Enhancer adoption caused by genomic insertion elicits interdigital Shh expression and syndactyly in mouse

Kousuke Mouri, Tomoko Sagai, Akiteru Maeno, Takanori Amano, Atsushi Toyoda, and Toshihiko Shiroishi

Mammalian Genetics Laboratory, Genetic Strains Research Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan; and Comparative Genomics Laboratory, Center for Information Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

Acquisition of new cis-regulatory elements (CREs) can cause alteration of developmental gene regulation and may introduce morphological novelty in evolution. Although structural variation in the genome generated by chromosomal rearrangement is one possible source of new CREs, only a few examples are known, except for cases of retrotransposition. In this study, we show the acquisition of novel regulatory sequences as a result of large genomic insertion in the spontaneous mouse mutation Hammer toe (Hm). Hm mice exhibit syndactyly with webbing, due to suppression of interdigital cell death in limb development. We reveal that, in the Hm genome, a 150-kb noncoding DNA fragment from chromosome 14 is inserted into the region upstream of the Sonic hedgehog (Shh) promoter in chromosome 5. Phenotyping of mouse embryos with a series of CRISPR/Cas9-aided partial deletion of the 150-kb insert clearly indicated that two different regions are necessary for the syndactyly phenotype of Hm. We found that each of the two regions contains at least one enhancer for interdigital regulation. These results show that a set of enhancers brought by the large genomic insertion elicits the interdigital Shh expression and the Hm phenotype. Transcriptome analysis indicates that ectopic expression of Shh up-regulates Chordin (Chrd) that antagonizes bone morphogenetic protein signaling in the interdigital region. Indeed, Chrd-overexpressing transgenic mice recapitulated syndactyly with webbing. Thus, the Hm mutation provides an insight into enhancer acquisition as a source of creation of novel gene regulation.

Significance

In this study, we reexamined an old mouse mutant named Hammer toe (Hm), which arose spontaneously almost a half century ago and exhibits a limb phenotype with webbing. We revealed that a 150-kb noncoding genomic fragment that was originally located in chromosome 14 has been inserted into a genomic region proximal to Sonic hedgehog (Shh), located in chromosome 5. This inserted fragment possesses enhancer activity to induce Shh expression in the interdigital regions in Hm, which in turn down-regulates bone morphogenetic protein signaling and eventually results in syndactyly and web formation. Since the donor fragment residing in chromosome 14 has enhancer activity to induce interdigital gene expression, the Hm mutation appears to be an archetypal case of enhancer adoption.

Author contributions: T.A. and T. Shiroishi designed research; K.M., T. Sagai, and A.M. performed research; K.M., T. Sagai, and A.T. analyzed data; and K.M. wrote the paper.

Data deposition: Microarray data have been deposited in the Gene Expression Omnibus (GEO) database of NCBI, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE98424). ATAC-seq data have been deposited in the DRA of DNA Data Bank of Japan (DDBJ), www.ddbj.nig.ac.jp/index-e.html (accession no. DRA006161).

See Commentary on page 839.

See Commentary on page 839.

1To whom correspondence should be addressed. Email: tshiroi@nig.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713339115/-/DCSupplemental.
expression in turn down-regulates bone morphogenetic protein (BMP) signaling, leading to a reduction of interdigital cell death and finally resulting in the syndactyly and interdigital webbing phenotype. Thus, we provide unequivocal evidence that interchromosomal translocation is able to generate enhancer activity for a developmental gene, which ultimately leads to morphological alteration.

**Results**

*Hm* was reported to be tightly linked to another old mutation named Hemimelic extra toe (Hx) that causes hemimelia and preaxial polydactyly. Only one recombinant was detected between *Hx* and *Hm* in a previous study (18). Since a single nucleotide substitution was identified in MFCS1 as the causative mutation of *Hx* (13, 14), we first sequenced MFCS1 from the *Hm* genome. The result did not show any kind of mutation in MFCS1.

