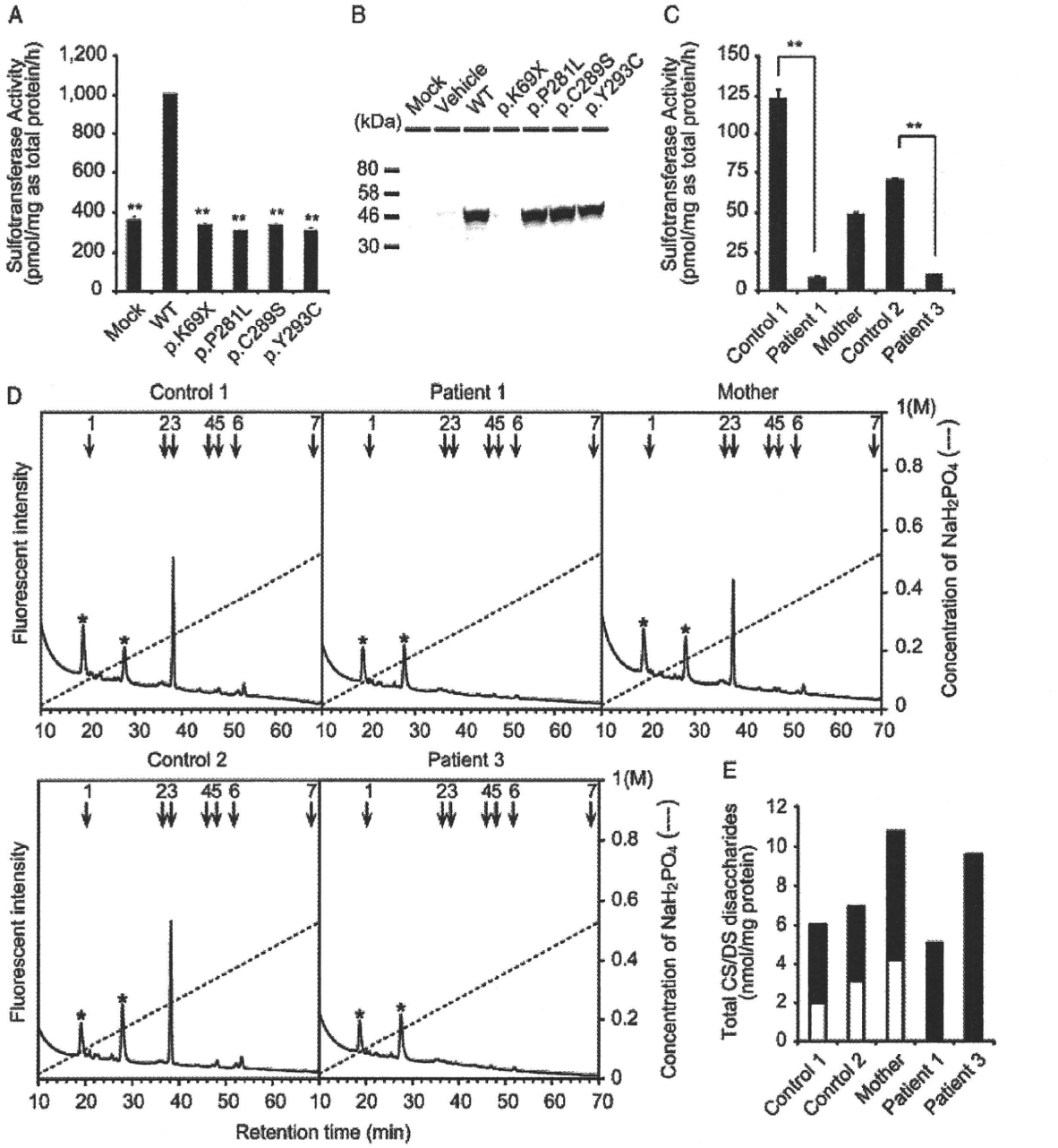

Figure 2. Glycobiological studies. **A:** Sulfotransferase activity of the recombinant D4ST1 of a control and four mutant forms: p.K69X, p.P281L, p.C289S, and p.Y293C. The activity is shown as the incorporation of [^{35}S]SO_4 from the donor [^{35}S]PAPS into the acceptor dermatan. Values are the mean ± SEM ($n = 3$) $^{**}P < 0.0001$ versus WT by one-way ANOVA with Dunnett's adjustment. **B:** Immunoblotting for the exogenous D4ST1 expression in COS-7 cell using the anti-V5 antibody. The mutant p.K69X was too small to be detected (predicted to be approximately 11 kDa). **C:** Sulfotransferase activity toward dermatan in the skin fibroblasts derived from patients 1 and 3, mother of patient 1, and sex- and age-matched healthy controls (controls 1 and 2 for patients 1 and 2, respectively) (mean ± SEM, $n = 3$). $^{**}P < 0.0001$ by two-tailed unpaired t test. **D:** Anion-exchange chromatograms of DS disaccharides obtained by CSase B digestion of GAG-peptides from the skin fibroblasts. The elution positions of authentic 2AB-labeled disaccharide standards are indicated by numbered arrows. 1, ΔHexUA-GalNAc; 2, ΔHexUA-GalNAc(6S); 3, ΔHexUA-GalNAc(4S); 4 ΔHexUA(2S)-GalNAc(6S); 5, ΔHexUA(2S)-GalNAc(4S); 6, ΔHexUA-GalNAc(4S,6S), and 7, ΔHexUA(2S)-GalNAc(4S,6S). Asterisks indicate impurities. **E:** The total amounts of CS and DS derived from skin fibroblasts. The total disaccharide contents of CS (black box) and DS (white box) were calculated based on the peak areas in the chromatograms of the digests with CSase AC and CSase B, respectively.

lysates of primary fibroblasts. The mean activity of the fibroblast lysates of the patients was significantly decreased to 6.7% (patient 1) and 14.5% (patient 3) of each age- and sex-matched control 1 and 2, respectively (Fig. 2C). These data indicate mutations in this study result in loss of function.

Glycosaminoglycan (GAG) Chain Analysis

The disaccharide compositions of DS and chondroitin sulfate (CS) in fibroblasts derived from patients and controls were analyzed by anion-exchange HPLC after digestion with

chondroitinase (CSase) B and a mixture of CSases AC-I and AC-II (CSase AC), respectively. CSase B cleaves the *N*-acetyl-D-galactosaminidic linkage in the GalNAc(4S)-L-iduronic acid (IdoUA) (±2S) sequences and yields unsaturated disaccharides, ΔHexUA (4,5-unsaturated hexuronic acid)-GalNAc(4S) and ΔHexUA(2S)-GalNAc (4S) [Sugahara and Mikami, 2007; Yoshida et al., 1993], where 2S and 4S represent 2-*O*-sulfate and 4-*O*-sulfate, respectively. The 4-*O*-sulfation of GalNAc residues is essential for recognition by CSase B. In contrast, CSase AC does not act on IdoUA-containing sequences but degrades the *N*-acetyl-D-galactosaminidic linkages in the GalNAc (±4S, ±6S)-D-glucuronic acid (GlcUA) sequences [Linhardt et al., 2006], where 6S stands for 6-*O*-sulfate. When digested with CSase B, 4-*O*-sulfated unsaturated disaccharides were detected only in the controls but not in the patients (Fig. 2D and Supp. Table S1), suggesting that D4ST1 is the major enzyme for 4-*O*-sulfation of DS in normal human skin fibroblasts and cannot be compensated functionally either by chondroitin 4-*O*-sulfotransferase 1 or 2.

To investigate whether the loss-of-function mutations in D4ST1 lead to an increase in nonsulfated dermatan, the GAG fraction prepared from the fibroblasts of the patients and controls were digested with CSase ABC, which cleaves all the *N*-acetyl-D-galactosaminidic linkages in CS/DS chains including the non-sulfated GalNAc-IdoUA sequence that is resistant to CSase B [Yoshida et al., 1993]. Notably, the proportion of the non-sulfated disaccharide ΔHexUA-GalNAc in the patients was very low (Supp. Table S1), suggesting that nonsulfated dermatan was negligible. Conversion of GlcUA to IdoUA most likely occurs even in the patients. However, the epimerization reaction is reversible that favors the GlcUA formation, and D4ST1 probably functions as the "4-*O*-sulfation lock" of IdoA [Malmström, 1984]. Hence, the defect in D4ST1 probably allowed back epimerization reactions converting IdoUA to GlcUA to form CS in the patients. The total amount of CS/DS disaccharides did not show a significant difference between the patients and controls (Fig. 2E). CS disaccharides produced by digestion with CSase AC increased markedly in the cell lysates of fibroblasts from the patients compared with those from the controls, suggesting the upregulation of CS chain biosynthesis (diverted from DS) in the patients.

Glycosaminoglycan of Decorin

We then examined decorin as a major DS proteoglycan in skin. Decorin directly binds to collagen via its core protein and the GAG side chains aggregate and function as interfibrillar bridges [Scott, 1996, 2003]. Decorin purified from the fibroblasts of the patients was resistant to CSase B, indicating that its GAG side chain consisted of only CS but no DS disaccharides, while decorin in the controls contained a hybrid CS/DS chain (Fig. 3A). Actually, the GAG side chain of decorin from the controls was mainly composed of DS disaccharides (approximately 95%), IdoUA-GalNAc(4S), and IdoUA(2S)-GalNAc(4S), but contained a small portion of CS disaccharide, GlcUA-GalNAc(4S) (Fig. 3B). In contrast, no DS disaccharides were detected in the decorin proteoglycan from the patients (Fig. 3B). Its GAG side chain most likely consisted of only CS disaccharides including 4-*O*-sulfated and 6-*O*-sulfated disaccharide units. These data and those shown in Figure 2E suggest that the DS chain of decorin proteoglycan in the patients has been replaced by CS.

TGF-β Signaling Is Unaltered in Fibroblasts Derived from Patients

DS proteoglycan is known to be related to TGF-β signaling [Hocking et al., 1998; Tiedemann et al., 2005; Yamaguchi et al., 1990]. As decorin neutralizes TGF-β1 activity [Yamaguchi et al., 1990], we examined whether *CHST14* aberrations would affect TGF-β1 signal transduction. We overexpressed wild-type or mutant *CHST14* cDNA in the fibroblasts from patients in which endogenous *CHST14* expression was negligible and found that the levels of *CHST14* expression from transgenes were approximately 300-fold those of endogenous mutant *CHST14* expression at 56 hr after transfection (Supp. Fig. S2A). To investigate whether *CHST14* overexpression would affect TGF-β signaling in the patient fibroblasts, we performed three different experiments. First, we measured the expression of *PAI1* and *SMAD7*, which are direct downstream targets of TGF-β. The expression of *PAI1* and *SMAD7* was significantly upregulated at 24 hr after TGF-β1 treatment. However, the degree of upregulation was not different between the cells transfected with empty or *CHST14* expression

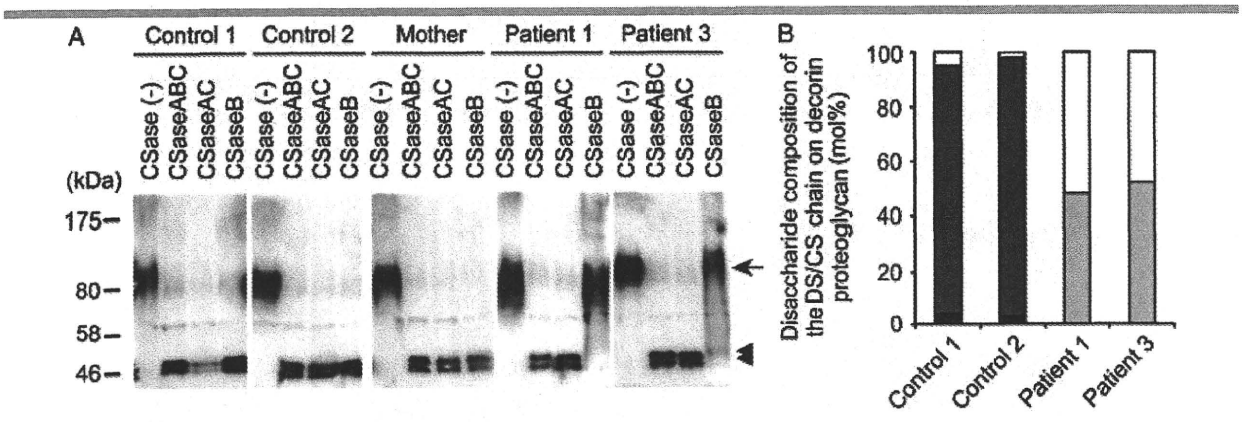

Figure 3. Disaccharide composition analysis of CS/DS chains of decorin. **A:** Immunoblot of decorin proteoglycans. Each serum-free conditioned medium of skin fibroblasts was digested with CSase ABC, CSase AC, CSase B, or buffer alone (—) and then subjected to Western blotting using an antihuman decorin antibody. The arrow and arrowheads indicate the decorin proteoglycans and the core protein of decorin without a CS/DS side chain, respectively. **B:** Proportion of the disaccharide units in the CS/DS hybrid chain in decorin proteoglycans secreted by the fibroblasts. White, light gray, dark gray, and black boxes are GlcUA-GalNAc(4S), GlcUA-GalNAc(6S), IdoUA-GalNAc(4S), and IdoUA(2S)-GalNAc(4S), respectively.

vectors (Supp. Fig. S2B). Second, TGF-β1 transmits signals to the nucleus via phosphorylation of SMAD2/3 proteins. Therefore, a reporter gene assay was performed using SBE4-luc vector, a TGF-β-responsive reporter containing four tandem copies of the SMAD-binding element (SBE). *CHST14* overexpression did not affect the reporter activity upregulated by TGF-β1 treatment (Supp. Fig. S2C). Third, we examined the level of phosphorylated SMAD2 proteins by Western blot analysis. Although TGF-β1 treatment clearly stimulated the phosphorylation of SMAD2, there were no significant differences between the cells transfected with empty or *CHST14* vectors (Supp. Fig. S2D). The same results from these three experiments were obtained using fibroblasts from another patient (data not shown). Furthermore, we confirmed that overexpression of the mutant D4ST1 enzymes identified in patients did not affect TGF-β signaling in normal fibroblasts (data not shown). These results imply that TGF-β signaling might not be changed in the fibroblasts from patients with a D4ST1 deficit.

Histopathological Examination

Hematoxylin and eosin (H&E)-stained light microscopy on patients' skin specimens showed that fine collagen fibers were predominant in the reticular to papillary dermis and normally thick collagen bundles were markedly reduced (Fig. 4). Electron microscopy showed that the collagen fibrils were dispersed in the reticular dermis, compared with regularly and tightly assembled ones observed in the control, whereas each collagen fibril was smooth and round, not varying in size and shape, similar to each fibril of the control (Fig. 4).

