***Title:*** Developmental hourglass and heterochronic shifts in fin and limb development

**Authors:** Koh Onimaru*[1,2], Kaori Tatsumi[1], Chiharu Tanegashima[1], Mitsutaka Kadota[1], Osamu Nishimura[1], Shigehiro Kuraku*[1]

**Affiliations:**

1. Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research (BDR), 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, 650-0047, Japan

2. Laboratory for Bioinformatics Research, RIKEN BDR, 1-2 Hirosawa, Wako City, Saitama, 351-0198, Japan

***Corresponding authors. Email: koh.onimaru@riken.jp, shigehiro.kuraku@riken.jp**

**Abstract**

How genetic changes are linked to morphological novelties and developmental constraints remains elusive. Here we investigate genetic apparatuses that distinguish fish fins from tetrapod limbs by analyzing transcriptomes and open chromatin regions (OCRs). Specifically, we compared mouse forelimb buds with the pectoral fin buds of an elasmobranch, the brown-banded bamboo shark (*Chiloscyllium punctatum*). A transcriptomic comparison with an accurate orthology map revealed both a mass heterochrony and hourglass-shaped conservation of gene expression between fins and limbs. Furthermore, open-chromatin analysis suggested that access to conserved regulatory sequences is transiently increased during mid-stage limb development. During this stage, stage-specific and tissue-specific OCRs were also enriched. Together, early and late stages of fin/limb development are more permissive to mutations than middle stages, which may have contributed to major morphological changes during the fin-to-limb evolution. We hypothesize that

the middle stages are constrained by regulatory complexity that results from dynamic and tissue-specific transcriptional controls.

## MAIN TEXT

### Introduction

The genetic mechanism of morphological diversity among multicellular organisms is of central interest in evolutionary biology. In particular, our understanding of how morphological novelties are linked to the emergence of their respective genetic apparatuses is limited (Rebeiz and Tsiantis, 2017). In addition, it is still unclear to what extent internal constraints, such as pleiotropy, affect evolvability (Wagner and Zhang, 2011). The fin-to-limb transition is a classic, yet still influential, case study that contributes to our understanding of morphological evolution. In general, tetrapod limbs are composed of three modules, the stylopod, zeugopod, and autopod, which are ordered proximally to distally (Figure 1A). In contrast, fish fins are often subdivided into different anatomical modules along the anterior–posterior axis—the propterygium, mesopterygium, and metapterygium (Figure 1A). Although it is still controversial how this different skeletal arrangement compares with the archetypal tetrapod limb, the autopod (wrist and digits) seems to be the most apparent morphological novelty during the fin-to-limb transition (Clack, 2009). Despite intensive comparative studies of developmental gene regulation, genetic machinery that differs between fins and limbs remains elusive. Instead, several studies revealed that autopod-specific regulation of *Hoxa13* and *Hoxd10–13*, which control autopod formation, is also conserved in non-tetrapod vertebrates (Davis et al., 2007; Freitas et al., 2007; Schneider et al., 2011), except that the expression domains of *Hoxa13* and *Hoxa11* are mutually exclusive in mouse and chick limbs while overlapping in examined fish fin buds (note that axolotl limbs also exhibit such fish-like overlap of these expression domains; Ahn and Ho, 2008; Metscher et al., 2005; Sakamoto et al., 2009; Woltering et al., 2019). Whereas several gene regulatory differences have been proposed to explain the anatomical difference between fins and limbs, these proposals have been exclusively focused on *Hox* genes (Kherdjemil et al., 2016; Nakamura et al., 2016; Sheth et al., 2012; Woltering et al., 2014). Therefore, a

genome-wide systematic study is required to identify the genetic differences between fish fins and tetrapod limbs.

There have been several difficulties that limit genetic comparisons between tetrapods and non-tetrapod vertebrates. For example, whereas zebrafish and medaka are ideal models for molecular studies, their rapid evolutionary speed and a teleost-specific whole-genome duplication hinder comparative analyses with tetrapods at both the morphological and genetic levels (Ravi and Venkatesh, 2008). This obstacle can be circumvented by using more slowly evolving species such as spotted gar, coelacanths, and elephantfish (also known as elephant shark, a cartilaginous fish that is not a true shark) with their genome sequences that have not experienced recent lineage-specific genome duplications and thus facilitate the tracing of the evolution of gene regulation (Amemiya et al., 2013; Braasch et al., 2016). However, the major disadvantage of these slowly evolving species is the inaccessibility of developing embryos. In contrast, although the eggs of sharks and rays (other slowly evolving species; Hara et al., 2018) are often more accessible, their genomic sequence information has not been available until recently. As a solution for these problems, this study used embryos of the brownbanded bamboo shark (referred to hereafter as the bamboo shark), because a usable genome assembly was recently published for this species (Hara et al., 2018). Importantly, its non-coding sequences seem to be more comparable with those of tetrapods than with teleosts (Hara et al., 2018). In addition, this species is common in aquariums and has a detailed developmental staging table, providing an opportunity to study embryogenesis (Onimaru et al., 2018). These unique circumstances of the bamboo shark enabled a comprehensive study to identify the genetic differences between fins and limbs.

In this study, to identify genetic differences between fins and limbs, we performed RNA sequencing (RNA-seq) analyses of developing bamboo shark fins and mouse limbs. Along with this transcriptomic comparison, we also generated an accurate orthology map between the bamboo shark and mouse. In addition, we applied an assay for transposase-accessible chromatin with high-throughput and chromatin accessibility analysis (ATAC-seq; Buenrostro et al., 2013) across a time series of mouse limb buds, which generated a high-quality data set the showing dynamics of open chromatin regions (OCRs; putative enhancers) during limb development. We also analyzed the evolutionary conservation of sequences in these OCRs to gain insights into the gene regulatory changes during the fin-to-limb transition.

