RESEARCH

Open Access



Functional evidence supports the potential role of *Tbx4*-HLEA in the hindlimb degeneration of cetaceans

Zhenhua Zhang¹, Yao Liu¹, Na Liang¹, Zhenpeng Yu¹, Luoying Deme¹, Duo Xu¹, Jia Liu¹, Wenhua Ren¹, Shixia Xu^{1*} and Guang Yang^{1,2*}

Abstract

The evolution of limb morphology plays an important role in animal adaptation to different ecological niches. To fully adapt to aquatic life, cetaceans underwent hindlimb degeneration and forelimb transformed into flipper; however, the molecular mechanisms underlying the limb changes in cetaceans remain unclear. We previous study had shown that the Tbx4 hindlimb enhancer A (Tbx4-HLEA) in cetaceans exhibited specific deletions and nucleotide substitutions, with significantly reduced regulatory activity. To further investigate whether cetacean HLEA has a potential impact on hindlimb development in vivo, a knock-in mouse model was generated by knocking in the homologous cetacean HLEA in the present study. Phenotypic analysis showed a significant reduction in hindlimb bud development in homozygous knock-in mice at embryonic day (E)10.5; however, the phenotypic difference was rescued after E11.5. Transcriptomic and epigenetic analyses indicated that the cetacean HLEA acts as an enhancer in the mouse embryos and significantly reduces the transcriptional expression levels of Tbx4 at E10.5, supporting that downregulation of cetaceans HLEA regulatory activity reduces the expression of Tbx4. Additionally, both the number of activated noncoding elements and chromatin accessibility near Tbx4 were increased in homozygous knock-in mice at E11.5. The functional redundancy of enhancers compensated for the functional defect of cetacean HLEA, rescuing the expression level of Tbx4, and may account for the phenotype restoration after E11.5. In conclusion, our study suggested that the evolution of cetacean HLEA may be an important link with relevant molecular mechanism for the hindlimb degeneration.

Keywords Cetaceans, Tbx4-HLEA, Hindlimb degeneration, Phenotypic rescue, Knock-in mice

*Correspondence: Shixia Xu xushixia78@163.com Guang Yang gyang@njinu.edu.cn ¹ Jiangsu Key Laboratory for the Conservation and Utilization of Biodiversity in the Middle and Lower Reaches of the Yangtze River, College of Life Sciences, Nanjing Normal University, Nanjing 210023, China ² Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

Introduction

In the complex natural environment, diversification of limb morphology is particularly important for better survival and movement of animals. Cetacea (whales, dolphins and porpoises), originated from a clade of ancestral artiodactyl about 50 million years ago, fully adapted to the aquatic environment through secondary entry into the water [1, 2]. In order to improve hydrodynamic swimming efficiency, their somatotype evolved into streamlined morphology, especially with the hindlimb degeneration and the flipper-like forelimb, with evidence for the dramatic change in somatotype from a wealth



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

of fossils during the Eocene (56 to 34 million years ago) [3-5]. Although the molecular mechanisms of vertebrate limb development have been minutely explained in model species such as mice and chickens, the relevant molecular evolutionary mechanisms underlying hindlimb degeneration in cetaceans remain largely unknown [6, 7].

Vertebrate limb development involves two signaling centers (the apical ectodermal ridge, AER and the zone of polarizing activity, ZPA) and three axes (the proximodistal axis, the antero-posterior axis, and the dorso-ventral axis) [8]. These important biological processes are tightly regulated by tissue-specific transcriptional regulators in vivo, meanwhile signaling pathways and genes that regulate limb development are highly conserved in vertebrates [9], making it difficult to explore the molecular mechanism underlying the evolution of diverse limbs in different vertebrates. It has been found that mutation or the absence of genes that regulate limb development can cause limb development defects. For example, the hindlimb-specific transcription factor Tbx4 plays an important role in the developmental initiation and morphological construction of hindlimb buds [10, 11]. In humans, Tbx4 haploinsufficiency caused small patella syndrome (SPS), a hindlimb specific defect [12]. Complete inactivation of *Tbx4* in mice resulted in embryonic lethality and impaired hindlimb development [13, 14]. However, in recent years, an increasing number of studies have shown that non-coding regulatory regions of the genome play an important role in the development of tissues and organs (such as limbs, eyes, phallus, heart, etc.) [15–18]. In one of these cases, two enhancers of Tbx4hindlimb enhancer A (HLEA) and hindlimb enhancer B (HLEB)—were identified as regulators of hindlimb development from E10.5 to E12.5, and deletion of HLEA led to downregulation of Tbx4 expression, resulting in abnormal hindlimb development [19].

Compared with other mammals, four specific deletion and substitutions have been identified in cetacean HLEA, and cellular functional experiments have shown that the ability of cetacean HLEA to drive reporter gene expression is significantly lower than that of mice [20]. Enhancers bond with transcription factors to regulate the expression of target genes [21]. ChIP-seq experiments demonstrated that *Pitx1*, a transcription factor that specifies hindlimb morphology, could bind to HLEA and HLEB to regulate Tbx4 expression [22]. The reduced enhancing activity of cetacean HLEA may be due to the loss of a binding site for a key transcription factor involved in limb development caused by sequence deletions. However, the functional significance of modified cetacean HLEA in the hindlimb degeneration and related mechanisms remains unresolved and requires further in vivo experiments.