Discussion

In this study, we identified a total four *CHST14* mutations (three missense and one nonsense) in six Japanese patients presenting with a new type of EDS [Kosho et al., 2010]. This disorder represents all hallmarks of EDS: skin hyperextensibility, joint hypermobility, and tissue fragility affecting skin, ligaments, joints, blood vessels, and internal organs [Steinmann et al., 2002]. Tentatively we categorized patients 1 and 2 before as EDS VIB, a subtype of kyphoscoliosis type without lysyl hydroxylase deficiency, based on cutaneous (hyperextensibility, bruisability, fragility with atrophic scars) and skeletal (generalized joint laxity, kyphoscoliosis, Marfanoid habitus) features as well as mild delay of motor development with hypotonia in infancy [Kosho et al.,

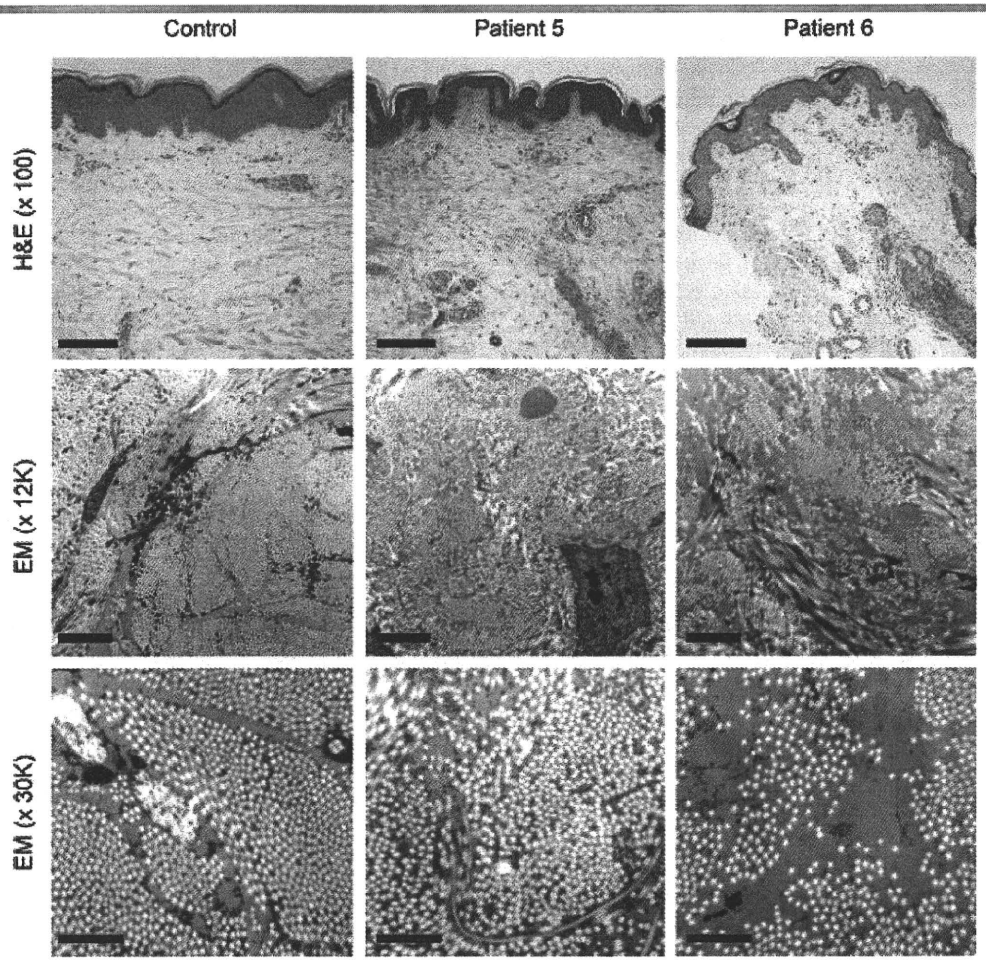

Figure 4. Pathological examination of dermal tissues from patients. H&E-stained light microscopy (upper picture) on skin specimens of patients 5 and 6 (× 100) shows that fine collagen fibers were present predominantly in the reticular to papillary dermis with marked reduction of normally thick collagen bundles. Electron microscopy (EM) (× 12,000 middle picture; × 30,000 lower picture) showed that collagen fibrils were dispersed in the reticular dermis, compared with the regularly and tightly assembled ones observed in the control subject. However, each collagen fibril was smooth and round, not varying in size, similar to the control. Scale bars indicate 500 μm (upper picture), 2 μm (middle), and 1 μm (lower).

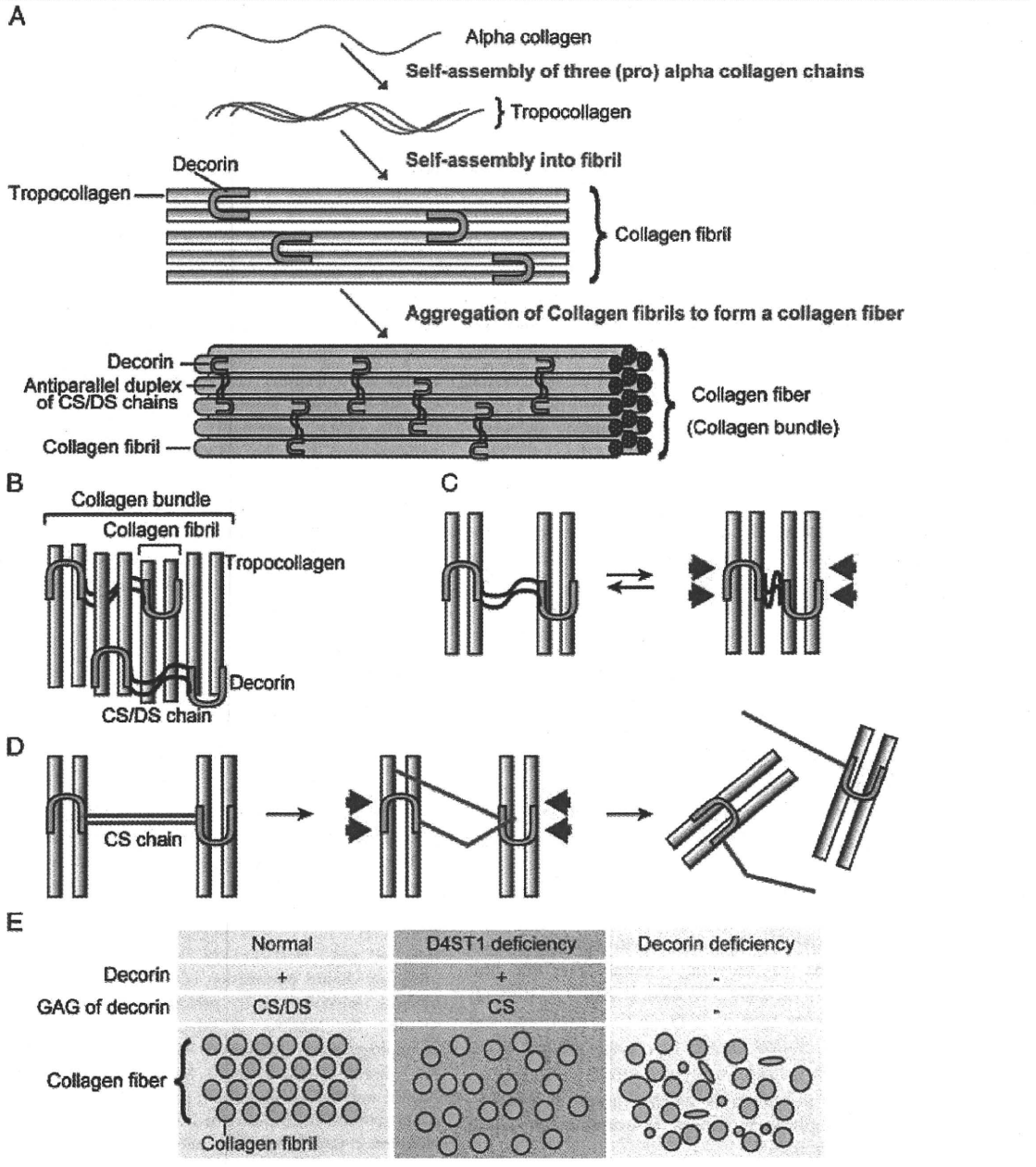

Figure 5. Putative model for abnormal collagen bundle assembly in this disease. **A:** The scheme of the collagen biosynthesis. Three alpha collagen chains are self-assembled into form a tropocollagen unit. Tropocollagens are packed via decorin to form a collagen fibril. Decorin core protein directly binds to collagen's particular amino sequence. Collagen fibrils assemble to form the collagen fiber, also called a collagen bundle when thick. **B:** The relationship between collagen fibril and decorin proteoglycan. Collagen fibrils are assembled into a collagen bundle by the antiparallel complex of the CS/DS hybrid GAG chains of decorin proteoglycan like a bridge to keep the space between fibrils and make collagen fiber tighter. **C, D:** The structural alteration of collagen fibers by mechanical compression in normal and affected states, respectively. CS/DS hybrid chains are able to bend against mechanical compression and rebound to form the original structure, because GAG chains function as suspension (C). D4ST1 defects result in absence of CS/DS hybrid chains in decorin (D). Replaced CS chains might not form collagen bundles properly. Even if they can form bundles, decorin CS chains cannot resist mechanical stresses and collagen fibrils get scattered after repetitive stresses irreversibly, as observed in patients (D). Red arrows indicate mechanical pressures. **E:** Comparisons of collagen fiber formation in normal (left), D4ST1 deficiency (middle), and decorin (core protein) deficiency (right). Collagen fibrils normally aggregate in line and form collagen fibers. The fibrils are round and uniform. In D4ST1 deficiency, collagen fibrils are scattered, although the shape and size of collagen fibrils are unchanged. Irregular shapes and sizes of collagen fibrils were seen in decorin null mice [Danielson et al., 1997], being apparently different from those in D4ST1 deficiency.

2005]. Patients in our series [Kosho et al., 2010] shared many clinical features with a Pakistani sister and brother reported by Steinmann et al. [1975], including down-slanting palpebral fissures, high palate, talipes equinovarus, progressive talipes vagus and planus, joint laxity, scoliosis, skin hyperextensibility, bruisability, and fragility with atrophic scars; hyperalgesia to pressure,

radiologically identified tall vertebral bodies and diaphysial narrowing of phalanges, metacarpals, and metatarsals; and delayed motor development. The sibs have been classified into EDS VIB [Steinmann et al., 2002], with the lysyl hydroxylase activity proved to be normal [Wenstrup et al., 1989]. Pathological findings were similar to those observed in our series: light microscopically, collagen bundles were interspersed with filamentous material that stained only faintly; and, electron microscopically, a great proportion of collagen fibrils was not integrated into bundles but dispersed into the ground substance [Steinmann et al., 1975].

CHST14 has recently been demonstrated as the causative gene for adducted thumb–clubfoot syndrome (ATCS; MIM# 601776), another autosomal recessive disorder [Dündar et al., 2009]. ATCS has been categorized as a new type of arthrogryposis, based on characteristic clinical pictures from birth to early childhood, including adducted thumbs and clubfoot as well as craniofacial dysmorphism (broad or bossed forehead, brachycephaly, late-closing large fontanelle, hypertelorism, downslanting palpebral fissures, blue sclerae, low-set posteriorly rotated or dysplastic ears, high or cleft palate, short neck), arachnodactyly with tapering fingers, cryptorchidism, inguinal hernia, atrial septal defect, kidney defects, cranial ventricular enlargement, and psychomotor retardation [Dündar et al., 1997, 2001; Janecke et al., 2001; Sonoda and Kouno, 2000]. In a recent study by Dündar et al. [2009], ATCS has been again categorized as a connective tissue disorder, based on additional clinical pictures from childhood to adolescence, including skin fragility, bruisability, and translucency; joint laxity, and osteopenia. In 11 ATCS patients from four families identified to date [Dündar et al., 2009], 5 died in early infancy or childhood: a male from a Austrian family (p.R213P) died shortly after birth due to respiratory failure [Janecke et al., 2001]; a female from a Turkish family (p.V49X) died at age 6 years [Dündar et al., 1997], and a female and two males from a Turkish family (p.[R135G]+[L137Q]) died before 4 months of age [Dündar et al., 2001]. ATCS patients may have wider and more severe manifestations than our series, implicating roles for DS not only in connective tissue maintenance but also in embryonic development [Dündar et al., 2009]. Furthermore, a skin specimen from a patient was interpreted as showing normal structure and ultrastructure [Dündar et al., 2009]. To date, it would be difficult to delineate whether a new type of EDS we have proposed and ATCS would be distinct clinical entities or a single clinical entity (D4ST1 deficiency) with variable inter- and intrafamilial expressions and with different presentations depending on the patients' ages at diagnosis. Longitudinal clinical information of ATCS patients would solve this issue.

Collagen bundles are composed of many collagen fibrils linked via antiparallel CS/DS chain complexes (Fig. 5A and B). The property of reversible deformation of proteoglycan comes from antiparallel anionic GAG chain formation [Scott, 2003]. Interestingly, the structure of DS chains is flexible, whereas that of CS chains is rigid, because L-IdoUA residues in DS can switch readily between almost equi-energetic $^1C_4$, $^2S_0$, and $^4C_1$ conformers; by comparison GlcUA in CS adopts purely the $^4C_1$ conformation [Casu et al., 1988; Catlow et al., 2008]. Transition from the CS/DS hybrid chain of decorin to a CS chain probably decreases the flexibility of the GAG chain and break the GAG antiparallel complex after compression stresses (Fig. 5C and D). This irreversible event could explain the progressive course of this disease. In the patients, the size and shape of the collagen fibrils seemed normal, whereas the collagen bundles were not properly organized. In decorin null mice, dermal collagen fibrils showed huge varieties of size and shape [Danielson et al., 1997]. These findings suggest that the core protein of decorin is important for collagen fibril formation, and that the CS/DS hybrid chain of decorin proteoglycan regulates the space between the collagen fibrils and forms collagen bundles as reported previously (Fig. 5E) [Scott, 1995]. Decorin is able to neutralize the growth-stimulatory activity of TGF-β1 in Chinese hamster ovary cells through the interaction of TGF-β1 with decorin via its core protein, not via a GAG chain [Yamaguchi et al., 1990]. In our transfection experiments with *CHST14* cDNAs, no significant effects on TGF-β signaling between wild-type and mutants were detected, implying the importance of the decorin core protein (not GAG) for TGF-β signaling (Supp. Fig. S2).

In conclusion, we have detected *CHST14* mutations causing a new type of EDS with distinct craniofacial characteristics, multiple congenital contractures, progressive joint and skin laxity, and multisystem fragility related manifestations. Abnormal collagen bundle formation would be a main pathology associated with a decorin GAG abnormality. Decorin GAG side chains in the patients consist of only CS but no DS disaccharides. CS/DS hybrid chains are more flexible than CS chains. Collagen bundles bound by CS chains instead of CS/DS chains in patients should be more fragile than those in controls. These findings underscore the important aspects of decorin proteoglycans in the extracellular matrix and provide new insights for human connective tissue disorders.