## Results

### Comparative transcriptome analysis

To compare the temporal dynamics of gene expression between bamboo shark fin and mouse limb development, we obtained RNA-seq data from a time series of growing fin and limb buds with three replicates (Figure 1B; Supplementary file 1 for the details of RNA-seq). We selected limb buds from embryonic day (E)9.5 to E12.5 mice because this is the period during which the major segments of the tetrapod limb—the stylopod, zeugopod, and autopod—become apparent. In particular, the presumptive autopod domain, which is a distinct structure in the tetrapod limb, is visually recognizable from E11.5. For the bamboo shark, we selected developing fin stages from as wide a time period as possible (Figure 1B). To perform fine-scale molecular-level comparison, we annotated its coding genes using BLASTP against several vertebrates (listed in the Materials and Methods) and our custom algorithm. As a result, 16443 unique genes from 63898 redundant coding transcripts were annotated as orthologous to known genes of vertebrates, among which 13005 genes were uniquely orthologous to mouse genes (Table 1 for details of the transcriptome assembly; Figure 1–figure supplement 1–3, Supplementary files 2 and 3 for gene annotations and Supplementary data for sequence information). The number of detected orthologs is reasonable when compared with other studies (e.g., Hao et al., 2020). The quality of the ortholog assignment, which was assessed by examining *Hox* and *Fgf* genes, showed that our custom algorithm is more accurate than other methods (Figure 1C; see Materials and Methods and Supplementary file 4 for details). Using this assembly for the bamboo shark and RefSeq genes for mice, the means and standard errors of the transcripts per million (TPM) values were calculated from three replicates (see Figure 1–figure supplement 4 for other normalization methods and Supplementary files 5 and 6 for the full list of TPM values). In addition, for most of the analyses, TPMs were scaled by setting the highest TPM in each gene of each species to ‘1’ (which we refer to as the Max 1 method) to capture temporal dynamics rather than absolute transcript amounts. Compared to using intact TPMs and other scaling methods, Max 1 is relatively sensitive to interspecific

differences in dynamically regulated gene expression (see Methods and Figure 1–figure supplement 5 and 6 for details).

With this transcriptome data set and gene annotation, we first validated our data by analyzing the expression profiles of *Hoxa* and *Hoxd* genes. In mouse limb development, *Hoxa* and *Hoxd* genes undergo two phases of global regulation (Deschamps and Duboule, 2017). During the first phase, *Hoxd* genes are regulated by an enhancer group located 3′ of the entire HoxD cluster, and the *Hoxd* genes are sequentially upregulated from 3′ to 5′. The outcome of this first phase helps to establish the arm and the forearm. During the second phase, enhancers located 5′ of the HoxD cluster start to activate expression of *Hoxd10* to *Hoxd13* in the presumptive autopod region (Hoxa genes are regulated in a similar manner; Deschamps and Duboule, 2017). As expected, we detected the two phases of *Hoxd* gene regulation in mouse limb transcriptomes; the expression levels of *Hoxd1* to *Hoxd8* were highest at E9.5 (the first phase regulation), and *Hoxd11* to *Hoxd13* were gradually upregulated later (the second phase regulation; Figure 1D). Interestingly, the expression levels of *Hoxd9* and *Hoxd10* were highest at E10.5, which probably represents the transitional state between the first and second global regulation (Andrey et al., 2013). A similar profile was observed for *Hoxa* genes (Figure 1D). As with mouse limb buds, we found similar phasic regulation of *Hoxa* and *Hoxd* genes in the bamboo shark fin transcriptome (Figure 1D), suggesting that these transcriptomic data cover comparable developmental stages between the two species at least with respect to *Hox* gene regulation.

The overall similarity in the temporal dynamics of *Hox* gene expression between the mouse limb bud and the bamboo shark fin bud is an expected result because the second phase of *Hoxd* gene regulation has been found to be conserved in the fins of many fish (Ahn and Ho, 2008; Davis et al., 2007; Freitas et al., 2007; Schneider et al., 2011; Tulenko et al., 2017). However, there are several differences that are worth noting. For example, in mouse limb buds, *Hoxd11* and *Hoxd12* expression was highest at E11.5, followed by further upregulation of *Hoxd13* at E12.5 (Figure 1D). In contrast, in bamboo shark fin buds, these three genes reached their peak expression simultaneously at [stage (st)]31 (Figure 1D). This led us to investigate further whether the quantitative collinearity of 5′ *Hoxd* genes, where the expression of *Hoxd13* is much higher than that of its neighboring *Hoxd* genes, whose transcription levels decrease with increasing distance from *Hoxd13* (Montavon et al., 2008), is conserved in the bamboo shark fin buds. First, as a confirmation of

the previous observation, we also found quantitative collinearity of *Hoxd* genes in our transcriptome data of mouse limb buds at E12.5 (Figure 1–figure supplement 7). However, the bamboo shark fin buds exhibited no clear relationship between the genomic loci and the expression levels of *Hoxd* genes at either st31 or st32 (Figure 1–figure supplement 7): *Hoxd12* expression was highest among its neighbors. *Hoxd9* showed the second highest expression, followed by *Hoxd10* and *Hoxd11*, which had roughly identical levels of transcripts. *Hoxd13* expression was lowest among these 5′ members. Given that quantitative collinearity is considered to be a consequence of the characteristic global regulation of the HoxD cluster in the mouse limb bud (Montavon et al., 2008), this result suggests that the bamboo shark fin bud may have a different mechanism for *Hoxd* gene regulation. Interestingly, a recent study also showed that the presumptive autopod domains of chick limb buds express nearly a same amount of *Hoxd13* and *Hoxd12* transcripts (Yakushiji-Kaminatsui et al., 2018), suggesting that quantitative collinearity is not a universal feature of fins and limbs, rather varies among species. Taken together, although the overall temporal dynamics of *Hox* gene expression are conserved between the mouse limb bud and the bamboo shark fin bud, some differences in the regulation of *Hox* genes may exist between species.