In this study, the homologous cetacean HLEA was knocked into the genome of mice using CRISPR/Cas9 technology. Compared with wild-type (WT) mice, the hindlimb buds of homozygous knock-in mice (HKM) were significantly reduced at E10.5. In combination with genomics, Cleavage Under Targets and Tagmentation (CUT&Tag) and RNA-seq, it was found that cetacean HLEA also functioned as an enhancer in mice and regulated the expression of many genes that control limb development. Furthermore, phenotypic difference in the hindlimb buds of HKM was rescued after E11.5. Transcriptome analysis showed that the number of significantly differentially expressed genes (DEGs) decreased sharply at E11.5, and in particular, it became almost undetectable at E12.5. Meanwhile, epigenomic analysis revealed an increased number of activated enhancers near Tbx4 in HKM at E11.5 compared with E10.5, suggesting that the functional redundancy of enhancers compensated for the functional defect of cetacean HLEA, thereby rescuing the expression level of target gene Tbx4, and further led to phenotypic recovery after E11.5. Therefore, the evolution of HLEA may be part of a molecular mechanism related to hindlimb degeneration in cetaceans.

Results

Construction of a cetacean Tbx4-HLEA knock-in mouse model

We designed a homologous recombination vector (HRV) containing a 1.1 kb sequence of cetacean HLEA (Fig. S1A). The HRV was then injected into C57BL/6J mouse zygotes to achieve editing of the HLEA locus, enabling phenotypic and omics analyses (Fig. 1A). The genome of the offspring obtained from mating heterozygous mice was extracted and used as a template for polymerase chain reaction (PCR), and the homozygous mice were successfully identified (Fig. S1B). In order to verify the integrity of the editing site, we designed the same pair of primers to amplify the 1.3 kb target sequence from the WT and HKM genome, and the complete cetacean HLEA and four specific deletions were successfully identified in HKM genome (Fig. 1B). In addition, H3K27ac and H3K4me1 for CUT&Tag were selected to further confirm if cetacean HLEA exhibited enhancer activity in HKM, and it was shown that significant signal peaks were observed in the target region (chr11:85876050-85877122) (Fig. 1C, and Table S1). Based on the analysis principle of DESeq (method), the enrichment ratios of H3K27ac and H3K4me1 were logFC=-0.315163887 and logFC=0.18613584, respectively, suggesting that the enhancer activity of HLEA was not species-specific between mice and cetaceans (Fig. 1D, and Table S1).



Fig. 1 Generating a knock-in mouse model with the cetacean HLEA enhancer and performing genotype identification. A Cas9, gRNA and homologous recombination vector were co-injected into C57BL/6J zygotes using the CRISPR/Cas9 system. B Based on PCR amplification and sequencing, the complete cetacean HLEA and four specific deletions were successfully identified in the HKM genome. C CUT&Tag demonstrated that the cetacean HLEA functions as an enhancer in mice (H3K27ac and H3K4me1, which are closely related to enhancer activity, were selected for CUT&Tag experiments. Left: WT, right: HKM). D The enrichment ratios of H3K27ac and H3K4me1 at the HLEA locus in the genomes of WT and HKM were calculated based on the DESeq analysis principles (details are provided in Methods section and Table S1)

Cetacean HLEA homozygous mice exhibited developmental delay in hindlimb buds at E10.5

To further investigate the effect of cetacean HLEA on limb development in mice, hindlimb buds were separated along the body axis for detailed characterization (Fig. 2A). The hindlimb buds of embryos (n=26) for both WT and HKM were measured at E10.5 (Figs. S2A, S2B, and Tables S2, S3, S4), and the results showed that

the hindlimb buds of HKM (right: Mean = 1.45524 mm^2 , left: Mean = 1.46457 mm^2) were significantly smaller than those of WT (right: Mean = 1.67924 mm^2 , left: Mean = 1.68793 mm^2 , right *p-value* = 0.0026 and left *p-value* = 0.0023) (Fig. 2B). To confirm whether this phenotypic difference is retained during later stages of limb development, the hindlimb bud morphology was characterized at E11.5 and E12.5 (Figs. S2C and S2D). The



Fig. 2 Morphometric analyses of the hindlimb in WT and HKM at different developmental stages. **A** Dissecting the hindlimb buds of WT and HKM embryos using Leica Microsystems (E10.5–E12.5). Scale bar, 1:2500 μ m. **B**–**D** ImageJ software was used to measure the area of hindlimb bud area for *t-test* analysis, **B** E10.5, **C** E11.5, **D** E12.5, Scale bar, 1:2500 μ m. **E** Double staining of the hindlimb skeleton elements with alcian blue (blue; cartilage) and alizarin red (violet; bone), top: E18.5 embryos (*n*=14), bottom: P1 embryos (*n*=12). Pg: pelvis girdle; Fe: femur; Ti: tibia; Fi: fibula, D1-D5: digit1, digit2, digit3, digit4, digit5. Scale bar, 1:2500 μ m. **F**, **G** Proportional results of hindlimb skeletal element lengths relative to body length (from the cervical to the caudal spine) at E18.5 and P1. The length of individual elements of the hindlimb skeleton was quantified using LAS X software (https://www.leica-microsystems.com.cn/cn/products/microscope-software/p/leica-las-x-ls/), while body length was measured with vernier calipers for *t-test* analysis, **F** E18.5, **G** P1. Scale bar, 1:2500 μ m. All the above measurements were performed in three independent replicates. The results were shown as Mean ± SD (standard deviation), and asterisks indicate significant differences (*: *p* < 0.05, **: *p* < 0.01, ns: Not significant)