Acknowledgments

We thank all the patients and their families for participating in this work. We also thank Ms. Y. Yamashita, Ms. T. Taniguchi, and Dr. S. Tominaga for their technical assistance, Drs. Y. Igawa, T. Miyahara (Shinshu University School of Medicine), Y. Seta, Y. Toki, K. Ono, T. Kosuda, A. Inoue (Tomioka General Hospital, Tomioka), T. Miyamoto (Meido Eye Clinic), K. Arai (Komachi Clinic), T. Shimizu (Gunma Spine Center), F. Miura, H. Ikei (Saku Central Hospital), N. Hoshino (Minami-Aiki-Mura Clinic), T. Muneta (Tokyo Medical and Dental University), N. Kurosawa (Tsuchiura Kyodo General Hospital), M. Nagasaka, M. Kato (Aichi Prefectural Colony Central Hospital), M. Kohyama (JA Hiroshima General Hospital), and T. Hattori (Aichi Children's Health and Medical Center) for providing clinical information of the patients. Dr. A. Kosaku (Dokkyo Medical University) kindly performed the electron microscopy. Grant sponsor: the Ministry of Education, Culture, Sports, Science and Technology; grant numbers: 21790341 (to N. Miyake), B-20390019 (to K.S.) and A-21249024 (to N. Matsumoto). Grant sponsor: the Japan Science and Technology Agency; grant number: 2014 (to N. Matsumoto). Grant sponsor: the Ministry of Health, Labour and Welfare, Japan; grant numbers: 2141039040 (to T.K. and N. Matsumoto), 20-S-3 (to S.I.). Grant sponsors: the Shinshu Association for the Advancement of Medical Sciences (to T.K.); a Grant-in-Aid for Exploratory Research of Young Scientists, Shinshu University (to T.K.); the Takeda Science Foundation (to N. Miyake); the Yokohama Foundation for Advancement of Medical Science; grant number: W2105 (to N. Miyake). Grant sponsor: the Hayashi Memorial Foundation for Female Natural Scientists; grant number: 09R198 (to N. Miyake). Grant sponsor: the Human Frontier Science Program; grant number; RGP0018/2005-C (to K.S.). This work has been done at Advanced Medical Research Center, Yokohama City University.

References

Beighton P, De Paepe A, Steinmann B, Tsipouras P, Wenstrup RJ. 1998. Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). Am J Med Genet 77:31–37.

Casu B, Petitou M, Provasoli M, Sinay P. 1988. Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans. Trends Biochem Sci 13:221–225.

Catlow KR, Deakin JA, Wei Z, Delehedde M, Fernig DG, Gherardi E, Gallagher JT, Pavao MS, Lyon M. 2008. Interactions of hepatocyte growth factor/scatter factor with various glycosaminoglycans reveal an important interplay between the presence of iduronate and sulfate density. J Biol Chem 283:5235–5248.

Danielson KG, Baribault H, Holmes DF, Graham H, Kadler KE, Iozzo RV. 1997. Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. J Cell Biol 136:729–743.

Dündar M, Demiryilmaz F, Demiryilmaz I, Kumandas S, Erkilic K, Kendirci M, Tuncel M, Ozyazgan I, Tolmie JL. 1997. An autosomal recessive adducted thumb-club foot syndrome observed in Turkish cousins. Clin Genet 51:61–64.

Dündar M, Kurtoglu S, Elmas B, Demiryilmaz F, Candemir Z, Ozkul Y, Durak AC. 2001. A case with adducted thumb and club foot syndrome. Clin Dysmorphol 10:291–293.

Dündar M, Muller T, Zhang Q, Pan J, Steinmann B, Vodopiutz J, Gruber R, Sonoda T, Krabichler B, Utermann G, Baenziger JU. 2009. Loss of dermatan-4-sulfotransferase 1 function results in adducted thumb-clubfoot syndrome. Am J Hum Genet 85:873–882.

Hiyama K, Okada S. 1975. Amino acid composition and physiochemical characterization of chondroitinase from *Arthrobacter aurescens*. J Biochem 78:1183–1190.

Hocking AM, Shinomura T, McQuillan DJ. 1998. Leucine-rich repeat glycoproteins of the extracellular matrix. Matrix Biol 17:1–19.

Janecke AR, Unsinn K, Kreczy A, Baldissera I, Gassner I, Neu N, Utermann G, Muller T. 2001. Adducted thumb-club foot syndrome in sibs of a consanguineous Austrian family. J Med Genet 38:265–269.

Kinoshita A, Sugahara K. 1999. Microanalysis of glycosaminoglycan-derived oligosaccharides labeled with a fluorophore 2-aminobenzamide by high-performance liquid chromatography: application to disaccharide composition analysis and exosequencing of oligosaccharides. Anal Biochem 269:367–378.

Kosho T, Miyake N, Hatamochi A, Takahashi J, Kato H, Miyahara T, Igawa Y, Yasui H, Ishida T, Ono K, Kosuda T, Inoue A, Kohyama M, Hattori T, Ohashi H, Nishimura G, Kawamura R, Wakui K, Fukushima Y, Matsumoto N. 2010. A new Ehlers-Danlos Syndrome with craniofacial characteristics, multiple congenital contractures, progressive joint and skin laxity, and multisystem fragility-related manifestations. Am J Med Genet A [Epub ahead of print].

Kosho T, Takahashi J, Ohashi H, Nishimura G, Kato H, Fukushima Y. 2005. Ehlers-Danlos syndrome type VIB with characteristic facies, decreased curvatures of the spinal column, and joint contractures in two unrelated girls. Am J Med Genet A 138A:282–287.

Linhardt RJ, Avci FY, Toida T, Kim YS, Cygler M. 2006. CS lyases: structure, activity, and applications in analysis and the treatment of diseases. Adv Pharmacol 53:187–215.

Malmström A. 1984. Biosynthesis of dermatan sulfate. II. Substrate specificity of the C-5 uronosyl epimerase. J Biol Chem 259:161–165.

Mao JR, Bristow J. 2001. The Ehlers-Danlos syndrome: on beyond collagens. J Clin Invest 107:1063–1069.

Michelacci YM, Dietrich CP. 1974. Isolation and partial characterization of an induced chondroitinase B from *Flavobacterium heparinum*. Biochem Biophys Res Commun 56:973–980.

Mikami T, Mizumoto S, Kago N, Kitagawa H, Sugahara K. 2003. Specificities of three distinct human chondroitin/dermatan N-acetylgalactosamine 4-*O*-sulfotransferases demonstrated using partially desulfated dermatan sulfate as an acceptor: implication of differential roles in dermatan sulfate biosynthesis. J Biol Chem 278:36115–36127.

Penc SF, Pomahac B, Winkler T, Dorschner RA, Eriksson E, Herndon M, Gallo RL. 1998. Dermatan sulfate released after injury is a potent promoter of fibroblast growth factor-2 function. J Biol Chem 273:28116–28121.

Scott JE. 1995. Extracellular matrix, supramolecular organisation and shape. J Anat 187(Pt 2):259–269.

Scott JE. 1996. Proteodermatan and proteokeratan sulfate (decorin, lumican/fibromodulin) proteins are horseshoe shaped. Implications for their interactions with collagen. Biochemistry 35:8795–8799.

Scott JE. 2003. Elasticity in extracellular matrix "shape modules" of tendon, cartilage, etc. A sliding proteoglycan-filament model. J Physiol 553(Pt 2):335–343.

Sonoda T, Kouno K. 2000. Two brothers with distal arthrogryposis, peculiar facial appearance, cleft palate, short stature, hydronephrosis, retentio testis, and normal intelligence: a new type of distal arthrogryposis? Am J Med Genet 91:280–285.

Steinmann B, Gitzelmann R, Vogel A, Grant ME, Harwood R, Sear CH. 1975. Ehlers-Danlos syndrome in two siblings with deficient lysyl hydroxylase activity in cultured skin fibroblasts but only mild hydroxylysine deficit in skin. Helv Paediatr Acta 30:255–274.

Steinmann B, Royce PM, Superti-Furga A. 2002. The Ehlers-Danlos syndrome. In: Royce PM, Stainmann B, editors. Connective tissue and heritable disorders. Hoboken, NJ: John Wiley & Sons Inc. p 431–523.

Sugahara K, Mikami T. 2007. Chondroitin/dermatan sulfate in the central nervous system. Curr Opin Struct Biol 17:536–545.

Tiedemann K, Olander B, Eklund E, Todorova L, Bengtsson M, Maccarana M, Westergren-Thorsson G, Malmstrom A. 2005. Regulation of the chondroitin/dermatan fine structure by transforming growth factor-beta1 through effects on polymer-modifying enzymes. Glycobiology 15:1277–1285.

Uyama T, Ishida M, Izumikawa T, Trybala E, Tufaro F, Bergstrom T, Sugahara K, Kitagawa H. 2006. Chondroitin 4-*O*-sulfotransferase-1 regulates E disaccharide expression of chondroitin sulfate required for herpes simplex virus infectivity. J Biol Chem 281:38668–38674.

Wenstrup RJ, Murad S, Pinnell SR. 1989. Ehlers-Danlos syndrome type VI: clinical manifestations of collagen lysyl hydroxylase deficiency. J Pediatr 115:405–409.

Yamagata T, Saito H, Habuchi O, Suzuki S. 1968. Purification and properties of bacterial chondroitinases and chondrosulfatases. J Biol Chem 243:1235–1523.

Yamaguchi Y, Mann DM, Ruoslahti E. 1990. Negative regulation of transforming growth factor-beta by the proteoglycan decorin. Nature 346:281–284.

Yasui H, Adachi Y, Minami T, Ishida T, Kato Y, Imai K. 2003. Combination therapy of DDAVP and conjugated estrogens for a recurrent large subcutaneous hematoma in Ehlers-Danlos syndrome. Am J Hematol 72:71–72.

Yoshida K, Arai M, Kohno Y, Maeyema KI, Myazono H, Kikuchi H, Morikawa K, Tawada A, Suzuki S. 1993. Activity of bacterial eliminases towards delmatan sulphates and dermatan sulphate proteoglycan. In: Scott JE, editor. Dermatan sulfate proteoglycans: chemistry, biology, chemical pathology. London: Portland Press. p 55–80.

Zawel L, Dai JL, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B, Kern SE. 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell 1:611–617.

nature genetics

Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome

Sarah B Ng[1,7], Abigail W Bigham[2,7], Kati J Buckingham[2], Mark C Hannibal[2,3], Margaret J McMillin[2], Heidi I Gildersleeve[2], Anita E Beck[2,3], Holly K Tabor[2,3], Gregory M Cooper[1], Heather C Mefford[2], Choli Lee[1], Emily H Turner[1], Joshua D Smith[1], Mark J Rieder[1], Koh-ichiro Yoshiura[4], Naomichi Matsumoto[5], Tohru Ohta[6], Norio Niikawa[6], Deborah A Nickerson[1], Michael J Bamshad[1–3] & Jay Shendure[1]

We demonstrate the successful application of exome sequencing[1–3] to discover a gene for an autosomal dominant disorder, Kabuki syndrome (OMIM%147920). We subjected the exomes of ten unrelated probands to massively parallel sequencing. After filtering against existing SNP databases, there was no compelling candidate gene containing previously unknown variants in all affected individuals. Less stringent filtering criteria allowed for the presence of modest genetic heterogeneity or missing data but also identified multiple candidate genes. However, genotypic and phenotypic stratification highlighted *MLL2*, which encodes a Trithorax-group histone methyltransferase[4]: seven probands had newly identified nonsense or frameshift mutations in this gene. Follow-up Sanger sequencing detected *MLL2* mutations in two of the three remaining individuals with Kabuki syndrome (cases) and in 26 of 43 additional cases. In families where parental DNA was available, the mutation was confirmed to be *de novo* ($n = 12$) or transmitted ($n = 2$) in concordance with phenotype. Our results strongly suggest that mutations in *MLL2* are a major cause of Kabuki syndrome.

Kabuki syndrome is a rare, multiple malformation disorder characterized by a distinctive facial appearance (**Supplementary Fig. 1**), cardiac anomalies, skeletal abnormalities, immunological defects and mild to moderate mental retardation. Originally described in 1981 (refs. 5,6), Kabuki syndrome has an estimated incidence of 1 in 32,000 (ref. 7), and approximately 400 cases have been reported worldwide. The vast majority of reported cases have been sporadic, but parent-to-child transmission in more than a half dozen instances[8] suggests that Kabuki syndrome is an autosomal dominant disorder. The relatively low number of cases, the lack of multiplex families and the phenotypic variability of Kabuki syndrome have made the identification of the gene(s) underlying this disorder intractable to conventional approaches of gene discovery, despite aggressive efforts.

We sequenced the exomes of ten unrelated individuals with Kabuki syndrome: seven of European ancestry, two of Hispanic ancestry and one of mixed European and Haitian ancestry (**Supplementary Fig. 1** and **Supplementary Table 1**). Enrichment was performed by hybridization of shotgun fragment libraries to custom microarrays followed by massively parallel sequencing[1–3]. On average, 6.3 gigabases of sequence were generated per sample to achieve 40× coverage of the mappable, targeted exome (31 Mb). As with our previous studies, we focused our analyses here primarily on nonsynonymous variants, splice acceptor and donor site mutations and coding indels, anticipating that synonymous variants were far less likely to be pathogenic. We also predicted that variants underlying Kabuki syndrome are rare, and therefore likely to be previously unidentified. We defined variants as previously unidentified if they were absent from all datasets used for comparison, including dbSNP129, the 1000 Genomes Project, exome data from 16 individuals previously reported by us[2,3] and 10 exomes sequenced as part of the Environmental Genome Project (EGP).

Under a dominant model in which each case was required to have at least one previously unidentified nonsynonymous variant, splice acceptor and donor site mutation or coding indel variant in the same gene, only a single candidate gene (*MUC16*) was shared across all ten exomes (**Table 1** and **Supplementary Table 2**). However, we considered *MUC16* as a likely false positive due to its extremely large size (14,507 amino acids). Potential explanations for our failure to find a compelling candidate gene in which newly identified variants were seen in all affected individuals included: (i) Kabuki syndrome is genetically heterogeneous and therefore not all affected individuals will have mutations in the same gene; (ii) we failed to identify all mutations in the targeted exome; and (iii) some or all causative mutations were outside of the targeted exome, for example, in noncoding regions or unannotated genes. To allow for a modest degree of genetic heterogeneity and/or missing data, we conducted a less stringent analysis by looking for candidate genes shared among subsets of affected individuals. Specifically, we searched

[1]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [2]Department of Pediatrics, University of Washington, Seattle, Washington, USA. [3]Seattle Children's Hospital, Seattle, Washington, USA. [4]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan. [5]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. [6]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Hokkaido, Japan. [7]These authors contributed equally to this work. Correspondence should be addressed to J.S. (shendure@u.washington.edu) or M.J.B. (mbamshad@u.washington.edu).

Received 28 April; accepted 21 July; published online 15 August 2010; doi:10.1038/ng.646

Table 1 Number of genes common to any subset of *x* affected individuals.