To investigate to what extent our bulk transcriptome data captured the processes of cellular differentiation, we also analyzed genes related to chrondrogenesis and myogenesis. As a result, we found that the chondrogenic pathway was at least partially conserved between bambooshark fin buds and mouse limb buds; the expression level of *Sox9* and *Runx3* (key transcription factors of chondrogenesis; Fowler and Larsson, 2020) increased relatively early, and that of *Acan* (a cartilage-specific proteoglycan; Fowler and Larsson, 2020) was upregulated later (Figure 1–figure supplement 8). In contrast, although *Nog* is known to be expressed in cartilaginous condensations in mouse limb buds (Brunet et al., 1998), we did not detect a *Nog* ortholog in either the fin transcriptome or the genome assembly of the bamboo shark. As for myogenesis, our transcriptome data captured both conserved and divergent myogenetic regulation: *Pax3* (a marker of myogenic precursor cells) was downregulated over developmental time, and the MyoD gene family (*Myog*, *Myod1*, *Myf5*) took turns for further differentiation (Chal and Pourquié, 2017). In contrast, whereas mouse limb buds showed upregulation of three myosin genes (*Myh3*, *Myh7*, *Myh8*) at E12.5, we detected the upregulation of only *Myh7* in bamboo shark fin buds. Again, we did not find *Myh3* and *Myh8* in

either the transcriptome or the genome assembly of the bamboo shark. These results suggest that our transcriptome data, even though based on bulk sampling of RNA, can reveal conserved and diverged cellular differentiation processes.

**Heterochronic gene expressions**

Next, to find differences in gene regulation between the two species, we performed a gene-by-gene comparison of expression dynamics with hierarchical clustering (Figure 2A). To find potential candidate genes that contribute to the different morphologies between fins and limbs, we annotated genes with mouse mutant phenotypes (see Supplementary file 7 for the full list of genes, expression data, and annotation). The result showed that 6701 genes were significantly expressed in only one of these species ("Fin-specific" and "Limb-specific" in Figure 2A; 3284 and 3417 genes, respectively). While the fin-specific gene group consisted of many uncharacterized genes, it included ones that are known to control only fish fin development (Fischer et al., 2003; Zhang et al., 2010), such as *And1* (TRINITY_DN62789_c1_g1_i3 in Supplementary data; ortholog of a coelacanth gene, XP_015216565) and *Fgf24* (TRINITY_DN92536_c7_g1_i2 in Supplementary data; ortholog of a coelacanth gene, XP_006012032). In the limb-specific gene group, several interesting genes were listed that exhibit abnormal phenotype in the mouse limb (e.g., *Bmp2, Ihh* and *Megf8*). However, the number of these species-specific genes is probably unreliable and overestimated because these groups also contain genes for which their orthology was not assigned correctly. We also detected 1884 genes that were upregulated during late stages of fin/limb development for both species, including genes that are well known to be expressed later during fin/limb development, such as the autopod-related transcription factors *Hoxd13* and *Hoxa13* and differentiation markers *Col2a1* and *Mef2c* ("Conserved, late1 and Conserved, late2" in Figure 2A). Intriguingly, 5388 genes exhibited heterochronic expression profiles; their expression levels were highest during the late stages of mouse limb bud development but were relatively stable expression throughout fin development ("Heterochronic1"; 3178 genes) or decreased during the late stages of fin development ("Heterochronic2"; 2223 genes; see Supplementary file 7 for the full list of genes and annotations). For validation, we examined the spatio-temporal expression pattern of three heterochronic genes that exhibit limb abnormality in mouse

mutants, *Aldh1a2* from Heterochronic1 and, *Hand2* and *Vcan* from Heterochronic2. *Aldh1a2* is upregulated in the interdigital web of mouse limb buds from E11.5 (Figure 2–figure supplement 1A) and known to positively regulate interdigital cell death (Kuss et al., 2009). On the other hand, in bamboo shark fin buds, *Aldh1a2* expression was initially uniform and was later restricted to the distal edge of fin buds (Figure 2–figure supplement 1A). *Hand2* and *Vcan* transcripts were upregulated in mouse forelimb buds at E12.5 and downregulated in bamboo shark fin buds at st32 (Figure 2B, C). Thus, the temporal transcriptomic profiles were consistent with spatial expression patterns.

For a comparison, we found relatively few genes that were downregulated over time in the mouse limb bud but were upregulated in the shark fin. There was a total of 241 such genes, but only 43 of them displayed a clear heterochrony (yellow empty box in Figure 2–figure supplement 1B and Supplementary file 8 for the list of the genes). Of those, *Fgf8* is particularly interesting as FGF8 plays a crucial role as a growth signal from the apical ectodermal ridge (AER) in mouse and chick limb buds (Lewandoski et al., 2000). As shown in Figure 2-figure supplement 1C, *Fgf8* expression was high during the early stages of limb buds and was gradually downregulated at later stages. In contrast, in bamboo shark fin buds, *Fgf8* was expressed very weakly (around 0.1 TPM) at st. 27 and st. 27.5 and was upregulated at later stages. Indeed, this late upregulation of *Fgf8* was also reported in the apical fin fold (roughly equivalent to the AER) of zebrafish pectoral fin buds (Nomura et al., 2006). In the zebrafish pectoral fin bud, *Fgf16* and *Fgf24* are upregulated earlier than *Fgf8* (Draper et al., 2003; Nomura et al., 2006). In addition, *Fgf4*, *Fgf9*, and *Fgf17* are expressed in the AER and have a redundant function in the mouse limb bud (Mariani et al., 2008). Therefore, we also examined these other *Fgf* genes and found that moderate expression of *Fgf9*, *Fgf16*, and *Fgf24* were detected in the early stages of bamboo shark fin buds (Figure 2–figure supplement 1C). Although we cannot infer the ancestral state of the expression pattern, the overlapping functions of these genes may have allowed subfunctionalization of the signaling molecules of the AER during vertebrate divergence. In sum, we detected mass heterochronic shifts in gene expression between bamboo shark fin buds and mouse forelimb buds. In particular, a mechanism to maintain upregulation of the expression of genes involved in early fin development may have been either gained in the tetrapod lineages or lost in the cartilaginous fish lineages.