hindlimb buds of HKM at E11.5 were not significantly different from those of the WT (Fig. 2C, Fig. S2B, and Tables S3, S4, S5). Since the division of the limb into stylopod, zeugopod and autopod became more pronounced at E12.5, and the length of embryo could not be effectively measured, we only measured the autopod of the hindlimb bud, and there was no significant difference between HKM and WT (Fig. 2D, Fig. S2B, and Tables S3, S4, S6). In addition, hypodactyly or polydactyly was not found in HKM compared with WT at E18.5 (Fig. 2E). By measuring the length of osteogenesis in the pelvic girdle, femur, tibia, and fibula (normalized to body size based on the length of the ossification), no significant difference was found between HKM and WT (Fig. 2F, Fig. S2E, and Tables S7, S8, S9). Compared with WT, postnatal day 1 (P1) HKM still did not exhibit hypodactyly or polydactyly, and meanwhile, there was no significant difference in the length of bone in the pelvic girdle, femur, tibia, and fibula (Fig. 2E, G, Fig. S2F, and Tables S9, S10, S11).

Cetacean HLEA regulated the expression of downstream genes resulting in hindlimb bud developmental delay at E10.5

RNA-seq analysis was performed for hindlimb buds from WT and HKM at E10.5-E12.5. Before RNA-seq analysis, HKM were confirmed using PCR and agarose gel electrophoresis (Fig. S3A). A total of 1214 DEGs were detected at E10.5, which was much more than that at E11.5 and E12.5, transcripts of 960 and 254 genes were up- and down-regulated respectively in HKM compared to WT (including the target gene *Tbx4*, FDR = 5.74E - 28, log2FC=-1.70773) (Fig. 3A, and Tables S12, S13 and S14). DEGs at E10.5 included many key genes that regulate limb development, such as Hoxd13, Hoxa13, Hand2, *Fgf*8, *Bmp*6, *Pax*3, etc. (Fig. 3B). The representative top 50 most significant genes with altered expression are shown (Fig. 3C). In addition, gene ontology (GO) enrichment analysis showed that DEGs at E10.5 were significantly enriched in biological processes associated with limb development. The downregulated genes were associated with embryonic digit morphogenesis, embryonic hindlimb morphogenesis, proximal/distal pattern formation, limb development and osteoblast differentiation (Fig. 3D), while upregulated GO terms included anterior/ posterior pattern specification and Notch signaling pathway (Fig. S3B). Meanwhile, KEGG analysis showed that DEGs at E10.5 were significantly enriched in MAPK signaling pathway, Wnt signaling pathway, Hippo signaling pathway, Notch signaling pathway and Dorso-ventral axis formation (Fig. 3E and Fig. S3C), suggesting that cetacean HLEA might play a key role for limb development in the knock-in mice. However, thirty-four representative genes closely related to limb development with significantly altered expression at E10.5 did not show significant differential expression at E11.5 and E12.5 (Fig. 3F, and Table S15). RT-PCR further confirmed down-regulation of *Tbx4*, *Hand2*, *Hoxd13* and *Bmp6*, and up-regulation of *Hoxb9*, *Rarb* and *Pax3* at E10.5 compared with WT (Fig. 3G). In the knock-in mouse model, *Hand2*, *Hoxd13* and target gene *Tbx4* began to be up-regulated at E11.5 compared with E10.5, and there was no significant difference compared with WT at E12.5 (Fig. 3H). Taken together, these data suggested that cetacean HLEA may have caused retardation of hindlimb bud development by regulating the expression of downstream genes at E10.5.

The functional redundancy of potential activated

enhancers near Tbx4 in HKM at E11.5 rescues its expression Epigenetic analysis revealed that at E11.5, the number of non-coding elements (including HLEA and potential enhancers) co-marked by H3K27ac and H3K4me1 near *Tbx4* was higher in HKM than in WT (Fig. 4A, and Table S16). Further analysis of the read counts in the same histone-marked regions in WT and HKM showed that, compared to WT, H3K27ac levels were significantly increased in HKM, whereas no significant difference was observed in H3K4me1 levels (Fig. 4B). The result suggested that the potential enhancers near Tbx4 in HKM may be in a more active state, thereby enhancing the regulatory potential for target genes. To further validate this phenomenon, Hoxd13, Msx2, and Twist1 were selected for an analysis of histone-marked noncoding elements. The results showed that at E11.5, compared to WT, the number of histone-marked non-coding elements near these genes remained largely unchanged in HKM (Fig. S4), which is sharply in contrast with the changes observed near *Tbx4*. This finding indirectly supported that the activation and increased activity of noncoding elements near Tbx4 may have been induced by the cetacean HLEA knock-in. Both RNA-seq analysis and RT-PCR showed that the expression of *Tbx4* was up-regulated at E11.5 compared to E10.5 in the knock-in mouse model, which might be related to the increased number of activated non-coding elements and the enhanced activity of potential enhancers near Tbx4. Gene scanning within 200 kb upstream and downstream of the Tbx4 transcription start site (TSS) identified a total of 7 genes, e.g., Bcas3, Tbx2, Gm11444, Tbx4, Brip1, Ints2, and Med13. These genes and activated non-coding elements were mapped, revealing that the majority of activated non-coding elements in the knock-in mice at E11.5 were located near *Tbx4*, potentially strengthening the regulation of Tbx4 expression compared to WT (Fig. 4C). In the RNA-seq data from E10.5 and E11.5, we found that among the seven genes mentioned above, only Gm11444 was not detected in the hindlimb bud, while only Tbx4