A subsequent linkage study narrowed the critical region of *Hm* to a 420-kb interval between two DNA markers, D5kng2 and D5kng3 (19) (Fig. 1A). We extended the linkage analysis based on intercrossovers of the *Hm* heterozygotes (*Hm/+*) that were generated from crosses of *Hm/+* and each of three inbred strains—KJR/Ms, BLG/Ms, and MSM/Ms—established from highly polymorphic wild mice (20) and also a backcross involving a Shh-coding knockout (KO) mouse (Fig. 1B). In total, we obtained 12 recombinants from these crosses and thereby assigned the DNA marker NIG6 to the most distal border and NIG2 to the most proximal border of the *Hm* region (Fig. 1B). This result confined the *Hm* region to the segment depicted by the green bar in Fig. 1A. By combining the results of the former linkage analysis (19) with these, we further confined *Hm* to a 270-kb interval between D5kng2 and NIG6 (Fig. 1A).

Because *Hm* is tightly linked to the *Shh* locus, we examined whether *Hm* acts in cis on *Shh*. From the same backcross progeny used in the linkage analysis, we obtained four recombinants between the *Shh*-coding KO allele and the DNA marker NIG2 (Fig. 1C). Two of these, referred to as cis1/B6 and cis2/B6, have a recombinant chromosome carrying the *Hm* region and the *Shh*-coding KO allele. They exhibited normal limb phenotype, indicating a cis-acting nature for *Hm*. This was confirmed by the finding that genetic restoration of the *Shh* coding sequence in the cis1 recombinant chromosome yielded the syntactically phenotype (Fig. 1D and E). Thus, this cis–trans test showed unequivocally that the regulation by *Hm* acts in cis on *Shh*.

Because the syntactically of *Hm* is a dominant phenotype and is likely caused by the gain of a new regulatory activity, we anticipated that removing the mutated region from the *Hm* genome could rescue its phenotype. Because two *Shh* enhancers, MFCS4 and MACS1, which are necessary for postnatal viability, are located in the distal side of the 270-kb *Hm* region between NIG4 and NIG6 (21, 22), we first removed 80 kb that does not contain the two enhancers (Fig. 1A, magenta bar). Mice with the deletion showed normal digits, indistinguishable from those of wild-type mice (Fig. 2 A–D). This result indicated that the causative mutation is contained within the 80 kb on the reference mouse genome (GRCm38/mm10), which does not contain any coding sequence (Fig. 1A).

Next, to explore the genomic variation responsible for *Hm* in this 80 kb, we amplified genomic fragments using tiling PCR primer pairs from *Hm*-homozygous and wild-type mice (Fig. 1A). We found that no fragment was amplified with the primer pair F specifically from the *Hm* genome, suggesting a structural variation in a long interspersed nuclear element (LINE) sequence that resides in the 80 kb. Subsequent inverse PCR identified the structural variation in this LINE sequence (Fig. S1B). Sequencing revealed that the fragment amplified by the inverse PCR contains abnormal junctions of Chr5 and Chr14 (Fig. S1C). We confirmed the presence of these junctions by genomic PCR (Fig. 2F). These results indicated that a 150-kb fragment residing in the intergenic region between Slitk1 and Slitk6 in Chr14 is inserted into the region upstream of *Shh* in Chr5 (Fig. 2G). No coding sequences or noncoding RNAs are annotated in this 150-kb fragment. We termed this fragment the *Hm* inserted sequence (abbreviated as HIS).

To determine whether the HIS insertion directly causes the syntactically phenotype of *Hm*, we removed the LINE sequence containing HIS from the *Hm* genome using the CRISPR/Cas9 system (Fig. 2G and Dataset S2). A mouse line with this deletion showed normal digits that were indistinguishable from those of wild-type mice (Fig. 2E). The homozygote of the HIS-deleted *Hm* mutant was viable and did not show any abnormal phenotype in the whole limb or in other organs. This result indicated that the original LINE sequence in the *Hm* region in Chr5 has no significant function in mouse development. Thus, we conclude that insertion of HIS was the direct cause of *Hm*.