**Comparison of SHH signaling pathways in limb and fin buds**

In tetrapod limbs, SHH controls growth and asymmetric gene expression along the anterior-posterior axis. Although previous studies have repeatedly implied a relatively delayed onset of *Shh* expression or a short signal duration in developing fins of several elasmobranch species (Dahn et al., 2007; Sakamoto et al., 2009; Yonei-Tamura et al., 2008), there has not been solid evidence to support such a delay due to the lack of systematic gene expression analysis and the poor staging system of these species. Because the heterochronic genes identified above include basic SHH target genes, such as *Ptch1* and *Gli1*, we reexamined the expression dynamics of *Shh* and its target genes in mouse limb and bamboo shark fin buds. Because HOX genes are the upstream factors relative to *Shh* transcription (Zeller et al., 2009), we used them as a potential reference for developmental time. We first found that *Shh* transcription was present by the earliest stages examined in both bamboo shark fin and mouse limb buds, and it peaked when the transcription level of *Hoxd9* and *Hoxd10* was highest, suggesting that there was no apparent heterochrony in *Shh* transcription timing at least between these two species (red rectangles in Figure 3A and B). In contrast, SHH target genes, such as *Ptch1/2*, *Gli1*, *Gremlin* and *Hand2 (Vokes et al., 2008)*, did show a relatively extended period of expression in mouse limb buds as compared with their expression in bamboo shark fin buds. Namely, whereas the expression peak of SHH target genes was concurrent with that of *Shh* in the bamboo shark fin bud, these SHH target genes were highly expressed in E11.5 limb buds, which is one day later than the *Shh* expression peak (yellow rectangles in Figure 3A and B; see Figure 3–figure supplement 1 for intact TPM values). This timing difference is also apparent when comparing the expression peak of *Hoxd11* and *Hoxd12*, which was concurrent with that of SHH target genes in mouse limb buds, but came after downregulation of SHH target genes in bamboo shark fin buds (green rectangles in Figure 3A and B). To confirm this observation, we performed whole-mount *in situ* hybridization for *Ptch1* and *Hoxd12* in mouse limb buds and bamboo shark fin buds. As previously reported (Lewis et al., 2001; Zákány et al., 2004), mouse limb buds showed a clear expansion of the expression domain of *Ptch1* (upper panel in Figure 3C) from E10.5 to E11.5, which is accompanied by the anterior extension of the *Hoxd12* expression domain (black arrowheads in Figure 3C). In contrast, *Ptch1* was expressed in the posterior domain of bamboo shark fin buds at st. 29 (white arrowheads in Figure 3D), but was substantially downregulated by st. 31, whereas the *Hoxd12*

expression domain extended anteriorly at this stage (black arrowheads in Figure 3D). These results were roughly consistent with the RNA-seq data. We cannot completely reject the possibility that this timing difference is due to the different physical time-resolution of data sampling between these species (six time points over 20 days in the bamboo shark and four time points over 4 days in the mouse). However, given that this data set captured the similar expression dynamics of HoxA/D clusters between these species (Figure 1D; also see Figure 4C) as well as the differentiation dynamics of myocytes and chondrocytes (Figure 1–figure supplement 8), these results quite likely represent an interesting difference in the transcriptional regulation of SHH downstream genes between fins and limbs.

**Hourglass-shaped conservation**

Several studies have reported a temporally heterogeneous diversification of embryonic transcriptomes, such that the middle stages are more conserved than early or late stages (e.g., Irie and Kuratani, 2011; Kalinka et al., 2010; Levin et al., 2012). These observations are considered to support the notion of the developmental hourglass (or egg timer), which has been proposed to explain the morphological similarity of mid-stage embryos based on developmental constraints, such as strong interactions between tissues or Hox-dependent organization of the body axis (Duboule, 1994; Raff RA, 1996). In addition, a previous transcriptomic analysis reported that the late stage of mammalian limb development has experienced relatively rapid evolution (Maier et al., 2017). To examine which developmental stages of fins and limbs are conserved, we calculated the distance between the fin and limb transcriptome data. As a result, four different distance methods that we examined consistently indicated that the limb bud at E10.5 and the fin buds at st27.5–30 tended to have a relatively similar expression profile (Figure 4A for a Euclidean distance measure and Figure 4–figure supplement 1 for other types of distance measures). In addition, the transcriptomic profile of all the stages of examined fin buds showed the highest similarity to that of E10.5 limb bud (Figure 4B). Therefore, the mid-stages of limb and fin buds tend to be conserved over 400 million years of evolution.