Subset analysis (any *x* of 10)	1	2	3	4	5	6	7	8	9	10
NS/SS/I	12,042	8,722	7,084	6,049	5,289	4,581	3,940	3,244	2,486	1,459
Not in dbSNP129 or 1000 Genomes	7,419	2,697	1,057	488	288	192	128	88	60	34
Not in control exomes	7,827	2,865	1,025	399	184	90	50	22	7	2
Not in either	6,935	2,227	701	242	104	44	16	6	3	1
Is loss-of-function (nonsense or frameshift indel)	753	49	7	3	2	2	1	0	0	0

The number of genes with at least one nonsynonymous variant (NS), splice-site acceptor or donor variants (SS) or coding indel (I) are listed under various filters. Variants were filtered by presence in dbSNP or 1000 Genomes (not in dbSNP129 or 1000 genomes) and control exomes (not in control exomes) or both (not in either); control exomes refer to those from 8 Hapmap[3], 4 FSS[3], 4 Miller[2] and 10 EGP samples. The number of genes found using the union of the intersection of *x* individuals is given.

for subsets of *x* out of 10 exomes having ≥1 previously unidentified variant in the same gene, with $x = 1$ to $x = 10$. For $x = 9$, $x = 8$ and $x = 7$, previously unidentified variants were shared in 3 genes, 6 genes and 16 genes, respectively (**Table 1**). However, there was no obvious way to rank these candidate genes.

We speculated that genotypic and/or phenotypic stratification would facilitate the prioritization of candidate genes identified by subset analysis. Specifically, we assigned a categorical rank to each individual with Kabuki syndrome based on a subjective assessment of the presence of, or similarity to, the canonical facial characteristics of Kabuki syndrome (**Supplementary Fig. 1**) and the presence of developmental delay and/or major birth defects (**Supplementary Table 1**). The highest-ranked individual was one of a pair of monozygotic twins with Kabuki syndrome. We then categorized the functional impact (that is, nonsense versus nonsynonymous substitution, splice-site disruption and frameshift compared to in-frame indel) of each newly identified variant in candidate genes shared by each subset of two or more ranked cases. Manual review of these data highlighted distinct, previously unidentified nonsense variants in *MLL2* in each of the four highest-ranked cases. After sequential analysis of phenotype-ranked cases with a loss-of-function filter, *MLL2* was the only candidate gene remaining after addition of the second individual (**Table 2**). We found no such variant in *MLL2* in the individual with Kabuki syndrome ranked fifth; hence, the number of candidate genes dropped to zero after the individual ranked fourth in the set (**Table 2**). However, we found a 4-bp deletion in the individual ranked sixth, and we found nonsense variants in the individuals ranked seventh and ninth. Thus, exome sequencing identified a nonsense substitution or frameshift indel in *MLL2* in seven of the ten individuals with Kabuki syndrome analyzed here.

Retrospectively, we applied a loss-of-function filter to the subset analysis of exome data (**Table 1**), and at $x = 7$, found *MLL2* to be the only candidate gene. We also developed a *post hoc* ranking of candidate genes based on the functional impact of the variants present (variant score) and the rank of the cases in which each variant was observed (case score). When this was applied to the exome data as a combined metric, *MLL2* emerged as the top candidate gene (**Supplementary Fig. 2**).

In parallel with these analyses, we applied genomic evolutionary rate profiling (GERP)[9] to the exome data. GERP uses mammalian genome alignments to define a rejected substitution score for each variant regardless of functional class. We have previously shown that the quantitative ranking of candidate genes by the rejected substitution scores of their variants can facilitate the exome-based analysis of Mendelian disorders[10]. Following subset analysis with GERP-based ranking, *MLL2* remained on the candidate list up to $x = 8$, ranking third in a list of 11 candidate genes at this threshold (**Table 3** and **Supplementary Fig. 3**). Notably, the additional *MLL2* variant contributing to this analysis (such that *MLL2* was still considered at $x = 8$) was a synonymous substitution with a rejected substitution score of 0.368 in the individual ranked fifth.

We sought to confirm all newly identified variants in *MLL2*, particularly because loss-of-function variants identified through massively parallel sequencing have a high prior probability of being false positives. All seven loss-of-function variants in *MLL2* were validated by Sanger sequencing. We further analyzed the three cases in which we did not initially find a loss-of-function variant in *MLL2*, first by array comparative genomic hybridization (aCGH) to determine any gross structural changes and then by Sanger sequencing of all exons of *MLL2* in case of false negatives by exome sequencing. Because an average of 96% of the coding bases in *MLL2* were called at sufficient quality and coverage for single nucleotide variant detection, we anticipated that any missed variants were more likely to be indels because of the higher coverage required for confident indel detection in short-read sequence data. Indeed, although aCGH did not find any structural variants in the region, Sanger sequencing did identify frameshift indels in two of these three cases (specifically, the cases ranked eighth and tenth).

Ultimately, loss-of-function mutations in *MLL2* were identified in nine out of ten cases in the discovery cohort (**Fig. 1**), making this gene a compelling candidate for Kabuki syndrome. For validation, we screened all 54 exons of *MLL2* in 43 additional cases by Sanger sequencing. Previously unidentified nonsynonymous, nonsense or frameshift mutations in *MLL2* were found in 26 of these 43 cases (**Fig. 1** and **Supplementary Table 3**). In total, through either exome sequencing or targeted sequencing of *MLL2*, 33 distinct *MLL2* mutations were identified in 35 of 53 families (66%) with Kabuki syndrome (**Fig. 1** and **Supplementary Table 3**). In each of 12 cases for which DNA from both parents was available, the *MLL2* variant was found to have occurred *de novo*. Three mutations were found in two individuals each. One of these three mutations was confirmed to have arisen *de novo* in one of the cases, indicating that some mutations in individuals with Kabuki syndrome are recurrent. In addition, *MLL2* mutations (resulting in p.4527K>X and p.5464T>M) were also identified in each of two families in which Kabuki syndrome was transmitted from parent to child.

Table 2 Number of genes common in sequential analysis of phenotypically ranked individuals

Sequential analysis	1	+2	+3	+4	+5	+6	+7	+8	+9	+10
NS/SS/I	5,282	3,850	3,250	2,354	2,028	1,899	1,772	1,686	1,600	1,459
Not in dbSNP129 or 1000 Genomes	687	214	145	84	63	54	42	40	39	34
Not in control exomes	675	134	50	26	13	13	8	5	4	2
Not in either	467	89	34	18	9	8	4	4	3	1
Is loss-of-function (nonsense/frameshift indel)	25	1	1	1	0	0	0	0	0	0

Variants were filtered as in **Table 1**. Exomes were added sequentially to the analysis by ranked phenotype; for example, column "+3" shows the number of genes at the intersection of the three top ranked cases (**Supplementary Fig. 1**). The gene with at least one NS/SS/I in all individuals is *MUC16*, which is very likely to be a false positive due to its extreme length (14,507 amino acids).

Table 3 Analysis of exome variants using genomic evolutionary rate profiling

GERP score analysis (at least *x* of 10)	1	2	3	4	5	6	7	8	9	10
Variant RS score > 0	7,176	2,360	754	269	106	39	20	11	3	1
MLL2 rank	3,732	1,232	399	136	47	14	6	3	NA	NA

The number of genes with at least a single previously unidentified variant with a rejected substitution score[10] > 0 in at least *x* individuals is given. A gene rank is assigned based on the average GERP score[9] over all newly identified variants with rejected substitution score > 0 in all affected individuals.

None of the additional *MLL2* mutations was found in 190 control chromosomes from individuals of matched geographical ancestry.

Our results strongly suggest that mutations in *MLL2* are a major cause of Kabuki syndrome. *MLL2* encodes a large 5,262-residue protein that is part of the SET family of proteins, of which Trithorax, the *Drosophila* homolog of MLL, is the best characterized[11]. The SET domain of MLL2 confers strong histone 3 lysine 4 methyltransferase activity and is important in the epigenetic control of active chromatin states[12]. In mice, loss of *Mll2* on a mixed 129Sv/C57BL/6 background slows growth, increases apoptosis and retards development, leading to early embryonic lethality due in part to misregulation of homeobox gene expression[13]. However, no morphological defects have been reported in $Mll2^{+/-}$ mice[13].

Most of the *MLL2* variants identified in individuals with Kabuki syndrome are predicted to truncate the polypeptide chain before translation of the SET domain. Though it is not certain whether Kabuki syndrome results from haploinsufficiency or from a gain of function at *MLL2*, haploinsufficiency seems to be the more likely mechanism. Deletion of chromosome 12q12–q13.2, which encompasses *MLL2*, has been reported in a child with characteristics of Noonan syndrome[14]. However, we re-analyzed this case using oligo aCGH (including 21 probes that cover *MLL2*) and found the distal breakpoint to be located ~700 kb proximal to *MLL2* (data not shown). Also, all of the pathogenic missense variants identified here are located in regions of *MLL2* that encode C-terminal domains. This suggests that missense variants elsewhere in *MLL2* may be better tolerated or, alternatively, may be embryonically lethal.

For the 18 of 53 cases for which no previously unidentified protein-altering variant was found, it is possible that noncoding or other missed mutations in *MLL2* are responsible for this disorder. Alternatively, Kabuki syndrome could be genetically heterogeneous, and further analysis of these cases by exome sequencing may elucidate additional genes for Kabuki syndrome and potentially explain some of the phenotypic heterogeneity seen in this disorder. Notably, 9 of 10 individuals in the discovery cohort (90%), but only 26 of 43 individuals in the replication cohort (60%), were ultimately found to have mutations in *MLL2*. It is therefore possible that the careful selection of canonical Kabuki cases for the discovery cohort enriched for a shared genetic basis. This underscores the importance of access to deeply phenotyped and well-characterized cases.

In summary, we applied exome sequencing of a small number of unrelated individuals with Kabuki syndrome to discover that mutations in *MLL2* underlie this disorder. As predicted in previous analyses[2,3], allowing for even a small degree of genetic heterogeneity or missing data substantially confounds exome analysis by increasing the number of candidate genes consistent with the model of inheritance. To facilitate the prioritization of genes under such criteria, we stratified data by ranked phenotypes and found that *MLL2* was prominent in the higher ranked cases. However, nine of the ten individuals with Kabuki syndrome in the discovery cohort were ultimately found to have *MLL2* mutations, such that stratification by phenotype was of less importance than originally appeared to have been the case. Nonetheless, the sequential analysis of ranked cases may have reduced the probability of confounding due to genetic heterogeneity. All of the *MLL2* mutations found in the discovery set via exome sequencing were loss-of-function variants. As a result, *MLL2* ranked highly among candidate genes assessed by predicted functional impact. Such a pattern will likely occur for some, but not all, Mendelian phenotypes subjected to this approach. We anticipate that the further development of strategies to stratify data at both the genotypic and phenotypic level will be critical for exome and whole-genome sequencing to reach their full potential as tools for discovery of genes underlying Mendelian and complex diseases.

URLs. RefSeq 36.3, ftp://ftp.ncbi.nlm.nih.gov/genomes/MapView/Homo_sapiens/sequence/BUILD.36.3/updates/seq_gene.md.gz; Phaster, http://www.phrap.org; SeattleSeq Annotation, http://gvs.gs.washington.edu/SeattleSeqAnnotation/; 1000 Genomes Project, http://www.1000genomes.org/page.php/; dbGaP accession, http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000295.v1.p1.

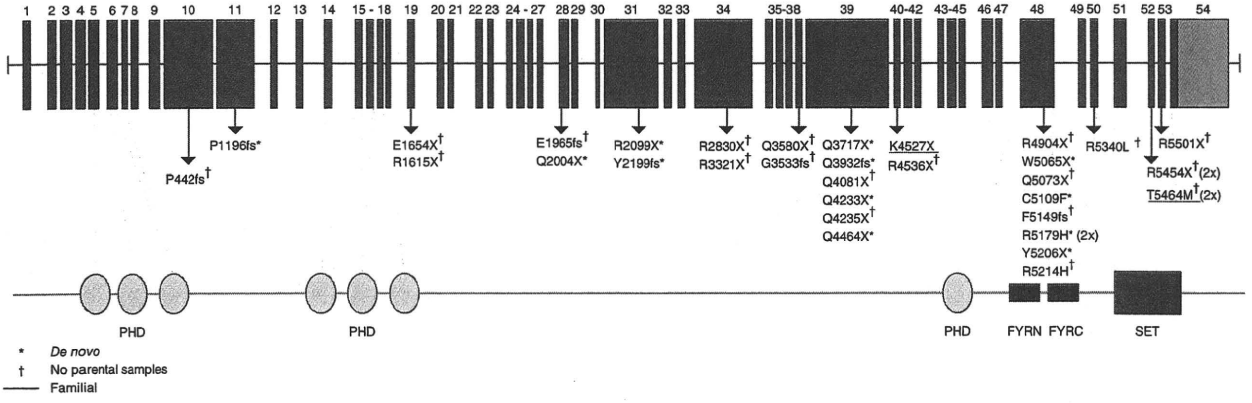

Figure 1 Genomic structure and allelic spectrum of *MLL2* mutations that cause Kabuki syndrome. *MLL2* is composed of 54 exons that encode untranslated regions (orange) and protein coding sequence (blue) including 7 PHD fingers (yellow), FYRN (green), FYRC (green) and a SET domain (red). Arrows indicate the locations of 32 different mutations found in 53 families with Kabuki syndrome including 20 nonsense mutations, 7 indels and 5 amino acid substitutions. Asterisks indicate mutations that were confirmed to be *de novo* and crosses indicate cases for which parental DNA was unavailable. The two underlined mutations were transmitted each within a family, from an affected parent to an affected child.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Exome data for the discovery cohort is available via the NCBI dbGaP repository under accession number phs000295.v1.p1.

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

ACKNOWLEDGMENTS

We thank the families for their participation and the Kabuki Syndrome Network for their support. We thank J. Allanson, J. Carey and M. Golabi for referral of cases and M. Emond for helpful discussion. We thank the 1000 Genomes Project for early data release that proved useful for filtering out common variants. Our work was supported in part by grants from the US National Institutes of Health (NIH)–National Heart, Lung, and Blood Institute (5R01HL094976 to D.A.N. and J.S.), the NIH–National Human Genome Research Institute (5R21HG004749 to J.S., 1RC2HG005608 to M.J.B., D.A.N. and J.S.; and 5RO1HG004316 to H.K.T.), NIH–National Institute of Environmental Health Sciences (HHSN273200800010C to D.N. and M.J.R.), Ministry of Health, Labour and Welfare (K.Y., N.M., T.O. and N.N.), Japan Science and Technology Agency (N.M.), Society for the Promotion of Science (N.M.), the Life Sciences Discovery Fund (2065508 and 0905001), the Washington Research Foundation and the NIH–National Institute of Child Health and Human Development (1R01HD048895 to M.J.B.). S.B.N. is supported by the Agency for Science, Technology and Research, Singapore. A.W.B. is supported by a training fellowship from the NIH–National Human Genome Research Institute (T32HG00035).