Fig. 3 RNA-seq analysis of downstream gene expression regulated by cetacean HLEA. **A** The number of DEGs between WT and HKM during E10.5-E12.5. **B** A volcano plot illustrating differentially regulated gene expression from RNA-seq analysis between WT and HKM at E10.5. Genes that are upregulated and downregulated are shown in red and blue, respectively. Values are presented as the log2 of tag counts. **C** The hierarchical clustering of the RNA-seq analysis results shows that the representative top 50 most significant genes with altered expression at E10.5. **D** Gene ontology (GO) functional clustering of genes that were downregulated for biological processes. **E** KEGG pathway analysis of downregulated targets in HKM transcriptome. **F** The hierarchical clustering of the RNA-seq analysis results shows that thirty-four representative genes closely related to limb development were differentially expressed at E10.5–E12.5. **G** RT-PCR validation analysis of the changes in the expression of *Hand2*, *Hoxd13*, and *Tbx4* from E10.5 to E12.5. Error bar represents the Mean ± SD and *p-value* was generated by *t-test.* **p* < 0.05; ***p* < 0.01



Fig. 4 Epigenetic analysis of hindlimb buds at E11.5 in HKM and WT: CUT&Tag. **A** Normalized H3K27ac (red), and H3K4me1 (blue) epigenetic signals in the region spanning 100 kb upstream and downstream of the *Tbx4* TSS locus for pooled samples per group at E11.5 in the hindlimb bud of HKM and WT. The inverted triangles indicate overlapping peaks of H3K27ac and H3K4me1, with blue inverted triangles marking HKM-specific overlapping peaks that are absent in WT. Furthermore, these HKM-specific peaks do not overlap with the coordinates in the ENCODE public dataset. Top: WT, bottom: HKM. **B** Comparative analysis of normalized read counts for the same potential enhancers co-marked by H3K27ac and H3K4me1 in WT and HKM. Data are mean ± standard, *p-value* was calculated using Student's *t*-test (*, *p* < 0.05; **, *p* < 0.01). Detailed analysis is available in the Materials and Methods and Table S16. **C** Potential activated elements and genes within 200 kb upstream and downstream of the *Tbx4* TSS were mapped to the genome. Top: WT E11.5, bottom: HKM E11.5. **D** Differential expression analysis of *Bcas3*, *Tbx2*, *Tbx4*, *Brip1*, *Ints2*, and *Med13* between WT and HKM from E10.5 (left) to E11.5 (right)

expression exhibited a downward trend compared to WT at E10.5. In contrast, the difference in expression level of *Tbx4* between HKM and WT disappeared at E11.5.

Similarly, the expression levels of *Bcas3*, *Tbx2*, *Brip1*, *Ints2* and *Med13* were neither up-regulated nor down-regulated at E11.5 (Fig. 4D). These results suggest that

the activated non-coding elements and the enhanced activity of potential enhancers in knock-in mice at E11.5 only enhanced *Tbx4* expression.

Discussion

Understanding the molecular mechanisms of limb development in cetaceans has attracted much attention from evolutionary biologists. Previous studies have demonstrated that cetaceans initiated hindlimb bud development at an early embryonic stage, but this process was disrupted due to the interruption of Fgf8 expression and the absence of Hand2 [23]. However, the reason for the interruption of *Fgf*8 expression and the absence of *Hand2* expression have not been addressed so far. With bioinformatics and comparative genomics, some genes regulating limb development, such as *Hoxd11* and *Hoxd13* [24, 25], were found to undergo accelerated evolution and showed specific amino acid substitution in cetaceans. Since gene functions are usually pleiotropic, for example, Hoxd13 not only controls the number of fingers but also plays an important role in suppressing prostate cancer metastasis [26, 27], it remains unclear whether genes with accelerated evolution do contribute to the evolutionary development of limb morphology. In addition, rapid evolution of conserved non-coding regions in genome have provided new insights into limb development in non-model animals [28–30]. Such a classic paradigm came from a study in snakes, which revealed that a 17bp deletion in the ZRS enhancer of Shh gene disrupted Shh protein expression and limb development [31, 32]. Most of the conserved non-coding elements closely related to genes regulating limb development in limb-free lizards had sequence variation [33, 34]. Although the conserved non-coding elements with sequence variation for the key genes of limb development in cetaceans were identified and their biological functions were verified using dual-luciferase reporter assay in the previous study [20, 35], it was still unclear whether they regulated the evolution and development of cetaceans limbs. The reasons were as follows: (a) Pleiotropic functions of conserved non-coding elements, for example, there were shared conserved noncoding elements in limb and penis development [16]. (b) Lack of in vivo experiments to provide clear phenotypic characteristics.

In this study, we successfully constructed a knock-in mouse model with the replacement of mice HLEA by cetacean HLEA. Analyses of the hindlimb morphology found that the hindlimb bud development of HKM was significantly smaller than that of WT at E10.5, however, such morphological difference disappeared from E11.5 to P1. RNA-seq analysis showed that *Hand2*, *Fgf8*, and the target gene *Tbx4* were down-regulated at E10.5. This result was consistent with previous studies on gene