To find factors that underlie the mid-stage conservation, we analyzed *Hox* genes, which were proposed to be responsible for the developmental hourglass (Duboule, 1994). We found that the comparison

of only *Hox* gene expression did not reproduce the hourglass-shaped conservation (Figure 4C), suggesting that other mechanisms constrain the middle stage of development. We further performed principal component analysis (PCA) of gene expression profiles to identify genes responsible for the hourglass-shaped conservation. The first component, PC1, distinguished transcriptome data mostly by species differences (Figure 4D). In contrast, PC2 was correlated with the temporal order of mouse limb buds (Figure 4D). PC2 was also weakly correlated with the temporal order of bamboo shark fin buds except at st27 (Figure 4D), but PC3 showed a clearer correlation (Figure 4E). These three components were mostly sufficient to reproduce the mid-stage conservation in Figure 4A (Figure 4–figure supplement 2A for the ratio of explained variables and 2B for the Euclidean distance measure). Interestingly, the plot with PC2 and PC3 roughly mirrored the hourglass-shaped conservation because the earliest and latest stages were placed more distantly than the middle stages in this representation (Figure 4E). Indeed, the major loadings of PC2 consisted of the conserved late expressed genes (C8) and the heterochronically regulated genes (C9 and C12) identified in Figure 2A (see Table 2 for the top 25 genes of PC2). Similarly, PC3 consisted of the conserved early genes (a part of C15) and the heterochronically regulated genes (C12 and C13; see Supplementary file 9 for the loadings of PC3 and others), suggesting that the presence of heterochronically regulated genes may at least partly contribute to the mid-stage conservation and the distant relationship between the early/late stages of fins and limbs. These results indicate that the mass heterochronic shift in gene expression, at least in part, contributes to the long distances between early- and late-stage expression profiles (Figure 4E).

Because a recent report suggests that pleiotropy of genes is related to hourglass-shaped conservation (Hu et al., 2017), we counted the number of genes with stage- or tissue-specific expression. Consistent with the previous report (Hu et al., 2017), we detected a relatively low number of stage-associated genes during the middle stages of mouse forelimb and bamboo shark fin development (Figure 4–figure supplement 2C). To evaluate the tissue specificity of genes, we first calculated Shannon entropy of gene expression patterns by analyzing RNA-seq data from 71 mouse tissues as released by the ENCODE project (Davis et al., 2018; Supplementary file 10 for the list of RNA-seq data). Namely, genes expressed only in a few tissues score lower with respect to entropy (thus, these genes are more specific). We counted genes with $1.0 \geq$ TPM and $0.65 \leq$ entropy and, again, found that the number of tissue-associated genes was relatively low at E10.5

(Figure 3F). Together, these results indicate an inverse correlation between the hourglass-shaped conservation and the number of tissue- and stage-specific genes.

**Open chromatin region (OCR) conservation**

Next, we systematically identified putative gene regulatory sequences involved in mouse limb development and sought a possible cause for the hourglass-shaped conservation in gene regulatory sequences. To this end, we applied ATAC-seq, which detects OCRs (putative active regulatory sequences), to time-series of forelimb buds at E9.5–E12.5 with three replicates. First, as a positive control, we found that ATAC-seq peaks that were determined by MACS2 peak caller covered 10 of 11 known limb enhancers of the HoxA cluster (Figure 5A and Figure 5–figure supplement 1), suggesting a high coverage of true regulatory sequences. Consistently, our ATAC-seq data showed relatively high scores for a quality control index, fraction of reads in peaks (FRiP), as compared with data downloaded from the ENCODE project (Davis et al., 2018; Figure 5B). Next, to examine evolutionary conservation, we performed BLASTN (Camacho et al., 2009) for the sequences in the ATAC-seq peaks against several vertebrate genomes. Reinforcing the result of the transcriptome analysis, we found that evolutionarily conserved sequences were most accessible at E10.5 (Figure 5C). To confirm this result, we also used a different alignment algorithm, LAST (Kiełbasa et al., 2011) with the bamboo shark and the alligator (Green et al., 2014) genomes. Alignment results for both analyses consistently indicated that the OCRs of E10.5 forelimb bud more frequently contained conserved sequences relative to those of other time points (Figure 5D; see Figure 5–figure supplement 2A and B for the absolute counts of conserved sequences). Therefore, activation of conserved gene regulatory sequences may be one of the proximate causes for the hourglass-shaped conservation of fin and limb transcriptome data.

**Temporal dynamics of open chromatin domains**

To further characterize the ATAC-seq peaks, we next performed a clustering analysis. Using one of the three replicates for each stage, we collected the summits of peaks and the surrounding 1400 bp and carried out

hierarchical clustering, which resulted in eight clusters (C1–C8; Figure 6A) that consisted of broad (C1 and C2), E11.5/E12.5-specific (C3 and C4), stable (C5 and C6), E10.5-specific (C7), and E9.5-specific (C8) peaks. The overall clustering pattern was reproducible by other combinations of replicates if its FRiP was ≥ 0.20 (Figure 6–figure supplement 1). Consistent with the above conservation analysis, E10.5-specific peaks frequently overlapped conserved sequences (Figure 5–figure supplement 2C and D).