To test whether the original HIS in the normal mouse Chr14 has enhancer activity, we performed a transgenic reporter assay using BAC clone RP23-444D12, which was isolated from the B6 genome library and harbors the HIS in Chr14 (Fig. 2G).
We made a transgene construct, introducing a LacZ reporter gene into the BAC clone. Transgenic mouse embryos with this construct showed reporter expression in the interdigital region at E13.5 (Fig. 2 H and I). The LacZ signals were more abundant on the ventral side than on the dorsal side. This difference in signal intensity along the dorsoventral axis correlates well with the fact that interdigital cell death in the Hm limb is more severe on the dorsal side than on the ventral side (11). Therefore, HIS in Chr14 of strain B6 appears to possess enhancer activity to elicit gene expression in the interdigital regions of mouse limbs.

To assess the function of the original HIS in wild type, we removed HIS from Chr14 using the CRISPR/Cas9 system. Mice homozygous for the deletion did not show any developmental abnormality, including limb phenotype (Dataset S2), suggesting no significant role for HIS in Chr14 in morphogenesis.

To find regulatory region(s) of HIS, we generated a series of partial deletion of HIS in the Hm genome by the CRISPR/Cas9 system (Fig. S2). Observation of the phenotype of these deletion mutants indicated that the middle part of HIS (region 1) was necessary to cause the syndactyly phenotype of Hm. Moreover, another part of HIS (region 2), which resides at the Shh side and shares no sequence similarity with region 1, was also required for syndactyly.

Next, to identify enhancers in HIS, we searched open chromatin regions by using ATAC-sequencing (ATAC-seq) with distal autopod cells from E13.5 embryos. We found seven peaks in the 150 kb of HIS (magenta bars of e1 to e7 in Fig. 2J). Among them, two peaks (e1 and e4) were not included in regions 1 and 2. We carried out transgenic reporter assays with the five peaks excluding e1 and e4 by inserting each fragment upstream of the hsp promoter and LacZ sequence (Dataset S3). Out of the five peaks to be tested, we found three autopod enhancers (Fig. 2 K–P). The e2 peak induced interdigital expression of LacZ (Fig. 2K). The e5 peak directed the expression of LacZ mainly in digits sometimes with a broad signal in the entire autopod, including interdigit regions (Fig. 2 M and N). The e7 peak induced interdigital expression of LacZ, although the signal was observed in only one embryo (Fig. 2P). The other two peaks induced no LacZ signals in the limb (Fig. 2 L and O).

In the CRISPR/Cas9-mediated deletion experiment of HIS, we obtained G0 mice with a more severe syndactyly phenotype than that in the original Hm heterozygote. Sequencing of PCR products amplified from the Hm genome with a primer pair for the 3′ junction indicated that a CRISPR/Cas9-mediated inversion of HIS had occurred in the G0 mouse (Fig. 3A and Dataset S2). We named this HIS inverted allele Hm-inv and found that mice heterozygous for Hm-inv showed neonatal lethality; in contrast, both heterozygotes and homozygotes of the original Hm are fully viable. The Hm-inv heterozygotes exhibited swollen soles and a cleft palate, which were not observed in the original Hm heterozygotes (Fig. 3B). Sagittal and transverse sections of X-ray micro-computed tomography (micro-CT) images showed that the volume of muscle fiber in the Hm-inv heterozygotes was increased in the swollen sole (Fig. 3 D–I). On the other hand, each phalange was normal (Fig. 3 D–I).

We examined the interaction between Hm and MFCS1, because MFCS1 is a sole enhancer necessary and sufficient for Shh expression in normal limb development (21). We eliminated MFCS1 from the Hm genome using the CRISPR/Cas9 system (Fig. S3A). Among the G0 pups generated, two individuals exhibited truncated limbs, resembling the phenotype of the MFCS1 KO mouse, indicating that MFCS1 was successfully
eliminated from both chromosomes in these two pups (Fig. S3B). We then crossed the G0 male to B6 and established a mouse line in which the MFCS1-eliminated chromosome (ΔMFCS1) is located at the cis-position relative to \textit{Hm}. All mice of this line (\textit{Hm}, ΔMFCS1/+ ) clearly exhibited syndactyly, similar to the original \textit{Hm} heterozygotes (Fig. S3C). Therefore, the presence of MFCS1 in cis relative to \textit{Hm} is not necessary to generate the syndactyly phenotype of \textit{Hm}. In other words, the limb enhancer MFCS1 is dispensable for \textit{Hm} action.