expression during the early development of the hindlimb bud in the Pantropical spotted dolphin (Stenella attenuata) embryos [23], suggesting that Fgf8 expression was not maintained during late development in embryonic period and that loss of Hand2 might lead to the failure of the hindlimb to express Shh and then not to establish ZPA. In addition, the expression level of genes closely related to limb initiation (Hoxb9, Cyp26b1, Isl1, etc.) [36] and development (Hoxd13, Shox2, Nog, etc.) [37, 38] were significantly altered at E10.5 but not at E11.5 and E12.5. Compared with WT, only 53 DEGs were detected at E11.5 (Fig. S5A), and few was detected at E12.5. The CUT&Tag assay showed that at E11.5, the number of activated non-coding elements near Tbx4 in HKM was more than in WT, and the activity of potential enhancers in the same region was higher in HKM than in WT. Studies have shown that non-coding elements marked by both H3K27ac and H3K4me1 are often candidate enhancers and that functional redundancy may exist among enhancers regulating the same gene expression [39, 40]. Therefore, we speculate that the newly acquired potential enhancers in HKM, along with the more active potential enhancers, may have acted in coordination to compensate for the weaken of HLEA function in cetaceans and thus rescued Tbx4 expression, which in turn would affect the expression levels of other genes related to limb development (Fig. S5B) and led to the recovery of limb development. Additionally, further analysis revealed that the coordinates of the potential activated enhancers near Tbx4 in HKM don't overlap with the known enhancer HLEB (chr11: 85960938-85964405) (Fig. 4C). More intriguingly, at E11.5, the HLEB region in both WT and HKM is marked solely by H3K4me1, without the presence of H3K27ac, suggesting that HLEB may be in a silent state at the time of our sampling. This observation provides further support for the hypothesis that the compensatory effect of HLEA may originate from the coordinated action of more active potential enhancers and newly acquired ones. However, we cannot rule out the possibility that HLEB may partially compensate for the weaken of HLEA function at later developmental stages. To explain the morphological change and recovery in the hindlimb bud development in knock-in mice, we hypothesized that some limb development-related redundant enhancers present in mice have become evolutionarily lost or not activated in cetaceans. These redundant enhancers became to function to make compensation for the morphological changes led by the incorporation of cetacean enhancers, and thus cause the late limb morphological recovery. In summary, based on the aforementioned results and hypotheses, it was implied that the functional attenuation of the cetacean HLEA may have been a potential factor in the regression

of their hindlimbs. However, this cannot be considered as the primary causal mechanism for the hindlimb reduction in cetaceans. Some other alternative mechanisms such as the passive decay of enhancers may also have played a role in the hindlimb reduction of cetaceans, which needs to be further investigated in the future.

In conclusion, our study provided evidence that cetacean HLEA played a role to regulate limb development in vivo by constructing a knock-in mice model and the phenomenon of development delay of hindlimb bud at E10.5. RNA-seq and CUT&Tag suggest that the sequence evolution of cetacean HLEA may be a potential molecular mechanism driving the degeneration of cetacean hindlimbs. However, the functional redundancy of enhancers rescued the development delay in the hindlimb of knock-in mice after E11.5, suggesting that the regulatory mechanism of the organism controls the development of tissues and organs is diverse (coding genes and conserved regulatory elements, CREs) [29] and complex (the functional pleiotropy of genes and CREs) [16]. In the course of evolution, cetaceans may have lost the functional redundancy of enhancers related to limb development, which determined their limb morphological changes (hindlimb degeneration and flipper-like forelimb), promoted a streamlined body, and improved hydrodynamic swimming efficiency to better adapt to the aquatic environment. Therefore, using multi-omics analyses to screen more key genes and core regulatory elements with sequence divergence and adaptive evolution signals, in combination with more in-depth functional experimental verification, is an important approach with great potential to uncover the evo-devo molecular mechanisms of cetacean limb changes.

Materials and methods

Knock-in mice model and identification

We knocked the cetacean HLEA into mice via CRISPR/ Cas9 system. Firstly, two sgRNAs (gRNA1: 5'GTATAG GCTTAATTAGC3', PAM: TGG, gRNA2: 5'CTTCAG AAAATGCAAAG3', PAM: AGG, chemical synthesis) targeting sequences near the insertion site were designed and transcribed in vitro, and the donor vector with the inserted fragment was designed and constructed in vitro. Secondly, Cas9 mRNA (GemPharmatech Co.,Ltd), sgRNA and donor vector were co-injected into zygotes. Thereafter, the zygotes were transferred into the oviduct of pseudopregnant ICR females at 0.5 days post-coitum. All the offsprings of ICR females (F0 mice) were identified by PCR and crossing positive F0 mice with WT to build up heterozygous mice. Finally, specific primers were designed based on the cetacean HLEA and mouse HLEA sequences to genotype the offspring of heterozygous knock-in mice and identify homozygous knock-in mice. Cetacean HLEA primers: Forward 5'AAGATGGCGGAC AGAGGCGGGGTTCCT3', Reverse 5'AGAGGCAAG CTGCAGTTCGGTTAACAGGTTAGA3'. Mice HLEA primers: Forward 5'GGCAGCTGCGGAGGTGGCTGT AAA3', Reverse 5'ACTGGAAACGGGCGCTTGCTC AGTGTT3'.

All mice in this study were raised in the SPF animal room of the Animal Resources Center of Nanjing Normal University (NJNU-ARC), and the management of experimental animals was conducted under the guidance of the Animal Use and Ethics Committee of NJNU (ethics approval number: IACUC-20200501). Mice were maintained under a standard 12 h light/dark cycle and monitoring according to NJNU-ARC policies and procedures.