To characterize the regulatory features of the clusters, we performed motif analysis in each cluster using HOMER (Heinz et al., 2010). First, it was convincing that stable peaks (C5 and C6) were enriched for the CTCF binding motif both in *de novo* motif discovery (Figure 6A) and known motif enrichment analysis (Figure 6-figure supplement 3), which is a major regulator of three-dimensional genomic structure. This result was consistent whether random genomic regions or other peak regions were used for the background (Figure 6-figure supplement 4). In addition, E11.5/E12.5-specific peak C3 was enriched for the HOX13 motif (Figure 6A), which was consistent with the increase in the expression of 5′ *Hox* genes (Figure 1D). C4 was also enriched for motifs similar to those of C3, but the HOX13 motif was detected only in known motif enrichment analysis (compare Figure 5-figure supplement 2 and 3). The enrichment of the HOX9 motif in E10.5-specific peaks (C7) was also consistent with our RNA-seq data, in which *Hoxd9* and *Hoxa9* expression levels peaked at E10.5 (Figure 1D). Interestingly, in E10.5-specific peaks (C7), the LHX1 binding motif was ranked at the top of the motif enrichment list (the closely related transcription factors *Lhx2*, *Lhx9*, and *Lmx1b* are required to mediate a signaling feedback loop between ectoderm and mesenchyme in limb development (Tzchori et al., 2009). C8 was enriched for motifs similar to those in C7 (e.g., COUP-TFII), but the top-ranked transcription factor in the *de novo* motif discovery analysis was VSX2, which has a very similar binding sequence to the LHX motif (Figure 6-figure supplement 2). The LHX motif was top-ranked in C8 for the known motif enrichment analysis (Figure 6-figure supplement 3). For a better understanding of the dynamics of transcription factor motifs, we counted the average number of the above detected motifs within the OCRs of each stage, which revealed a transitional increase in LHX and HOX9 motifs at E10.5 and a gradual increase in the motifs detected in C3 over the developmental stages (Figure 6-figure supplement 4).

In addition, Gene Ontology (GO) analysis for the peaks in each cluster revealed that the constitutively accessible peaks (C5, C6) were closely located to genes annotated with “cellular components”

(Supplementary file 11). Interestingly, the dynamically regulated peaks (C3, C4, C7, C8) were associated with genes with “developmental process”, “multicellular organism development”, and “anatomical structure morphogenesis” (Supplementary file 11), suggesting that these dynamic OCRs regulate developmental genes. Together, these results suggest that there are E10.5-specific transient OCRs that exhibit several characteristics including their evolutionary conservation, the presence of LHX and HOX9 motifs and a close relation with developmental genes.

To confirm the results from the above clustering analysis, we also determined the genomic regions that showed a statistically significant increase or decrease in the ATAC-seq signal within a day by using all replicates. As a result, ATAC-seq signals were most increased during the transition from E9.5 to E10.5 in the mouse limb bud. From E10.5 to E11.5, the total number of decreased and increased signals was highest, indicating that the OCR landscape was most dynamically changing at E10.5 (Figure 6B). In contrast, relatively few significant changes were observed from E11.5 to E12.5. Thus, in contrast to the transcriptome analysis, stage-specific gene regulatory sequences are likely to be most accessible at E10.5. Moreover, by comparing the peaks of each cluster identified above with ATAC-seq peaks of other cells and tissues released by the ENCODE project (Davis et al., 2018; Supplementary file 10 for the full list of cells and tissues), we discovered that the C7 cluster (E10.5-specific peaks) contained more peaks that did not overlap with those of other cells and tissues. Again, in contrast to the transcriptome analysis, the data suggest that gene regulatory sequences that are accessible only at E10.5 tend to be limb-specific (Figure 6C). Taken together, these analyses revealed a unique regulatory landscape of forelimb buds at E10.5, which is enriched for evolutionarily conserved stage-specific and tissue-specific OCRs.

## Discussion

In this work, we applied transcriptomics and chromatin accessibility analysis to systematically study genetic changes that differentiate fins from limbs. Because of the slow sequence evolution and the embryo availability of the bamboo shark, we were able to compare transcriptional regulation of genes with high accuracy and found both heterochronic shifts and hourglass-shaped conservation of transcriptional regulation

between fin and limb development. Here, we discuss the interpretations, limitations, and implications of these results.

Our time-series transcriptome data indicated that a remarkable number of genes that exhibit the highest expression during the late stages of mouse limb bud development are decreased during the late stages of bamboo shark fin development (Figure 2). The simplest hypothesis for this mass heterochronic shift is that the later stages of limb development gained expression of one or a few upstream transcription factor(s) or signaling molecules that collectively regulate this group of genes. Interestingly, we also observed relatively extensive expression of the downstream targets of the SHH signaling pathway in mouse limb buds, as compared with bamboo shark fin buds (Figure 3). Because SHH-independent regulation of its target genes through the GLI3-HOX complex was previously reported (Chen et al., 2004), the mismatch between the peak expression of *Shh* and its target genes may be caused by such SHH-independent regulatory mechanisms that are absent in bamboo shark fin development. Given that direct and genetic interactions of GLI3 and HOX have a significant impact on autopod formation, the emergence of this interaction may be a key component of the mass heterochronic shift and the acquisition of autopod-related developmental regulation in the tetrapod lineages. However, because we compared only two species, it is equally possible that the late stages of shark fin development lost this SHH-independent gene regulation. Alternatively, given that the evolutionary distance between these two species is >400 million years, it is also possible that every one of these genes independently shifted their expression to the later stages of limb development or to the early stages of shark fin development. Further taxon sampling and functional analyses will reveal the relation between the mass heterochronic shift and the emergence of the autopod.