To determine the effect of HIS on Shh regulation, we analyzed the expression level of Shh mRNA, using qRT-PCR, in the \textit{Hm} mutant limb at different stages from E11.5 to E15.5. We first examined Shh expression at the early embryonic stage (E11.5), when Shh expression remains in the ZPA. Expression was observed in both the \textit{Hm} homozygote and the wild-type embryo, but its level was 2.6-fold higher in the \textit{Hm} homozygote (Fig. 4A). After Shh expression ceases in the ZPA, interdigital cell death proceeds from E12.5 to E15.5. During these later stages, significantly higher Shh expression was observed in the \textit{Hm} limb than in the wild-type limb (Fig. 4B).

We examined the dose effect of the Shh expression level on phenotypic severity at E13.5. We found that the expression level correlated well with dose of the \textit{Hm} allele, and the \textit{Hm-inv} allele showed an extremely high level of Shh expression (Fig. 4C). When HIS was removed from Chr5, the heterozygote (HmΔHIS/+ ) did not show ectopic expression of Shh. Thus, the level of ectopic Shh expression correlated with the severity of syndactyly.

Next, we analyzed the localization of ectopic Shh expression in the autopod. RT-PCR of mRNA prepared from the interdigital region at E13.5 showed ectopic expression of Shh in the \textit{Hm} homozygote but not in the wild-type embryo (Fig. 4D). This ectopic Shh expression was not detected in situ hybridization in the \textit{Hm} homozygote, probably due to its lower expression level. On the other hand, in situ hybridization of the \textit{Hm-inv/+ } embryo clearly showed ectopic Shh expression in the interdigital region (Fig. 4E and F, yellow arrowhead), reflecting the high level of Shh mRNA (Fig. 4C). Shh expression was more abundant on the ventral side than the dorsal side, consistent with the \textit{LacZ} expression pattern of the BAC transgenic mouse (Fig. 2f), as well as with the biased interdigital cell death along the dorsoventral axis.

There was no obvious difference in the Shh expression pattern in the limb bud between the \textit{Hm} homozygote and wild-type embryo at the early stage of limb development (E11.5) (Fig. 4G and H). This result is consistent with the finding that \textit{Hm} did not exhibit abnormality in digit number. Taken together, these results indicate that HIS acts as a new regulatory region that induces Shh expression in the interdigital region at the later stages of limb development when Shh expression has normally ceased in the ZPA.

BMP signaling positively regulates interdigital cell death (23). To investigate whether the ectopic Shh expression in \textit{Hm} affects BMP signaling, we examined the expression level of Msx2, a BMP downstream target (24, 25). In situ hybridization showed clearly that Msx2 was down-regulated in the \textit{Hm} mutant (Fig. 5A and B). This result suggested that the ectopic Shh expression suppresses interdigital cell death through BMP down-regulation. To uncover the molecular changes of BMP signaling in the developing \textit{Hm} limb, we analyzed the interdigital transcriptome by microarray and found a total of 892 differentially expressed genes, including \textit{Shh} and Msx2 (\log2(\textit{Hm}/\textit{WT}) > 1; Fig. 5C). We focused on those involved in BMP signaling. In contrast to Msx2, Msi1 expression level was unchanged (Fig. 5D). There was no marked change in expression level of the BMP and BMP receptor genes. On the other hand, among BMP antagonists, only Chordin (\textit{Chrd}) was included in the 892 differentially expressed genes. To test the possibility that the up-regulation of \textit{Chrd} suppresses BMP signaling in the \textit{Hm} limb, we generated a transgenic mouse that over-expresses \textit{Chrd} in the limb. We observed soft tissue syndactyly, similar to the \textit{Hm} phenotype, in the transgenic mouse (Fig. 5 E and F, n = 3/6). This successful phenocopying of syndactyly by the ectopic expression of \textit{Chrd} suggests that \textit{Chrd} changes may also mediate the down-regulation of BMP signaling in the \textit{Hm} limb.