AUTHOR CONTRIBUTIONS

The project was conceived and the experiments were planned by M.J.B., D.A.N. and J.S. The review of phenotypes and the sample collection were performed by M.J.B., M.C.H., M.J.M., K.Y., N.M., T.O. and N.N. Experiments were performed by S.B.N., K.J.B., A.E.B., C.L., H.C.M., J.D.S., M.J.R., E.H.T. and H.I.G. Ethical consultation was provided by H.K.T. Data analysis was performed by A.W.B., M.J.B., K.J.B., G.M.C., S.B.N. and J.S. The manuscript was written by M.J.B., S.B.N. and J.S. All aspects of the study were supervised by M.J.B. and J.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
2. Ng, S.B. *et al.* Exome sequencing identifies the cause of a Mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
3. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
4. FitzGerald, K.T. & Diaz, M.O. MLL2: A new mammalian member of the *trx/MLL* family of genes. *Genomics* **59**, 187–192 (1999).
5. Niikawa, N., Matsuura, N., Fukushima, Y., Ohsawa, T. & Kajii, T. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. *J. Pediatr.* **99**, 565–569 (1981).
6. Kuroki, Y., Suzuki, Y., Chyo, H., Hata, A. & Matsui, I. A new malformation syndrome of long palpebral fissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. *J. Pediatr.* **99**, 570–573 (1981).
7. Niikawa, N. *et al.* Kabuki make-up (Niikawa-Kuroki) syndrome: a study of 62 patients. *Am. J. Med. Genet.* **31**, 565–589 (1988).
8. Courtens, W., Rassart, A., Stene, J.J. & Vamos, E. Further evidence for autosomal dominant inheritance and ectodermal abnormalities in Kabuki syndrome. *Am. J. Med. Genet.* **93**, 244–249 (2000).
9. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
10. Cooper, G.M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods* **7**, 250–251 (2010).
11. Prasad, R. *et al.* Structure and expression pattern of human *ALR*, a novel gene with strong homology to *ALL-1* involved in acute leukemia and to *Drosophila* trithorax. *Oncogene* **15**, 549–560 (1997).
12. Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol. Cell. Biol.* **27**, 1889–1903 (2007).
13. Glaser, S. *et al.* Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423–1432 (2006).
14. Tonoki, H., Saitoh, S. & Kobayashi, K. Patient with del(12)(q12q13.12) manifesting abnormalities compatible with Noonan syndrome. *Am. J. Med. Genet.* **75**, 416–418 (1998).

ONLINE METHODS

Cases and samples. For exome sequencing, we selected ten individuals of self-reported European, Hispanic or mixed European and Haitian ancestry with Kabuki syndrome from ten unrelated families. Phenotypic data were collected from review of medical records, phone interviews and photographs. All participants provided written consent, and the Institutional Review Boards of Seattle Children's Hospital and the University of Washington approved all studies. The clinical characteristics of the 43 individuals in the validation cohort who had been diagnosed with Kabuki syndrome have been reported previously[7]. Subjective assessment and ranking of the Kabuki phenotype was based on pictures of each subject (**Supplementary Fig. 1**) and clinical information (**Supplementary Table 1**). Informed consent was obtained for publication of each of the facial photos shown.

Exome definition, array design and target masking. We targeted all protein-coding regions as defined by RefSeq 36.3. Entries were filtered for the following: (i) CDS as the feature type, (ii) transcript name starting with "NM_" or "-", (iii) reference as the group_label, (iv) not being on an unplaced contig (for example, 17|NT_113931.1). Overlapping coordinates were collapsed for a total of 31,922,798 bases over 186,040 discontiguous regions. A single custom array (Agilent, 1M features, aCGH format) was designed to have probes over these coordinates as previously described[3], except here, the maximum melting temperature (T_m) was raised to 73 °C.

The mappable exome was also determined as previously described[3] using this RefSeq exome definition instead. After masking for 'unmappable' regions, 30,923,460 bases were left as the mappable target.

Targeted capture and massive parallel sequencing. Genomic DNA was extracted from peripheral blood lymphocytes using standard protocols. Five micrograms of DNA from each of ten individuals with Kabuki syndrome was used for construction of a shotgun sequencing library as described previously[3] using paired-end adaptors for sequencing on an Illumina Genome Analyzer II (GAII). Each shotgun library was hybridized to an array for target enrichment; this was then followed by washing, elution and additional amplification. Enriched libraries were then sequenced on a GAII to get either single-end or paired-end reads.

Read mapping and variant analysis. Reads were mapped and processed largely as previously described[3]. In brief, reads were quality recalibrated using Eland and then aligned to the reference human genome (hg18) using Maq. When reads with the same start site and orientation were filtered, paired-end reads were treated like separate single-end reads; this method is overly conservative and hence the actual coverage of the exomes is higher than reported here. Sequence calls were performed using Maq and these calls were filtered to coordinates with ≥8× coverage and consensus quality ≥ 20.

Indels affecting coding sequences were identified as previously described[3], but we used phaster instead of cross_match and Maq. Specifically, unmapped reads from Maq were aligned to the reference sequence using phaster (version 1.100122a) with the parameters -max_ins:21 -max_del:21 -gapextend_ins:-1 -gapextend_del:-1 -match_report_type:1. Reads were then filtered for those with at most two substitutions and one indel. Reads that mapped to the negative strand were reverse complemented and, together with the other filtered reads, were remapped using the same parameters to reduce ambiguity in the called indel positions. These reads were then filtered for (i) having a single indel more than 3 bp from the ends and (ii) having no other substitutions in the read. Putative indels were then called per individual if they were supported by at least two filtered reads that started from different positions. An 'indel reference' was generated as previously described[3], and all the reads from each individual were mapped back to this reference using phaster with default settings and -match_report_type:1. Indel genotypes were called as previously described[3].

To determine the novelty of the variants, sequence calls were compared against 16 individuals for whom we had previously reported exome data[2,3] and 10 EGP exomes. Annotations of variants were based on NCBI and UCSC databases using an in-house server (SeattleSeqAnnotation). Loss-of-function variants were defined as nonsense mutations (premature stop) or frame-shifting indels. For each variant, we also generated constraint scores as implemented in GERP[10].

***Post hoc* ranking of candidate genes.** Candidate genes were ranked by summation of a case score and variant score. The case score was calculated by counting the total number of Kabuki exomes in which a variant was identified at a given gene, weighted for case rank from 1 to 10. For example, the top ranked case was weighted by a factor of 10, whereas the case ranked tenth was weighted by a factor of 1. The variant score was calculated by first counting the total number of nonsense, nonsynonymous and synonymous variants across the ten Kabuki exomes and assigning a prior probability of the occurrence of each variant type per gene based upon the target of 18,918 genes. Next, for each candidate gene shared among two or more Kabuki exomes, the scores for each newly identified variant were summed across the gene. The case score and variant score were summed as the candidate gene score.

Mutation validation. Sanger sequencing of PCR amplicons from genomic DNA was used to confirm the presence and identity of variants in the candidate gene identified via exome sequencing and to screen the candidate gene in additional individuals with Kabuki syndrome.

Array comparative genomic hybridization (CGH). Samples were hybridized to commercially available whole-genome tiling arrays consisting of one million oligonucleotide probes with an average spacing of 2.6 kb throughout the genome (SurePrint G3 Human CGH Microarray 1x1M, Agilent Technologies). Twenty-one probes on this array covered *MLL2* specifically. Data were analyzed using Genomics Workbench software according to the manufacturer's instructions.

ARTICLE

SMOC1 Is Essential for Ocular and Limb Development in Humans and Mice

Ippei Okada,[1,14] Haruka Hamanoue,[1,2,14] Koji Terada,[3] Takaya Tohma,[4] Andre Megarbane,[5] Eliane Chouery,[5] Joelle Abou-Ghoch,[5] Nadine Jalkh,[5] Ozgur Cogulu,[6] Ferda Ozkinay,[6] Kyoji Horie,[7] Junji Takeda,[7,8] Tatsuya Furuichi,[9,10] Shiro Ikegawa,[9] Kiyomi Nishiyama,[1] Satoko Miyatake,[1] Akira Nishimura,[1] Takeshi Mizuguchi,[1,15] Norio Niikawa,[11,12] Fumiki Hirahara,[2] Tadashi Kaname,[13] Koh-ichiro Yoshiura,[12] Yoshinori Tsurusaki,[1] Hiroshi Doi,[1] Noriko Miyake,[1] Takahisa Furukawa,[3] Naomichi Matsumoto,[1,*] and Hirotomo Saitsu[1,*]

Microphthalmia with limb anomalies (MLA) is a rare autosomal-recessive disorder, presenting with anophthalmia or microphthalmia and hand and/or foot malformation. We mapped the MLA locus to 14q24 and successfully identified three homozygous (one nonsense and two splice site) mutations in the SPARC (secreted protein acidic and rich in cysteine)-related modular calcium binding 1 (*SMOC1*) in three families. *Smoc1* is expressed in the developing optic stalk, ventral optic cup, and limbs of mouse embryos. *Smoc1* null mice recapitulated MLA phenotypes, including aplasia or hypoplasia of optic nerves, hypoplastic fibula and bowed tibia, and syndactyly in limbs. A thinned and irregular ganglion cell layer and atrophy of the anteroventral part of the retina were also observed. Soft tissue syndactyly, resulting from inhibited apoptosis, was related to disturbed expression of genes involved in BMP signaling in the interdigital mesenchyme. Our findings indicate that SMOC1/Smoc1 is essential for ocular and limb development in both humans and mice.

Introduction

Microphthalmia with limb anomalies (MLA [MIM 206920]), also known as Waardenburg anophthalmia syndrome or ophthalmoacromelic syndrome, is a rare autosomal-recessive disorder first described by Waardenburg.[1] It is characterized by ocular anomalies ranging from mild microphthalmia to true anophthalmia and by limb anomalies such as oligodactyly, syndactyly, and synostosis of the 4th and 5th metacarpals.[2–4] The genetic cause for MLA has remained unknown.

It is widely known that secreted signaling molecules such as Sonic hedgehog (Shh), wingless-type MMTV integration site family (Wnt), transforming growth factor β (Tgf-β), bone morphogenetic proteins (Bmps), and fibroblast growth factor (Fgf) are involved in the development of many organs and tissues, including the eyes and limbs.[5,6] In particular, mutations in *BMP4* (MIM 112262) have resulted in anophthalmia with systemic manifestations, including polydactyly and/or syndactyly (also known as micropthalmia, syndromic 6, MCOPS6 [MIM 607932]),[7] highlighting importance of BMP signaling in both the developing eye and limb.

SMOC1 (MIM 608488), which encodes SPARC (secreted protein acidic and rich in cysteine)-related modular calcium binding 1, is a member of the SPARC (also known as BM-40) matricellular protein family that modulates cell-matrix interaction by binding to many cell-surface receptors, the extracellular matrix, growth factors, and cytokines.[8,9] SMOCs are extracellular glycoproteins with five domains: an N-terminal follistatin-like (FS) domain, two thyroglobulin-like (TY) domains, a domain unique to SMOC, and an extracellular calcium-binding (EC) domain.[9] SMOC1 is widely expressed in various tissues with localization to basement membranes.[9,10] Although the biological function of SMOC1 remains largely unknown, it has been recently reported that *Xenopus* smoc protein, the ortholog of human SMOC1, acts as a BMP antagonist,[11] suggesting that human SMOC1 can also modulate BMP signaling.

Here, we demonstrate that *SMOC1* mutations cause MLA. We also show that *Smoc1* null mice recapitulated

[1]Department of Human Genetics, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan; [2]Department of Obstetrics and Gynecology, Yokohama City University Graduate School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan; [3]Department of Developmental Biology, Osaka Bioscience Institute, 6-2-4 Furuedai, Suita, Osaka 565-0874, Japan; [4]Division of Pediatrics, Okinawa Prefectural Nanbu Medical Center & Children's Medical Center, 118-1 Ikyoku, Arakawa, Haebaru, Okinawa 901-1193, Japan; [5]Medical Genetics Unit, St. Joseph University, Beirut 1104-2020, Lebanon; [6]Department of Pediatrics, Ege University Faculty of Medicine, 35100 Bornova-Izmir, Turkey; [7]Department of Social and Environmental Medicine, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan; [8]Center for Advanced Science and Innovation, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan; [9]Laboratory for Bone and Joint Disease, Center for Genomic Medicine, RIKEN, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan; [10]Laboratory Animal Facility, Research Center for Medical Sciences, Jikei University School of Medicine, 3-25-8, Nishi-Shimbashi, Minato-ku, Tokyo 105-8461, Japan; [11]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Ishikari-Tobetsu, Hokkaido 061-0293, Japan; [12]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Sakamoto 1-12-4, Nagasaki 852-8523, Japan; [13]Department of Medical Genetics, University of the Ryukyus Faculty of Medicine, 207 Uehara, Nishihara, Okinawa 903-0215, Japan

[14]These authors contributed equally to this work

[15]Current address: Laboratory of Biochemistry and Molecular Biology, National Cancer Institute, National Institutes of Health, Building 37, Room 6050, Bethesda, MD 20892, USA

*Correspondence: naomat@yokohama-cu.ac.jp (N.M.), hsaitsu@yokohama-cu.ac.jp (H.S.)

DOI 10.1016/j.ajhg.2010.11.012.

MLA phenotypes, indicating that *SMOC1* plays essential roles in both eye and limb development in humans and mice.

Subjects and Methods

Subjects

A total of four families with one or two cases of MLA were analyzed in this study, including three previously reported families (A, B, and C).[12,13] Family X from Turkey, which has been previously described,[14] was newly recruited to this study. Detailed clinical information of all the patients is available in the literature,[12,14] and phenotypes of patients with confirmed mutations are summarized in Table S1 (available online). A total of five affected and 16 unaffected members from the four families were analyzed in the linkage study. Genomic DNA was obtained from peripheral-blood leukocytes with the use of QuickGene 610-L (Fujifilm, Tokyo, Japan) after informed consent had been given. Experimental protocols were approved by the institutional review board of Yokohama City University School of Medicine.

SNP Genotyping, and Fine Mapping with Short Tandem Repeat Markers

Whole-genome SNP genotyping, with the use of GeneChip Human Mapping 50K Array XbaI (Affymetrix, Santa Clara, CA), and fine mapping of possible candidate regions, with the use of additional microsatellite markers, were performed as previously described.[12,15] The list of primers used for fine mapping is presented in Table S2.

Linkage Analysis

Multipoint linkage analyses using aligned SNPs were performed with ALLEGRO software.[16] Two-point linkage analyses of candidate regions were performed with the LINKAGE package MLINK (FASTLINK software, version 5.1). In each program, an autosomal-recessive model of inheritance with complete penetrance and a disease-allele frequency of 0.001 were applied.