Phenotypic analysis

For embryos at stages E10.5 to E12.5, we strictly followed the mating time of the female and male mice for sampling. Specifically, the female and male mice were housed together at 5 p.m. on the first day, and vaginal plugs were checked before 9 a.m. the next day, which was designated as E0.5. Embryos were then collected at 6 p.m. on days 10, 11, and 12 of pregnancy, and these embryos were considered to be at developmental stages E10.5, E11.5, and E12.5, respectively. All sampling was conducted according to the aforementioned time points in this study. Hindlimb buds at E10.5-E12.5 were separated along the body axis under Leica Microsystems (M165FC) to remove as much excess muscle tissue as possible. The isolated hindlimb buds were photographed at the same magnification (Leica Microsystems Ltd M165FC 1x). To characterize the hindlimb buds for different samples, the area of limb buds was measured using ImageJ software (http://rsbweb.nih.gov/ij/) at the same scale bar (1:2500 µm) [41]. For E18.5 and postnatal day 1 mice, alcian blue/alizarin red (Sangon Biotech (Shanghai) Co., Ltd. A600298, A506786) was used to stain the cartilage and osteogenesis of the limbs [42]. Again, pictures were taken at the same magnification (Leica Microsystems Ltd M165FC 0.73×), and the same scale bar was used to measure osteogenesis length. All the above measurements were performed in three independent replicates. To facilitate the *t-test* (the significance is indicated as *p < 0.05), each set of measurements was tested for normal distribution, and the results are presented in the supplementary material (Tables S3, S4, S8, S9, and S11).

RNA-sequencing analysis

E10.5–E12.5 embryos from WT and HKM were collected from different pregnant mice and divided into three independent replicates. In each replicate, the same number of hindlimb buds was pooled from littermate embryos. Total RNA of each sample was extracted according to the instruction manual of the TRlzol Reagent (Life technologies, California, USA). RNA integrity and concentration were checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). The mRNA was isolated by NEB-Next Poly (A) mRNA Magnetic Isolation Module (NEB, E7490). The cDNA library was constructed following the manufacturer's instructions of NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, E7530) and NEBNext Multiplex Oligos for Illumina (NEB, E7500). The constructed cDNA libraries were sequenced on a flow cell using an Illumina HiSeq[™] sequencing platform. Low quality reads, such as only adaptor, unknown nucleotides >5%, or Q20 < 20% (percentage of sequences with sequencing error rates <1%), were removed by perl script. The clean reads that were filtered from the raw reads were mapped to mice (mm10) genome (GRCm38) using Tophat2 [43] software. The aligned records from the aligners in BAM/SAM format were further examined to remove potential duplicate molecules. Gene expression levels were estimated using FPKM values (fragments per kilobase of exon per million fragments mapped) by the Cufflinks software [44]. DESeq [45] and Q-value were employed and used to evaluate differential gene expression between WT and knock-in mice. After that, gene abundance differences between those samples were calculated based on the ratio of the FPKM values. The false discovery rate (FDR) control method was used to identify the threshold of the *P-value* in multiple tests in order to compute the significance of the differences. Here, only gene with an absolute value of log2 ratio \geq 2 and FDR significance score < 0.05 were used for subsequent analysis. Genes were compared against various protein database by BLASTX, including the National Center for Biotechnology Information (NCBI) non-redundant protein (Nr) database, Swiss-Prot database with a cut-off E-value of 1E-5. Furthermore, genes were searched against the NCBI non-redundant nucleotide sequence (Nt) database using BLASTn by a cut-off *E-value* of 1E–5. Genes were retrieved based on the best BLAST hit (highest score) along with their protein functional annotation. To annotate the gene with gene ontology (GO) terms, the Nr BLAST results were imported into the Blast2GO program [46]. GO annotations for the genes were obtained by Blast2GO. This analysis mapped all of the annotated genes to GO terms in the database and counted the number of genes associated with each term. Perl script was then used to plot GO functional classification for the unigenes with a GO term hit to view the distribution of gene functions. The obtained annotation was enriched and refined using TopGo (R package). The gene sequences were also aligned to the Clusters of Orthologous Group (COG) database to predict and classify functions [47]. KEGG pathways were assigned to the assembled sequences by perl script.

Quantitative reverse-transcription PCR (qRT-PCR)

E10.5-E12.5 embryos from WT and HKM were collected from different pregnant mice for three independent experiments. Per group was the same number pooled hindlimb buds from littermate embryos. Total RNA was extracted according to the instruction manual of the RNA isolater Total RNA Extraction Reagent (Vazyme R401-01) and reverse-transcribed into cDNA using HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme R223-01) according to the manufacturer's instructions. Expression of mRNAs was determined using ChamQ SYBR qPCR Master Mix (2× without ROX) (Vazyme Q321-02). The relative expression level of the target was calculated using the comparative Ct method. GAPDH was used as an internal control to normalize sample differences. Significance was assessed by *t*-test (*: p < 0.05, **: p < 0.01). The sequences of the primers used for RT-PCR analysis are presented in Table S17.

Identification of potential enhancers: CUT&Tag analysis

E10.5–E11.5 embryos from WT and HKM were collected from different pregnant mice. For each sample used for sequencing, an equal number of hindlimb buds from different female offspring was pooled. The genotype of all embryos was recharacterized prior to the experiments. We analyzed the number of enhancers within 100 kb of the upstream and downstream of the *Tbx4* TSS using CUT&Tag. H3K27ac and H3K4me1, which are closely related to enhancer activity [40, 48], were selected to perform CUT&Tag. The overlapping regions of H3K27ac and H3K4me1 signal peaks were used to define activated enhancer.