Related to the potential changes in regulation of SHH target genes, by analyzing catshark fin buds, we previously proposed that the expression domains of genes that are positively regulated by SHH might have expanded anteriorly during the fin-to-limb transition (Onimaru et al., 2015). We speculated that this expression changes may be linked to the loss of pro- and mesopterygial elements. Recently, this hypothesis was partially supported by another group who compared the gene expression pattern of lungfish and cichlid fin development (Woltering et al., 2020), where lungfish fin buds seem to exhibit an intermediate condition between non-sarcopterygian fish fins and tetrapod limbs in terms of gene expression distribution along the

anterior-posterior axis. This group particularly emphasized that the absence and presence of the dynamics of the anterior expansion of *Hoxd13* expression correlate with the difference between the metapterygial morphologies of lungfish and tetrapods (also see Johanson et al., 2007 for a conflicting report). However, the significance of changes in *Hoxd13* expression remains unclear because of the following two reasons: a) *Hoxd13* expression pattern seems to quite vary among species—the anterior expansion of *Hoxd13* expression has been observed in the fin buds of the little skate, the small-spotted catshark, and *Polyodon* (Davis et al., 2007; Freitas et al., 2007; Nakamura et al., 2015), while not in those of zebrafish and cichlids (Ahn and Ho, 2008; Woltering et al., 2020) and b) in fish fins, the expression domain of *Hoxa13,* whose function is mostly redundant with *Hoxd13,* commonly spans from anterior to posterior regions in fish fin buds like as tetrapod limbs (Davis et al., 2007; Freitas et al., 2007; Nakamura et al., 2016). Therefore, while changes in *Hoxd13* expression domain are likely to contribute to some degree of anatomical diversity, their impact is questionable in the context of the fin-to-limb transition. Nevertheless, our previous study and Woltering et al. commonly suggest that the anterior expansion of gene expression domains is likely associated with the substantial anatomical changes during the fin-to-limb transition. As discussed above, we speculate that the mass heterochronic shifts that we observed in the present study may be related to the gain of SHH-independent regulation of its target genes. Therefore, whether the anterior expansion of SHH-target gene expression is related to the mass heterochronic shifts will be one of the interesting questions to address in the future.

We observed that gene expression profiles are most highly conserved between bamboo shark fin buds at st. 27.5–30 and mouse forelimb buds at E10.5 (Figure 4). Consistent with this result, our chromatin accessibility analysis reveals that OCRs at E10.5 tend to contain evolutionarily conserved sequences (Figure 5). Whereas transcriptomic conservation during the middle of embryonic development has been reported by many groups using different species (e.g., Irie and Kuratani, 2011; Kalinka et al., 2010), analysis of regulatory sequence conservation during embryonic development has been either incomplete or controversial. For example, by analyzing histone acetylation marks on several developing organs in mouse embryos, Nord et al. proposed regulatory sequences active at E11.5 are exposed by the highest evolutionary constraint (Nord et al., 2013). However, they used stem cell lines as the substitutes for organs at early stages.

Another study showed that genes expressed at the segmentation stage of zebrafish embryos tended to be surrounded by highly conserved non-coding sequences (Piasecka et al., 2013). Although their results are in line with our present study as discussed below, they did not show that these highly conserved non-coding sequences were indeed active at the segmentation stage. In addition to these studies, there is a conflicting observation that early, instead of middle, embryonic stages tend to be regulated by conserved OCRs (Uesaka et al., 2019). Therefore, our present study is the first to convincingly show a clear correlation of conservation status between transcriptomic data and OCRs. Our results suggest that evolutionary constraints on the gene regulatory apparatus are present during the middle stage of fin and limb development. What drives the hourglass-shaped conservation is still under debate. Interestingly, we found that stage- and tissue-specific OCRs were enriched in this conserved period, during which a relatively low number of stage- and tissue-specific genes were expressed (Figure 6). These quite contrasting observations imply that the mid-stage limb development is enriched for pleiotropic genes controlled by multiple tissue-specific enhancers, including limb-specific ones, rather than by constitutive promoters that often regulate housekeeping genes. Therefore, we speculate that, at least in the case of limb development, complex regulatory sequences that execute spatiotemporally specific transcriptional controls over pleiotropic genes constrain the evolvability of this particular period of morphogenesis, probably due to the vulnerability of complex regulation to genetic mutations.

In conclusion, the present study provides insights for the evolutionary origin of gene regulation that differentiates fins from limbs. In particular, comparative transcriptional analyses prompted us to hypothesize that mass heterochronic shifts of gene expression may have occurred during the fin-to-limb evolution. In addition, both transcriptome and open chromatin data point to an evolutionary constraint during mid-stage limb development, likely owing to gene regulatory complexity. Although these hypotheses require further taxon sampling and experimental tests, this study opens up new prospects for understanding not only the genetic basis of the fin-to-limb transition but also the general nature of morphological evolution.

## Materials and Methods

447 **Key resources table**

| Key Resources Table | | | | |
|---|---|---|---|---|
| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
| gene (*Mus musculus*) | Hand2 | ENSEMBL | ENSMUST00000040104.4 | N/A |
| gene (*Mus musculus*) | Vcan | ENSEMBL | ENSMUST00000109546.8 | N/A |
| gene (*Mus musculus*) | Aldh1a2 | ENSEMBL | ENSMUST00000034723.5 | N/A |
| gene (*Mus musculus*) | Ptch1 | ENSEMBL | ENSMUST00000192155.5 | N/A |
| gene (*Mus musculus*) | Hoxd12 | ENSEMBL | ENSMUST00000001878.5 | N/A |
| gene (*Chiloscyllium punctatum*) | Hand2 | ENSEMBL | Chipun0004250/g4250.t1/ TRINITY_DN85524_c0_g1_i1 | N/A |
| gene (*Chiloscyllium punctatum*) | Vcan | This paper | Chipun0003941/g3941.t1/ TRINITY_DN95522_c0_g1_i8 | N/A |
| gene (*Chiloscyllium punctatum*) | Hoxd12 | This paper | Chipun0005654/g5654.t1/TRINITY_DN85970_c0_TRINITY g1_i1 | N/A |
| gene (*Chiloscyllium punctatum*) | Ptch1 | This paper | Chipun0003320/g3320.t1/TRINITY_DN92499_c0_g1_i3 | N/A |
| gene (*Chiloscyllium punctatum*) | Aldh1a2 | This paper | Chipun0010503/g10503.t1/TRINITY_DN81423_c0_g1_i1 | N/A |
| strain, strain background ( *Mus musculus*) | C52BL/6 | Laboratory for Animal Resources and Genetic Engineering RIKEN, | N/A | N/A |
| antibody | Anti-Digoxigenin-AP, Fab fragments (Sheep) | Millipore Sigma | Cat# 11093274910 | polyclonal (1:4000) |
| sequence-based reagent | Mus musculus Hand2 forward primer | This paper | PCR primers | ACCAAACT CTCCAAGA TCAAGACA CTG |
| sequence-based reagent | Mus musculus Hand2 reverse primer | This paper | PCR primers | TTGAATACT TACAATGTT TACACCTT C |
| sequence-based reagent | Mus musculus Vcan forward | This paper | PCR primers | TGCAAAGA TGGTTTCA TTCAGCGA |