Mutation Analysis of Candidate Genes

All coding exons and exon-intron boundaries of *RAD51L1* (MIM 602948), *ACTN1* (MIM 102575), *ERH* (MIM 601191), *SRSF5* (MIM 600914), *DCAF5* (MIM 603812), *COX16*, *EXD2*, *GALNTL1*, *SLC39A9*, *KIAA0247*, *MED6* (MIM 602984), *TTC9* (MIM 610488), *MAP3K9* (MIM 600136), and *SMOC1* (transcript variant 1, GenBank accession number NM_001034852.1) were analyzed in the probands of families A, C, and X. The transcript variant 2 of *SMOC1* (GenBank accession number NM_022137.4) is 3 bp shorter than the variant 1, leading to an in-frame amino acid deletion at position 431. PCR was cycled 35 times at 94°C for 30 s, at 60°C for 30 s, and at 72°C for 30–90 s in a total volume of 20 μl containing 30 ng genomic DNA as a template, 0.5 μM forward and reverse primers, 200 μM each deoxyribonucleotide triphosphate (dNTP), 1 × ExTaq buffer, and 0.25 U ExTaq (Takara). All primers were designed with Primer3 software. Detailed information of primers is available upon request. PCR products were purified with ExoSAP (USB) and sequenced with BigDye Terminator 3.1 (Applied Biosystems) on a 3100 Genetic Analyzer. Sequences of patients were compared to reference genome sequences in the UCSC Genome Browser (February 2009 assembly) with Seqscape software, version 2.1 (Applied Biosystems).

Animals

Smoc1 mutant mice, created with the use of the *Sleeping Beauty* transposon system, have been previously described.[17] Line PV384 was provided by the RIKEN BioResource Center through the National BioResource Project of MEXT, Japan. Three independent mouse lines (no. 1 to no. 3), each with a single insertion in intron 1 of *Smoc1*, were bred as heterozygotes. Lines 1 and 3 were backcrossed for at least four generations to a C57BL/6J background. Line 2 was maintained with a mixed background of C57BL/6J and ICR. We mainly analyzed line 1, but we confirmed similar phenotypes in lines 2 and 3. Animals were housed in accordance with protocols approved by the Institutional Animal Care and Use Committee at Yokohama City University, School of Medicine. PCR genotyping of mice was performed with the use of genomic DNA from yolk-sac, ear, or tail biopsies. The following primers were used: PV384-WF, 5′-AAAGGCTGGGAATTGTTTG A-3′; PV384-WR, 5′-TGCAGCTGAAACTGTCTCTCC-3′; PV384-MF, 5′-TGTCCTAACTGACTTGCCAAA-3′. The PV384-WF/PV384-WR primers amplified a 441 bp wild-type (WT) product, and the PV384-MF/PV384-WR primers amplified a 218 bp mutant product.

Southern Hybridization

Genomic DNA was extracted from livers or tail biopsies of PV384 heterozygous (*Smoc1*$^{Tp/+}$) mice via standard protocols. The gene-trap insertions were analyzed by Southern hybridization with the use of 10 μg of *Sac*I-, *Nde*I-, BglII-, and *Eco*RI-digested DNA. The probe (451 bp), which hybridized to the internal ribosome entry site (IRES) in the gene-trap vector, was synthesized with the DIG PCR Probe Synthesis Kit (Roche) with the use of the following primers: 5′-CTAACGTTACTGGCCGAAGC-3′ and 5′-CCCAGATCAGATCCCATACAA-3′. Hybridization, washing, and detection of probes were performed according to the manufacturer's protocol. Images were captured with the FluorChem system (Alpha Innotech).

Cloning of Gene-Trap Insertion Sites

After identification of aberrant DNA fragments by Southern hybridization, *Nde*I-, *Sac*I-, and *Eco*RI-digested DNA from PV384 mice was fractionated by electrophoresis, and appropriately sized fragments containing *Ol1* (*other locus 1*), *Ol2*, and *Ol3* were isolated with a QIAEXII Gel Extraction Kit (QIAGEN). The isolated DNA was self-ligated by Ligation High ver.2 (Toyobo), precipitated with ethanol, and dissolved in 20 μl EB buffer (QIAGEN). Inverse PCR was performed in 25 μl reactions, containing 2 μl ligated DNA, 1 × PCR buffer for KOD FX, 0.4 mM each dNTP, 0.5 μM each primer, and 0.5 U KOD FX DNA polymerase (Toyobo). Primers common to *Ol1*, *Ol2*, and *Ol3* were as follows: Inv-F, 5′- AT CGCCAGTTCTGTATGAACGGTCTGGTCTT-3′; Inv-R, 5′-CCCTC TTTACGTGCCAGCCATCTTAGAGATAC-3′. Confirmatory PCR of gene-trap insertion sites for *Ol1*, *Ol2*, and *Ol3* loci was performed with the use of the following primers: *Ol1*-F, 5′-GAGTGGTATTCA TTGGATTCTGCTGAT-3′; *Ol2*-F, 5′-AAATCCAGCTGGCCAACAGA CTAAG-3′; *Ol3*-F, 5′-TTGCCGGGTAGACTCTATCAAGAACCA-3′; TBAL-R, 5′-CTTGTGTCATGCACAAAGTAGATGTCC-3′. Primer sets of *Ol1*-F/TBAL-R, *Ol2*-F/TBAL-R, and *Ol3*-F/TBAL-R could amplify 175 bp, 607 bp, and 767 bp products, respectively. These PCR primer pairs were also used for genotyping of mice harboring a single insertion at the *Smoc1* locus.

Confirmation of Promoter- and Poly(A)-Trapped Transcripts

Whole embryos at embryonic day 10.5 (E10.5) and E11.5 were stored in RNAlater solution (QIAGEN). Total RNA was extracted from WT, *Smoc1*$^{Tp/+}$, and *Smoc1*$^{Tp/Tp}$ embryos with the use of RNeasy Plus Mini (QIAGEN). One microgram total RNA was subjected to reverse transcription with the use of a PrimeScript 1st Strand Synthesis Kit with random hexamers (Takara). A control reaction with no reverse transcriptase was included in each experiment. PCR was performed in 20 μl reactions, containing 1 μl cDNA, 1 × PCR Buffer for KOD FX, 0.4 mM each dNTP, 0.3 μM each primer, and 0.4 U KOD FX (Toyobo). Primers used are listed below: *Smoc1*-F, 5′-GTCCCCACCTCCCCAAGTGCTTTGA-3′; *LacZ*-R, 5′-TGCCAAAAGACGGCAATATGGTGGAAA-3′; *GFP*-F, 5′-TCACATGGTCCTGCTGGAGTTCGTGAC-3′; *Smoc1*-R, 5′-ACACTTGCTCTGGCCAGCATCTTTGCAT-3′. Primer sets of *Smoc1*-F/*Smoc1*-R, *Smoc1*-F/*LacZ*-R, and *GFP*-F/*Smoc1*-R could amplify native *Smoc1* (366 bp), promoter-trapped transcripts (Tp-LacZ, 500 bp) and poly(A)-trapped transcripts (Tp-GFP, 308 bp), respectively. The PCR conditions were 98°C for 10 s, 68°C for 1 min, for 30 cycles. Primers for *ACTB*[18] were used as an internal control. PCR for *ACTB* was cycled 20 times at 94°C for 20 s, 60°C for 20 s, and 72°C for 30 s in a total volume of 10 μl containing 0.5 μl cDNA, 0.4 μM each primer, 0.2 mM each dNTP, 1 × ExTaq buffer, and 0.5 U ExTaq HS (Takara). All PCR products were electrophoresed on 2% agarose gels.

In Situ Hybridization

Embryos were collected between E9.5 and E13.5. Whole-mount in situ hybridization was carried out as previously described.[19,20] Two fragments of *Smoc1* cDNA were obtained as probes by RT-PCR, with the use of total RNA extracted from livers of E16.5 mouse embryos, and subcloned into pCR4-TOPO (Invitrogen). Primer sequences were as follows: probe 1-F, 5′-GTCTGCTCACGCCCCACT-3′; probe 1-R, 5′-CCTGAACCATGTCTGTGGTG-3′; probe P-F, 5′-CAGGAACAGGAAAGGGAAGA-3′; probe P-R, 5′-AAGGGAAAACCACACAGCAC-3′. PCR products were 1023 bp and 1578 bp, corresponding to nucleotide positions 275–1297 and 1849–3426 of the mouse *Smoc1* cDNA (GenBank accession number NM_001146217.1), respectively. The cDNA fragment amplified with probe P-F and probe P-R primers was identical to the probe used in a previous report.[21] Digoxigenin-labeled sense and antisense riboprobes were synthesized with the use of a digoxigenin RNA labeling kit (Roche). These two different antisense probes demonstrated identical staining patterns, and the control sense probes showed no staining. The expression pattern was confirmed with more than three embryos. In addition, the following probes were used: *Bmp2* (gift from Y. Takahashi),[22] *Sox9* (gift from A. Yamada),[22] *Bmp7* (gift from E.J. Robertson), and *Msx2* (gift from Dr. R.E. Maxson, Jr). The numbers of embryos examined were as follows (numerical quantity for WT, *Smoc1*$^{Tp/+}$, and *Smoc1*$^{Tp/Tp}$, respectively, shown in parentheses): *Msx2* (2, 1, 3) at E11.5; *Bmp2* (3, 0, 3), *Bmp7* (3, 0, 3), *Msx2* (3, 0, 3), and *Sox9* (2, 1, 3) at E12.5; *Bmp2* (1, 2, 3), *Bmp7* (2, 1, 3), *Msx2* (1, 2, 3), and *Sox9* (1, 3, 4) at E13.5. Stained embryos were cleared in glycerol to enable images to be produced with a VHX-1000 digital microscope (Keyence).

Histology

Heads of embryos and newborns were fixed overnight in 4% paraformaldehyde in PBS at 4°C. These embryos were then washed in PBS. Frozen samples were serially sectioned at 16 μm (E14.5) and 20 μm (P0). The numbers of eyes examined (WT, *Smoc1*$^{Tp/+}$, *Smoc1*$^{Tp/Tp}$) were as follows: coronally sectioned at E14.5 (8, 10, 12), coronally sectioned at P0 (8, 10, 6), horizontally sectioned at P0 (2, 2, 4). For evaluation of ventral atrophy of the retina, only the coronally sectioned eyes were used. TB staining was performed according to standard protocols. Forelimbs of mice were fixed in 4% paraformaldehyde in PBS, decalcified in 10% EDTA, and embedded in paraffin. Forelimbs were serially sectioned at 4 μm and stained with hematoxylin and eosin.

Evaluation of Optic Nerve Diameter

The palatine and orbital bones were carefully removed to expose the optic chiasm and optic nerve. During the dissection process, 4% paraformaldehyde in PBS was frequently applied onto the gaps between the bone and optic nerve. Xylene cyanol was applied to enhance the outline of optic nerves at poastnatal day 0 (P0). Photographs of optic nerves were taken with a VHX-1000 digital microscope, and the diameter was measured for right and left optic nerves with the bundled software included with the VHX-1000 instrument.

Skeletal Staining

For skeletal preparations, mice were fixed in 99.5% ethanol after removal of the skin and viscera. Cartilage tissues were stained with 0.015% alcian blue and 20% acetic acid in 75% ethanol for three days at 37°C. After dehydration with 99.5% ethanol for three days, bones were stained with 0.002% alizarin red in 1% KOH. Then skeletons were cleared in 1% KOH for several weeks. For P14 mice, soft tissues were dissolved in 2% KOH before alizarin red staining.

Nile Blue Staining

For the study of apoptosis of hindlimbs at E13.5 and E14.5, Nile blue (NB) staining was performed on the basis of a previously described protocol,[23] except that staining was performed at 37°C (not room temperature). Apoptosis was determined by NB-stained (deceased) cells. After rinsing in Tyrode solution, hindlimbs of control (WT and heterozygous littermates) and homozygous mice were evaluated. Photographs of dorsal aspects were taken with a VHX-1000 digital microscope. Experiments were repeated three times, and reproducible representative results are presented.

Statistical Analysis

Statistical analyses were performed with the use of non-repeated-measures ANOVA followed by Dunnett's post hoc test. The results are given as mean ± standard deviation, and the threshold p value for statistical significance was 0.01.

Results

Identification of Homozygous *SMOC1* Mutations

We have previously mapped the MLA locus to a 422 kb region at 10p11.23 by analyzing three families (one Japanese family [A] and two Lebanese families [B and C]). This region contained only one gene, *MPP7*, in which no mutations were found.[12] After a new Turkish family (X) was added to the analysis, the MLA locus was again searched by homozygosity mapping to the consanguineous families (X, B, and C) and haplotype mapping to family A for detection of compound-heterozygous mutations; however, we could not detect any common regions

among the four families. We then focused on identifying common regions in any three of the four families to allow for locus heterogeneity (Table S3).

A locus at 14q24.1-q24.2, which showed the highest LOD score (3.936) among the candidate regions larger than 2.0 Mb, was highlighted among families A, C, and X. This locus was analyzed with the use of additional microsatellite markers, and a 3.0 Mb region containing 24 genes was identified (Figures 1A and 1B). A total of 14 genes were sequenced, and homozygous mutations were found in *SMOC1*: c.718C>T (p.Gln240X) in family A, c.664+1G>A in family C, and c.378+1G>A in family X (Figures 1C and 1D). All of these homozygous mutations were cosegregated with the disease phenotype, and the parents of the individuals with these mutations were heterozygous carriers (Figure 1C). We could not find any mutations in *SMOC1* in family B, in which MLA is unlinked to the 14q24.1-q24.2 locus. Interestingly, in family A haplotypes of paternal and maternal alleles, each having the same mutation, are completely different (data not shown), suggesting that the same mutation may have occurred in separate events. The c.718C>T mutation was not detected in 289 healthy Japanese controls, including 100 Okinawa islanders. The other two mutations were not detected in ethnically matched controls (54 Lebanese and 99 Turkish subjects, respectively), nor in 289 Japanese controls. The two splice-donor-site mutations (c.664+1G>A and c.378+1G>A) are predicted to abolish a donor site, as predicted by ESE-finder, NetGene2, HSF2.4.1, SpliceView, and BDGP analysis (Table S4). Thus, the three mutations are likely to lead to a loss of functional SMOC1.

Smoc1 Expression in the Developing Eye and Limb in Mice

For the examination of *Smoc1* expression in the developing eye and limb, whole-mount in situ hybridization of mouse embryos was performed. *Smoc1* was expressed in the forebrain, midbrain, hindbrain, pharyngeal arch, somites, and forelimb buds at E9.5 (Figure 2A). At E10.5, *Smoc1* expression was observed in the optic stalk (Figure 2B), and at E11.5, expression was localized to the closure site of the optic cup (Figure 2C). Expression of *Smoc1* in developing limbs between E10.5 and E11.5 was observed in both dorsal and ventral regions, with a broader pattern of expression in dorsal regions, but expression was not detected in the most anterior, posterior, and distal parts of limb buds (Figures 2D and 2E). Expression coinciding with chondrogenic condensation was observed at E12.5 (Figure 2F), and expression then became restricted to future synovial joint regions at E13.5 (Figure 2G). This dynamic expression suggests that *Smoc1* plays a critical role in ocular and limb development.