CUT&Tag

CUT&Tag assays were performed were performed by Wuhan IGENEBOOK Biotechnology *Co., Ltd* (http:// www.igenebook.com) according to the previously described [49]. Briefly, 50,000 cell was harvested and incubated with concanavalin A coated magnetic beads for 15 min at room temperature (RT). Then, beadbound cells was resuspended and incubated with the appropriate primary antibody (H3K4me1, CST5326; H3K27ac, ab4729) and IGG (2729s, CST) overnight at 4 °C. Afterwards, Goat anti-rabbit secondary antibody (ab206, abcam) was added to this mixture for 30 min at RT. The magnet stand was used to remove unbound antibodies. pA-Tn5 adapter complex was added at this step at RT for 1 h, then, remove unbound pA-Tn5 protein. Next, cells were resuspended in tagmentation buffer at 37 °C for 1 h. After that, immunoprecipitated DNA was saved. To amplify libraries, 21 µL DNA was mixed with 2 µL of a universal i5 and a uniquely barcoded i7 primer, using a different barcode for each sample. A volume of 25 µL NEB Next HiFi 2×PCR Master mix was added and mixed. The sample was placed in a Thermocycler with a heated lid using the following cycling conditions: 72 °C for 5 min (gap filling); 98 °C for 30 s; 14 cycles of 98 °C for 10 s and 63 °C for 30 s; final extension at 72 °C for 1 min and hold at 8 °C. Library was sequenced on Illumina novaseq 6000 with PE 150 method. Trimmomatic (version 0.36) was used to filter out low-quality reads [50]. WT clean reads were mapped to the mm10 genome by Bwa (version 0.7.15), while HKM clean reads were mapped to the mm10 genome-1 which had undergone sequence substitution in the region of chr11:85376050-86377122 [51]. Samtools (version 1.3.1) was used to remove potential PCR duplicates [52]. MACS2 software (version 2.1.1.20160309) was used to call peaks by default parameters (bandwidth, 300 bp; model fold, 5, 50; *p-value*, 0.00001). Read coverage in candidate enhancer regions was calculated for different samples, and differences between the two groups were assessed. Statistical analyses were conducted using GraphPad Prism (GraphPad Software, Boston, MA, USA). Two-tailed tests were performed to determine *p-values*, with statistical significance set at p < 0.05 [53, 54]. Subsequently, based on the analysis principle of Deseq [55], the enrichment ratio of H3K27ac and H3K4me1 marking HLEA in the genomes of WT and HKM were calculated.

$$FC (Fold Change) = \frac{norm(HKM)}{normWT}$$

Data standardization makes data comparable across different samples:

$$Norm = reads_{num} \times \frac{1000000}{clean_{reads}}$$

reads_{num} is the Number of reads enriched in target region. clean_{reads} is the total number of Reads obtained by filtering Raw Reads. log2FC is the ratio of norm between two samples (groups) was taken as the logarithm base 2. An absolute value greater than 1 indicates a difference.

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s13227-025-00239-5.

Supplementary Material 1 Supplementary Material 2 Page 11 of 12

Acknowledgements

We thank all the members of Jiangsu Key Laboratory for the Conservation and Utilization of Biodiversity in the Middle and Lower Reaches of the Yangtze River for the daily discussion.

Author contributions

G.Y, W.R and S.X conceptualized and supervised the study. Z.Z performed all sample collection, phenotype and data analysis, RNA extraction, RT-PCR validation, and prepared the original draft. Y.L, D.X, and J.L were responsible for mouse husbandry and genotype identification. N.L, Z.Y, and L.D were responsible for the collection and preliminary analysis of enhancer HLEA. G.Y revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was financially supported by the Key Project of the National Natural Science Foundation of China (grant no. 32030011), the National Key Research and Development Program, Ministry of Science and Technology (grant no. 2022YFF1301600), and the National Natural Science Foundation of China (grant nos. 32070409, 32270453), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the Qinglan Project of Jiangsu Province, and the PI Project of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (grant no. GML2021GD0805).

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

All animal experiments in this manuscript were conducted according to the Animal Care and Use Committee of Nanjing Normal University (ethics approval number: IACUC-20200501).

Consent for publication

Not applicable.

Competing interest

The authors declare no competing interests.

Received: 18 September 2024 Accepted: 13 March 2025 Published online: 22 March 2025

References

- Lambert O, et al. An amphibious whale from the middle eocene of peru reveals early south pacific dispersal of quadrupedal cetaceans. Curr Biol. 2019;29(8):1352-1359.e3.
- Thewissen JGM, et al. Whales originated from aquatic artiodactyls in the Eocene epoch of India. Nature. 2007;450(7173):1190–4.
- Beatty BL, et al. Transition of Eocene whales from land to sea: evidence from bone microstructure. PLoS ONE. 2015;10(2): e0118409.
- Thewissen JGM, et al. Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. Nature. 2001;413(6853):277–81.
- Gingerich PD, et al. Origin of whales from early artiodactyls: hands and feet of Eocene protocetidae from Pakistan. Science. 2001;293(5538):2239–42.
- Tickle C, Münsterberg A. Vertebrate limb development—the early stages in chick and mouse. Curr Opin Genet Dev. 2001;11(4):476–81.
- López-Ríos ZRJ, Zuniga A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. Nat Rev Genet. 2009;10(12):845–58.
- Capdevila J, Belmonte JCI. Patterning mechanisms controlling vertebrate limb development. Annu Rev Cell Dev Biol. 2001;17(1):87–132.
- Swank S, Sanger TJ, Stuart YE. (Non)Parallel developmental mechanisms in vertebrate appendage reduction and loss. Ecol Evol. 2021;11(22):15484–97.