\textbf{Discussion}

Enhancer Adoption by Genomic Insertion in \textit{Hm}. Genetic mechanisms underlying enhancer acquisition are currently categorized into three types: de novo acquisition by accumulation of
nucleotide substitutions, insertion of a TE, and acquisition by chromosomal rearrangement (3). The third type, in which a preexisting enhancer for one gene is captured by another gene, is termed enhancer adoption (26). While de novo acquisition requires a long time for a new enhancer to emerge, the latter two mechanisms, insertion of a TE and chromosomal rearrangement, can take place rapidly and saltationally. Many cases of enhancer acquisition by TE insertion have been reported (27, 28). For example, retrotransposition of a short interspersed element created a new enhancer to regulate Fgf8 expression in mouse forebrain and may have contributed to the evolution of the mammalian neuronal networks (29). By contrast, there are very few reports that enhancer adoption due to chromosomal rearrangement has occurred in evolution (30, 31).

This study revealed that a 150-kb genomic fragment, which was derived from Chr14, was inserted into the region upstream of Shh in Chr5 of the Hm genome. As a consequence, Shh was ectopically expressed in the interdigital region and led to syndactyly and web formation. This syndactyly phenotype is very similar to the previously reported case of enhancer adoption by Shh (26). Removing HIS from Chr5 in the Hm genome restored normal Shh expression and limb morphology, which demonstrated that the insertion of HIS is the direct cause of syndactyly with webbing. Thus, this study provides direct evidence that enhancer adoption can cause morphological alteration. Our BAC-transgenic reporter assay revealed that HIS in Chr14 of wild-type mouse also has enhancer activity in the interdigital region (Fig. 2 A and I). The microarray assay detected expression of Slitrk6, which neighbors HIS, in the interdigital region of wild-type mouse (Dataset S5), implying that HIS at the original position in Chr14 has tissue-specific enhancer activity. On the other hand, Slitrk6 KO mice show a reduction of innervation in the cochlea but no external abnormality (32). Given that removing HIS from mouse Chr14 did not cause any visible phenotype, any regulatory activity of the original HIS in Chr14 has little impact on limb development.

The result of a series of partial deletion experiments clearly showed that at least two regions in HIS are necessary for the syndactyly phenotype of Hm (Fig. S2). One of them, region 1, contains the interdigital enhancer e2 (Fig. 2K). On the other hand, region 2 contains two enhancers, e5 and e7, which induce reporter expression broadly in autopod and specifically in interdigits, respectively (Fig. 2 M, N, and P). These results imply that neither one of the three enhancers represents the total regulatory activity of HIS. Instead, it is most likely that a cooperative activity of these enhancers is necessary for the interdigital regulation of Shh in Hm. It is consistent with recent reports that many genes are regulated by multiple enhancers (33–35). Thus, in general, insertion of a large genomic fragment such as HIS can cause enhancer adoption, bringing a set of enhancers, which eventually leads to morphological change.

Recent whole-genome sequencing analyses have revealed a vast number of structural variants, a median of 18.4 Mbp per diploid genome, in human individuals. It is interesting to note that they are enriched around expression-quantitative trait loci (36). In addition, potential enhancers exist throughout the vertebrate genome (37). Taking all these findings into account, it is possible that numerous sequences in human genomes have potential enhancer activity and that enhancer adoption by chromosomal rearrangement occurs frequently in human populations. Recent reports that certain genetic diseases are caused by enhancer adoption (26, 38) support this assertion.