Ocular and Limb Anomalies in *Smoc1* Null Mice

To investigate the pathological basis of MLA due to the loss of *SMOC1* function, we obtained *Smoc1* mutant mice, PV384.[17] PV384 mice possess gene-trap insertions in the *Smoc1* locus and in three other loci. After PV384 mice were bred with C57BL/6J or ICR mice, we obtained three independent lines (no. 1 to no. 3), each with a sole insertion in intron 1 of *Smoc1* (Figure S1). We mainly analyzed line 1, but we confirmed similar phenotypes in lines 2 and 3. Heterozygous mutant mice ($Smoc1^{Tp/+}$) were healthy and fertile. Homozygous mice ($Smoc1^{TP/TP}$) were null mutants, as they showed no native transcript of *Smoc1* (Figure S1E). Homozygous mice were viable at P0; however, they did not survive beyond the first 3 wks of life (Figure 3B). Their growth was retarded in comparison to WT and heterozygous littermates at P0 and P14 (Figures 3A and 3C). Developmental defects in eyes and optic nerves were evident at E14.5. Homozygous mice had relatively small eyes, and histological examinations revealed aplasia or hypoplasia of optic nerves (in 10 of 12 optic nerves), atrophy of the anteroventral part of the retina (in 11 of 12 eyes), and extension of the retinal pigmented epithelium (RPE) to the optic nerve (in 10 of 12 eyes) (Figures 3D–3I). These abnormalities were also observed at P0 (aplasia or hypoplasia of optic nerves [in 7 of 10 optic nerves], retinal atrophy [in 6 of 6 eyes], and RPE extension [in 3 of 6 eyes with identifiable optic nerves]) (Figures 3J–3M). WT or heterozygous littermates did not show any such abnormalities, except that a few eyes of heterozygous mice showed extension of the RPE at E14.5, but not at P0 (in 2 of 10 and 0 of 12 eyes, respectively). Toluidine blue (TB) staining showed ganglion cell layers that were thinned and irregular to varying degrees in homozygous mice, suggesting a reduced number of retinal ganglion cells (Figures 3J–3K′). Thus, *Smoc1* is required for axon sprouting, elongation, or maintenance of retinal ganglion cells.[24] Hypoplasia of optic nerves was further quantitatively confirmed by macroscopic examination: the average diameter of optic nerves of homozygous mice was significantly smaller than that of WT and heterozygous littermates at P0 and P14 (Figures 3L–3Q). These data clearly demonstrate that loss of *Smoc1* in mice affects development of the body, retina, and optic nerves, in a manner similar to that seen in MLA patients.[3,4]

Newborn homozygous mice could be readily identified by their hindlimb syndactyly and pes valgus, whereas no abnormalities were observed in WT and heterozygous pups (Figure 4 and Table 1). Interestingly, the severity of syndactyly varied between mouse lines: line 1 exclusively showed soft tissue syndactyly, whereas line 2 frequently showed four digits (Figures 4F and 4J). Skeletal preparations with alcian blue and alizarin red revealed that the foot with four digits had four phalanx and five metatarsals with fusion to each other (Figure 4K). Thus the *Smoc1* null mutation resulted in a spectrum of phenotypes, from soft tissue syndactyly to four fused digits, probably due to different genetic backgrounds. Bowed tibiae and hypoplastic fibulae were also consistently observed in homozygous mice (Figures 4H and 4L). The articulation between

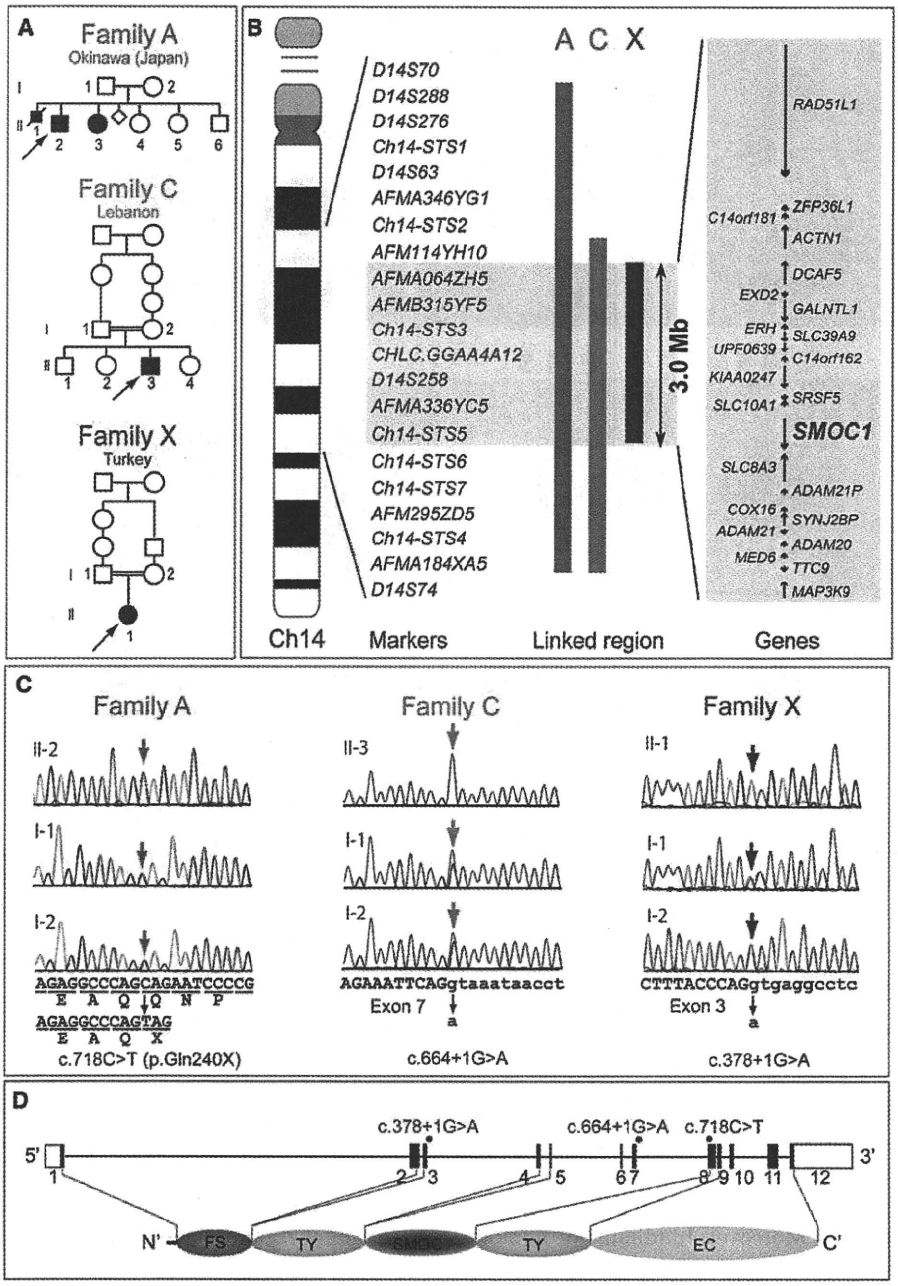

Figure 1. Genetic Analysis of Three Families with Members Affected by Microphthalmia with Limb Anomalies
(A) Pedigrees of the three families.
(B) Linkage analysis with SNPs and microsatellite markers on chromosome 14. From left to right: chromosome ideogram, genetic markers, linked regions of the three families, and genes mapped to the shortest overlapping linked region (between *AFM114YH10* and *Ch14-STS6* [UCSC coordinates, Feb. 2009: chromosome 14: 68,388,190–71,347,908 bp]).
(C) Sequences of mutations identified in each family. Affected patients in family A have a homozygous nonsense mutation (c.718C>T). Patients in families C and X have distinct homozygous splice-donor site mutations (c.664+1G>A and c.378+1G>A, respectively). For all mutations, parents of affected patients are heterozygous carriers, without exception. Sequences of the exon and intron are presented in upper and lower cases, respectively.
(D) At the top is a depiction of a schematic representation of *SMOC1* consisting of 12 exons (UTR and coding exons are indicated by open and filled rectangles, respectively). The locations of three mutations are indicated by red dots. At the bottom, the functional domains of SMOC1 are depicted. Abbreviations are as follows: FS, the follistatin-like domain; TY, the thyroglobulin-like domain; SMOC, the domain unique to SMOC; and EC, the extracellular calcium-binding domain.

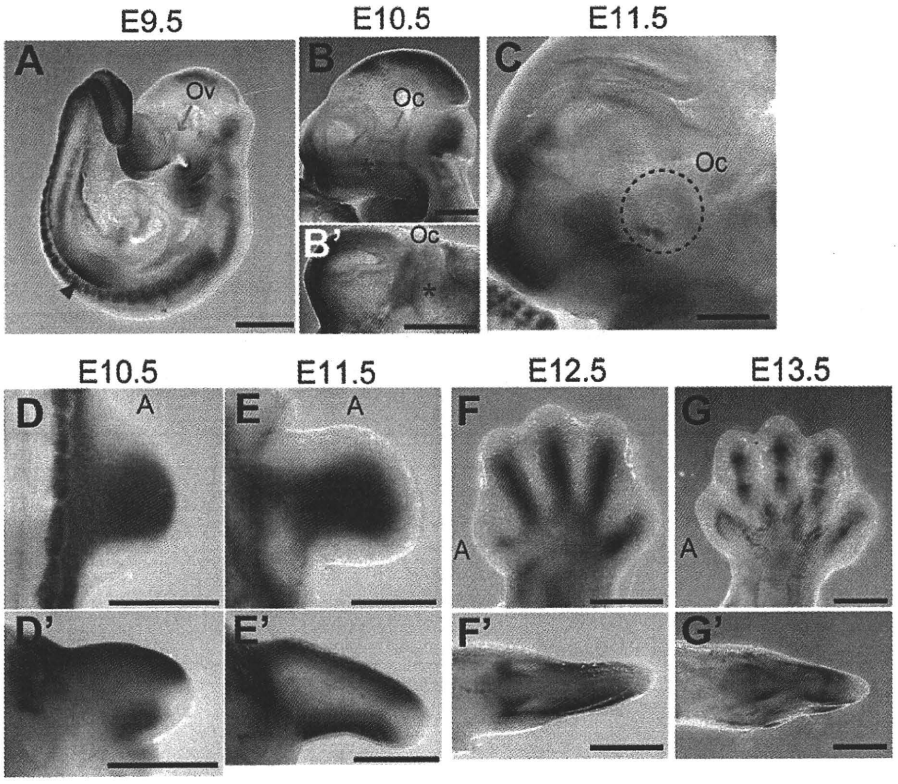

Figure 2. *Smoc1* Expression in Mouse Embryos
Lateral views of embryos (A–C) and a ventral view of the left part of the head (B′, lateral view is shown at the top).
(A) At E9.5, *Smoc1* was expressed in the forebrain, midbrain, hindbrain, pharyngeal arch, somites, and forelimb buds (magenta arrowhead), but not in the optic vesicle (Ov, blue arrow).
(B and B′) Expression in the optic stalk became evident at E10.5 (magenta asterisks), but was not evident in the optic cup (Oc, blue arrow).
(C) Expression was restricted to the closure site of the optic cup (dashed circle) at E11.5.
(D–G) Dorsal and (D′–G′) posterior view of the right hindlimbs (dorsal view is shown at the top in D′–G′). The anterior side is indicated by an A. (D and D′) At E10.5, *Smoc1* was more widely expressed in the dorsal part of the limb bud than in the ventral part. *Smoc1* expression is undetected in the most anterior, posterior, and distal parts of the limb bud. (E and E′) At E11.5, ventral expression was broader than that in the previous stage. (F and F′) At E12.5, expression was detected in areas consistent with chondrogenic condensation. (G and G′) At E13.5, *Smoc1* expression became restricted to future joint regions. Scale bar represents 500 μm.

tibia/fibula and calcanea of homozygous mice appeared malpositioned (Figures 4G and 4K), which might contribute to pes valgus. At P14, soft tissue syndactyly was also evident in most forelimbs of homozygous mice (Figures 4M–4O). Moreover, hindlimbs of homozygous mice showed synostosis between the 4^{th} and 5^{th} metatarsals (Figure 4T), which is observed in both the hands and the feet of MLA patients. Thus, many limb anomalies of MLA patients were recapitulated in *Smoc1* null mice (Table S1).

Reduced Interdigital Apoptosis and Disturbed BMP Signaling

Among the various abnormalities caused by loss of *Smoc1* function, we focused on soft tissue syndactyly, which was commonly observed in both fore- and hindlimbs of null mutants. It is possible that the syndactyly is caused by failed apoptotic regression of the interdigital mesenchyme. To examine this hypothesis, hindlimbs were stained with NB sulfate at E13.5 and E14.5, the time when interdigital apoptosis is most evident. In control embryos (WT and heterozygous littermates), NB-stained apoptotic cells were identified in the interdigital mesenchyme, where regression of the interdigital webbing occurs in the distal region (Figures 5A and 5C). By contrast, the number of apoptotic cells in the mesenchyme between digits 2 and 3 and digits 3 and 4 was dramatically reduced in homozygous mice at E13.5 and E14.5, along with persistent webbing in the distal region (Figures 5B and 5D, magenta asterisk). BMP signaling is involved in apoptosis of the interdigital mesenchyme.[25,26] *Bmp2*, *Bmp7*, and *Msx2*, a direct target of BMP signaling, were strongly expressed in the interdigital mesenchyme of control hindlimbs at both E12.5 and E13.5. However, the expression of these three genes was profoundly reduced and perturbed in hindlimbs of homozygous mice (Figures 5E–5J). These data suggest that inhibition of apoptosis is spatiotemporally correlated to reduced and/or disturbed expression of genes involved in BMP signaling in the interdigital mesenchyme.