- Duboc V, Logan MPO. Pitx1 is necessary for normal initiation of hindlimb outgrowth through regulation of Tbx4 expression and shapes hindlimb morphologies via targeted growth control. Development. 2011;138(24):5301–9.
- 11. Rodriguez-Esteban C, et al. The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity. Nature. 1999;398(6730):814–8.
- Kerstjens-Frederikse WS, et al. TBX4 mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. J Med Genet. 2013;50(8):500–6.
- Naiche LA, Papaioannou VE. Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. Development. 2003;130(12):2681–93.
- Naiche LA, Papaioannou VE. Tbx4 is not required for hindlimb identity or post-bud hindlimb outgrowth. Development. 2007;134(1):93–103.
- Roscito JG, et al. Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. Nat Commun. 2018;9(1):4737.
- Infante CR, et al. Shared enhancer activity in the limbs and phallus and functional divergence of a limb-genital cis-regulatory element in snakes. Dev Cell. 2015;35(1):107–19.
- 17. Roscito JG, et al. Recapitulating evolutionary divergence in a SingleCisregulatory element is sufficient to cause expression changes of the lens GeneTdrd7. Mol Biol Evol. 2021;38(2):380–92.
- Yuan X, et al. Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. Nat Commun. 2018;9(1):4977.
- Menke DB, Guenther C, Kingsley DM. Dual hindlimb control elements in theTbx4gene and region-specific control of bone size in vertebrate limbs. Development. 2008;135(15):2543–53.
- Liang N, et al. Divergence of Tbx4 hindlimb enhancer HLEA underlies the hindlimb loss during cetacean evolution. Genomics. 2022;114(2): 110292.
- Arnold M, Stengel KR. Emerging insights into enhancer biology and function. Transcription. 2023;14(1–2):68–87.
- Infante CR, et al. Pitx1 broadly associates with limb enhancers and is enriched on hindlimb cis-regulatory elements. Dev Biol. 2013;374(1):234–44.
- Thewissen JGM, et al. Developmental basis for hind-limb loss in dolphins and origin of the cetacean bodyplan. Proc Natl Acad Sci. 2006;103(22):8414–8.
- Li J, et al. Accelerated evolution of limb-related gene Hoxd11 in the common ancestor of cetaceans and ruminants (cetruminantia). Genes Genomes Genetics. 2020;10(2):515–24.
- Wang Z, et al. Adaptive evolution of 5'HoxD genes in the origin and diversification of the cetacean flipper. Mol Biol Evol. 2008;26(3):613–22.
- Gottschalk A, et al. HOXD13-associated synpolydactyly: Extending and validating the genotypic and phenotypic spectrum with 38 new and 49 published families. Genet Med. 2023;25(11): 100928.
- Xu F, et al. HOXD13 suppresses prostate cancer metastasis and BMP4induced epithelial-mesenchymal transition by inhibiting SMAD1. Int J Cancer. 2021;148(12):3060–70.
- Liu X, et al. A single-nucleotide mutation within the TBX3 enhancer increased body size in Chinese horses. Curr Biol. 2022;32(2):480-487.e6.
- 29. Peng C, et al. Large-scale snake genome analyses provide insights into vertebrate development. Cell. 2023;186(14):2959-2976.e22.
- Wu W, et al. Genomic adaptations for arboreal locomotion in Asian flying treefrogs. Proc Natl Acad Sci. 2022;119(13): e2116342119.
- Leal F, Cohn MJ. Loss and re-emergence of legs in snakes by modular evolution of sonic hedgehog and HOXD enhancers. Curr Biol. 2016;6(21):2966–73.
- Kvon EZ, et al. Progressive loss of function in a limb enhancer during snake evolution. Cell. 2016;167(3):633-642.e11.
- Wang Z, et al. Developmental regulation of conserved non-coding element evolution provides insights into limb loss in squamates. Sci China Life Sci. 2023;66(10):2399–414.
- Roscito JG, et al. Convergent and lineage-specific genomic differences in limb regulatory elements in limbless reptile lineages. Cell Rep. 2022;38(3): 110280.
- Sun L, et al. Evolutionary genetics of flipper forelimb and hindlimb loss from limb development-related genes in cetaceans. BMC Genomics. 2022;23(1):797.

- Royle SR, Tabin CJ, Young JJ. Limb positioning and initiation: an evolutionary context of pattern and formation. Dev Dyn. 2021;250(9):1264–79.
- Neufeld SJ, Wang F, Cobb J. Genetic interactions between Shox2 and Hox genes during the regional growth and development of the mouse limb. Genetics. 2014;198(3):1117–26.
- 38. Khan S, et al. Brachdactyly instigated as a result of mutation in GDF5 and NOG genes in Pakistani population. Pak J Med Sci. 2018;34(1):82.
- Kvon EZ, et al. Enhancer redundancy in development and disease. Nat Rev Genet. 2021;22(5):324–36.
- Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci. 2010;107(50):21931–6.
- Al-Ghadban S, et al. Dilated blood and lymphatic microvessels, angiogenesis, increased macrophages, and adipocyte hypertrophy in lipedema thigh skin and fat tissue. J Obes. 2019;2019:1–10.
- 42. Booth M, et al. An automated technique for double staining of bone and cartilage in fetal mouse skeletal specimens using alizarin red S and Alcian blue. Biotech Histochem. 2021;97(3):222–7.
- Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):1–13.
- Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
- 45. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.
- Conesa A, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.
- Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33–6.
- Kang Y, et al. Histone H3K4me1 and H3K27ac play roles in nucleosome eviction and eRNA transcription, respectively, at enhancers. FASEB J. 2021;35(8): e21781.
- Kaya-Okur HS, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun. 2019;10(1):1930.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
- 51. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
- Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
- Eresen A, et al. Sorafenib plus memory like natural killer cell combination therapy in hepatocellular carcinoma. Am J Cancer Res. 2024;14(1):344–54.
- 54. <Approaches-to-modeling-of-biological-experimental-data-with-Graphpad-prism-software.pdf>. Volume 13; 2018.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):1–21.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.