The limb enhancer MFCSI is known to topologically interact with the Shh promoter beyond the long distance (39), and we found a small or negligible effect of HIS on the enhancer activity of MFCSI. This implies that frequent enhancer adoption can occur at the genome-wide scale, adding novel expression domains without disturbing original enhancer activity.

We found that the HIS inversion in Chr5 dramatically changed Shh expression level and resulted in a much severer morphological defect than that of the original Hm heterozygote. In addition to syndactyly, the inversion caused other abnormalities including increased muscle volume in the autopod and cleft palate, the latter of which is perhaps the cause of perinatal lethality. Few studies to date have documented orientation-dependent enhancer activity (40, 41). Recently, it was reported that orientation-dependent CTCF binding sites determine topological DNA looping for promoter–enhancer interaction and thereby exert proper gene regulation (42). We could not identify any CTCF binding sites in the original HIS in Chr14 by in silico searching of public databases. It is possible that another unknown orientation-sensitive sequence, rather than CTCF binding sites, exists in HIS.

**Ectopic Shh Expression and Downstream Targets.** Ectopically expressed Shh in the Hm mutant up-regulates Chrd expression and down-regulates Msx2, a BMP signaling marker (Fig. 5), BMP signaling positively regulates interdigital cell death (43, 44). Since CHRD antagonizes BMP signaling (45), it is reasonable that up-regulation of Chrd via ectopic Shh expression results in inhibition of interdigital cell death. Indeed, our transgenic expression of Chrd clearly recapitulated syndactyly, consistent with earlier findings that ectopic expression of Gremlin (Grem) and Noggin (Nog), two other BMP antagonists, also inhibited interdigital cell death (46, 47). This molecular model of syndactyly in Hm is consistent with previous reports that ectopic expression of Shh suppresses interdigital cell death (26, 48).

Proper spatiotemporal gene regulation is essential for executing the developmental program. When the ectopic expression level of Shh was elevated by the inversion of HIS, digit patterning did not change. It is known that ectopic Shh expression in the anterior region of the early stage limb bud causes polydactyly (14, 17). In another mutant, Short digits (Dsh), ectopic Shh expression in the joint region causes brachydactyly with shortened digit length (49). Thus, both the timing and the pattern of ectopic expression of Shh are crucial for proper limb development. A change of Shh expression induces different downstream targets of Shh. In the early limb bud (approximately E9.0 to E12), in contrast to CHRD, another BMP antagonist, GREM, is induced.
under the control of SHH signaling (50). This sensitivity of GREM to SHH signaling disappears approximately when Shh expression ceases in the ZPA, which results in termination of the SHH–GREM–FGF feedback loop (51). This switching of sensitivity from early to later stages in limb development likely alters the downstream pathway of SHH signaling. In the Hm limb, we could not detect any difference in Fgf4 expression, suggesting that early developmental events relying on the SHH–FGF feedback loop terminated properly in the Hm limb.

Reduction of interdigital cell death, which leads to interdigital web formation, arose recurrently in independent lineages of different taxa. In this convergent evolution, gain of Grem expression was purportedly a molecular mechanism by which interdigital webbing emerged in the limbs of duck and bat (46, 52). We propose that gain of expression of Chrd, another BMP antagonist, may also have contributed to interdigital web formation in nature.

Materials and Methods

Microarray data have been deposited in the GEO of NCBI under accession no. GSE98424 and are available in Dataset S5. ATAC-seq data have been deposited in the DRA of DDBJ under accession no. DRA006161. An extended description of the materials and methods is provided in SI Materials and Methods.

Animal experiments in this study were approved by the Animal Care and Use Committee of National Institute of Genetics (NIG).

ACKNOWLEDGMENTS. We thank Y. Mizushima, K. Fukunaga, A. Kondo, Y. Suzuki, and H. Nakazawa for generating transgenic mice and mutants; Dr. A. McMahon for the Shh riboprobe and Dr. R. Hill for the Mx2 riboprobe; the staff of Comparative Genomics Laboratory at NIG for supporting genome sequencing. This study was supported by Grants-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JSPS) (Grants JP15J06985, JP17K15162, and JP17K19411).