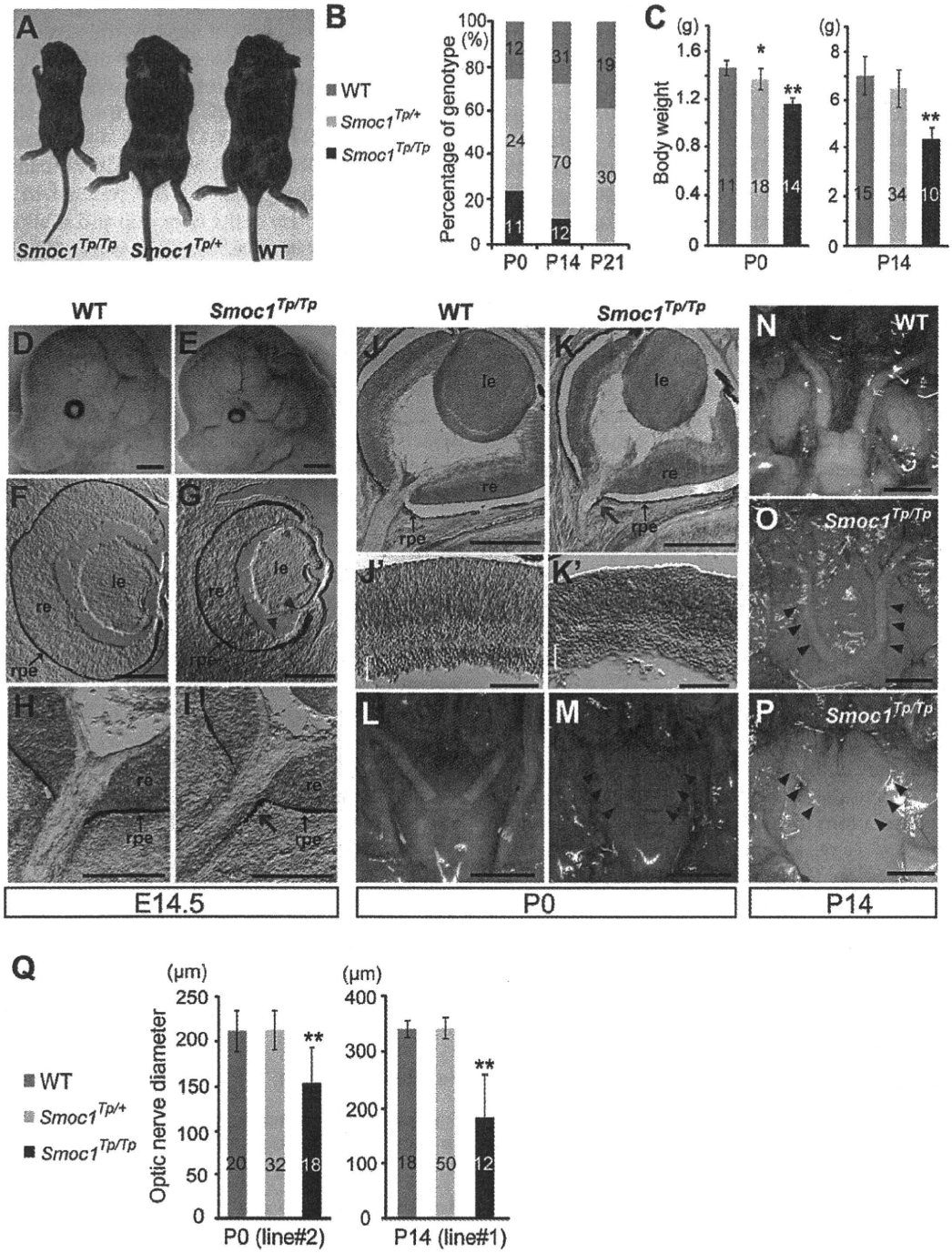

Figure 3. Growth and Ocular Phenotypes of *Smoc1* Null Mice

(A) Representative *Smoc1*$^{Tp/Tp}$ mouse, showing a small body in comparison to *Smoc1*$^{Tp/+}$ and WT littermates.

(B) Genotypes of living pups during the first 3 wk of life.

(C) Body weight of pups of each genotype at P0 (left panel) and P14 (right panel).

(D and E) Relatively small eyes were evident in *Smoc1*$^{Tp/Tp}$ mice in comparison to WT mice.

(F–K′) Coronal sections of eyes at E14.5 (F–I) and P0 (J–K′) with TB staining (H, I, and J–K′). (F-I) Atrophy of the anteroventral part of the retina (G, magenta arrowheads, dorsal view shown at the top), hypoplastic optic nerve, and extension of the RPE to the optic nerve (I, magenta arrow) in *Smoc1*$^{Tp/Tp}$ mice at E14.5. (J and K) Hypoplastic optic nerve and RPE extension in *Smoc1*$^{Tp/Tp}$ mice at P0 (K, magenta arrow). Note that sections in which optic nerves appeared most thick are presented in (H–K). (J′–K′) In higher-magnification views of (J and K), a thinned and irregular ganglion cell layer (white brackets) was observed in *Smoc1*$^{Tp/Tp}$ mice. Abbreviations are as follows: le, lens; re, retina; rpe, retinal pigmented epithelium.

(L–P) Ventral views of the brain showing optic nerves at P0 (L and M) and P14 (N–P), showing various degrees of optic nerve hypoplasia.

Discussion

In a previous report, we performed parametric linkage analysis with three families (families A, B, and C) and found 16 loci showing a LOD score ($\theta = 0.000$) higher than 3.0. Additional microsatellite markers highlighted only one locus, 10p11.23.[12] However, no mutations were found in the candidate gene *MPP7*.[12] By recruiting a new family (family X) to this study, we successfully found homozygous mutations in *SMOC1* in families A, C, and X. In family B, no *SMOC1* mutations were found, indicating the genetic heterogeneity in MLA. Patients with *SMOC1* mutations and *Smoc1* null mice showed similar limb anomalies, such as oligodactyly, syndactyly, synostosis of 4th and 5th metacarpals, hypoplasia of fibula, and bowed tibia. Oligodactyly, syndactyly, and synostosis of 4th and 5th metacarpals are common in MLA patients.[2–4] However, hypoplastic fibula and bowed tibia are less common in patients with MLA, as four out of 34 MLA patients showed these anomalies in the previous report.[3] Although one patient with a *SMOC1* mutation from family C did not show bowed tibia and hypoplastic fibula, these anomalies could be features specific to *SMOC1* mutations. Further *SMOC1* analysis of other MLA patients should delineate the phenotypic consequences caused by *SMOC1* mutations.

Accumulating evidence suggests that BMP signaling plays crucial roles in early eye vesicle and limb patterning, skeletal formation, and apoptosis of the interdigital mesenchyme,[25–29] and mutations involving BMP signaling cause human malformations including ocular, limb, and skeletal anomalies.[7,30–33] Here, we present genetic evidence that *SMOC1* is essential for ocular and limb development in humans and mice. Furthermore, *Xenopus smoc* can inhibit BMP signaling,[11] suggesting that *SMOC1/Smoc1* can also modulate BMP signaling in humans and mice. Indeed, we observed reduced and/or disturbed expression of genes involved in BMP signaling in the interdigital mesenchyme in *Smoc1* null mice, and limb and ocular abnormalities associated with loss of *Smoc1* function are consistent with phenotypic consequences of disturbed BMP signaling. Conditional inactivation of *Bmp2* in the limb showed 3/4 syndactyly, and a similar deficiency of both *Bmp2* and *Bmp7* resulted in malformed fibulae in mice.[25] Moreover, mice deficient in *Fmn1*, a repressor of BMP signaling, showed four digits, fused metatarsal bones, and an absence of fibulae in the hindlimbs,[34] suggesting the importance of altered BMP signaling in these features. Concerning ocular phenotypes, haploinsufficiency of mouse *Bmp4* resulted in a decreased number of ganglion layer cells and absence of the optic nerve similar to *Smoc1* null mice,[35] indicating that altered BMP signaling is also involved in the ocular phenotype. Interestingly, knockdown experiments of *smoc* by antisense morpholino in *Xenopus* showed absence or severe deformity of the eye and other anterior structures, which were accompanied by aberrant expression of *otx2*, *tbx2* in the eye field.[11] Mutations of *OTX2* (MIM 600037) cause micropthalmia, syndromic 5 (MCOPS5 [MIM 610125]) in humans.[36] Moreover, targeted disruption of *Tbx2* resulted in a marked reduction in the size of the optic cup and a failure of optic nerve formation in mice.[37] Thus, it is possible that loss of *SMOC1* function could alter the expression of *OTX2* and *TBX2* (MIM 600747) by disturbing BMP signaling in human developing eyes.

It is unknown how the loss of functional SMOC1, a BMP antagonist, leads to reduced expression of genes involved in BMP signaling in the interdigital mesenchyme in *Smoc1* null mice. In the case of *Fmn1*-deficient mice, the loss of the repressor of BMP signaling resulted in downregulation of *Fgf4* and *Shh* and in upregulation of *Gremlin* expression at E10.5, and absence of apoptosis of the interdigital mesenchyme between the two middle digits at E13.5.[34] Thus, there is a possibility that loss of SMOC1 could cause the imbalance among BMP, SHH, and FGF signaling, which would subsequently lead to reduced and/or disturbed expression of genes involved in BMP signaling in the interdigital mesenchyme. In fact, we observed reduced expression of *Msx2* in the progressive zone of hindlimbs at E11.5 (Figure S2). Moreover, expression of *Sox9*, the initial cartilage condensation marker, showed abnormal limb patterning, suggesting that SMOC1 may affect BMP signaling even at early stages of limb development (Figure S3). Further examinations are required for understanding spatial and temporal actions of SMOC1/Smoc1 protein during limb development.

In conclusion, our data demonstrate that *SMOC1/Smoc1* is an essential player in both ocular and limb development in humans and mice and give further support to the crucial roles of BMP signaling in these systems.

Supplemental Data

Supplemental Data include three figures and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

We would like to thank the patients and their families for their participation in this study. We thank Yoshiko Takahashi (Nara Institute of Science and Technology) and Atsushi Yamada (Showa University) for providing the *Bmp2* and *Sox9* probes; Elizabeth J. Robertson (University of Oxford) and Makoto Ishibashi (Kyoto University) for the *Bmp7* probe; Robert E. Maxson, Jr. (University of Southern California Keck School of Medicine) for the *Msx2*

(Q) Optic nerve diameter. Optic nerves were significantly hypoplastic in $Smoc1^{Tp/Tp}$ mice in comparison to WT and $Smoc1^{Tp/+}$ littermates. The numbers of pups (B and C) or eyes (Q) corresponding to each genotype are indicated within bars. Error bars indicate standard deviation: *$p < 0.01$, compared with WT. **$p < 0.01$, compared with WT and $Smoc1^{Tp/+}$. Scale bars represent 1 mm (D, E, and L–P), 200 μm (F–I), 500 μm (J and K), and 100 μm (J′ and K′).

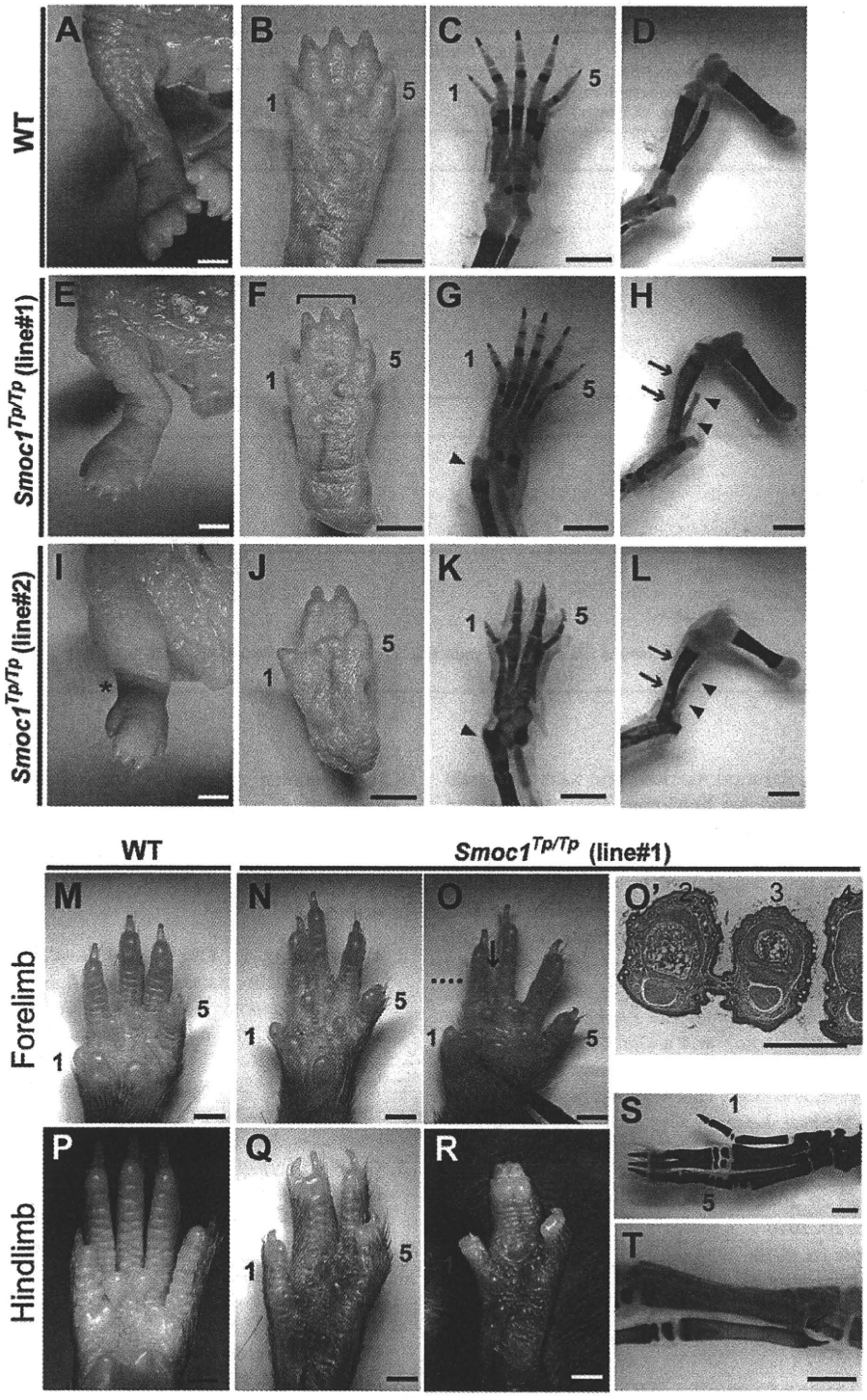

Figure 4. Limb Phenotypes of *Smoc1* Null Mice

Limbs of WT (A–D, M, and P) and *Smoc1*$^{Tp/Tp}$ mice (E–L, N–O′, and Q–T) at P0 (A–L) and P14 (M–T). Digit identities are indicated by the numbers 1 (thumb, anterior) and 5 (little finger, posterior). Skeletal staining with alcian blue and alizarin red is presented (C, D, G, H, K, L, S, and T). *Smoc1*$^{Tp/Tp}$ mice showed pes valgus (E and I), soft tissue syndactyly (F and G), and four digits with metatarsal fusion (J and K). Malposition of the articulation between the tibia/fibula and the calcanea (G and K, magenta arrowheads), bowed tibia (magenta arrows), and hypoplastic fibula (arrowheads) of *Smoc1*$^{Tp/Tp}$ mice (H and L) were observed. 2/3 soft tissue syndactyly (N) and 2/3 webbing (O) were evident in forelimbs of *Smoc1*$^{Tp/Tp}$ mice. (O′) A transverse section taken at the level indicated by the dashed line in (O) showed 2/3 webbing. 2/3 syndactyly (Q), 2/3/4 syndactyly (R), synostosis between the 2nd and 3rd proximal phalanx and metatarsals (S), and synostosis between the 4th and 5th metatarsals (T, arrow), observed in the hindlimbs of *Smoc1*$^{Tp/Tp}$ mice. Scale bars represent 1 mm (A–O and P–T) or 500 μm (O′).