| | | | | |
|---|---|---|---|---|
| | primer | | | CAC |
| sequence-based reagent | Mus musculus Vcan reverse primer | This paper | PCR primers | ACACGTGC AGAGACCT GCAAGATG CTG |
| sequence-based reagent | Mus musculus Aldh1a2 forward primer | This paper | PCR primers | ACCGTGTT CTCCAACG TCACTGAT GAC |
| sequence-based reagent | Mus musculus Aldh1a2 reverse primer | This paper | PCR primers | TCTGTCAG TAACAGTAT GGAGAGCT TG |
| sequence-based reagent | Mus musculus Hoxd12 forward primer | This paper | PCR primers | CTCAACTT GAACATGG CAGTGCAA GTG |
| sequence-based reagent | Mus musculus Hoxd12 reverse primer | This paper | PCR primers | AGCTCTAG CTAGGCTC CTGTTTCAT GC |
| sequence-based reagent | Mus musculus Ptch1 forward primer | This paper | PCR primers | GGGAAGG CAGTTCAT TGTTACTGT AACTG |
| sequence-based reagent | Mus musculus Ptch1 reverse primer | This paper | PCR primers | TGTAATAC GACTCACT ATAGGTCA GAAGCTGC CACACACA GGCATGAA GC |
| sequence-based reagent | *Chiloscyllium punctatum* Hand2 forward primer | This paper | PCR primers | ACCAGCTA CATTGCCT ACCTCATG GAC |
| sequence-based reagent | *Chiloscyllium punctatum* Hand2 reverse primer | This paper | PCR primers | CACTTGTT GAACGGAA GTGCACAA GTC |
| sequence-based reagent | *Chiloscyllium punctatum* Vcan forward primer | This paper | PCR primers | AGCTTGGG AAGATGCA GAGAAGGA ATG |
| sequence-based reagent | *Chiloscyllium punctatum* | This paper | PCR primers | AGAGCAGC TTCACAAT |

| | | | | |
|---|---|---|---|---|
| | Vcan reverse primer | | | GCAGTCTC TGG |
| sequence-based reagent | *Chiloscyllium punctatum* Aldh1a2 forward primer | This paper | PCR primers | TTGAACTT GTACTAAG TGGTATCG CTG |
| sequence-based reagent | *Chiloscyllium punctatum* Aldh1a2 reverse primer | This paper | PCR primers | AGGATGTG AACATTAG GCTGACCT CAC |
| sequence-based reagent | *Chiloscyllium punctatum Hoxd12* forward primer | This paper | PCR primers | GCCAGTAT GCAACAGA TCCTCTGA TGG |
| sequence-based reagent | *Chiloscyllium punctatum Hoxd12* reverse primer | This paper | PCR primers | CTAATGAC CTGTTGTA CTTACATTC TC |
| sequence-based reagent | *Chiloscyllium punctatum Ptch1* forward primer | This paper | PCR primers | TTCAGCCA GATTGCAG ATTACATCA ACC |
| sequence-based reagent | *Chiloscyllium punctatum Ptch1* reverse primer | This paper | PCR primers | TTCTCTGT GTTTCACA TTCAACGT CCTG |
| commercial assay or kit | Nextera DNA Sample Preparation Kit | Illumina | Cat# FC-121-1031 | |
| commercial assay or kit | TruSeq Stranded mRNA LT Sample Prep Kit | Illumina | Cat# RS-122-2101 | |
| software, algorithm | Trinity | https://github.com/trinityrnaseq/trinityrnaseq | RRID:SCR_013048 | N/A |
| software, algorithm | Bowtie2 | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml | RRID:SCR_016368 | N/A |
| software, algorithm | BWA | http://bio-bwa.sourceforge.net/ | RRID:SCR_010910 | N/A |
| software, algorithm | MACS2 | https://github.com/macs3-project/MACS | RRID:SCR_013291 | N/A |
| software, algorithm | HOMER | http://homer.ucsd.edu/homer/motif/ | RRID:SCR_010881 | N/A |
| software, algorithm | RSEM | https://github.com/deweylab/RSEM | RRID:SCR_013027 | N/A |
| software, algorithm | scikit-learn | https://scikit-learn.org/stable/ | RRID:SCR_002577 | N/A |

**Animals**

Animal experiments were conducted in accordance with the guidelines approved by the Institutional Animal Care and Use Committee (IACUC), RIKEN Kobe Branch, and experiments involving mice were approved by IACUC (K2017-ER032). The eggs of brownbanded bamboo shark (*C. punctatum*) were kindly provided by Osaka Aquarium Kaiyukan and were incubated at 25°C in artificial seawater (MARINE ART Hi, Tomita Pharmaceutical Co., Ltd.) and staged according to the published staging table (Onimaru et al., 2018). For mouse embryos, C52BL/6 timed-pregnant females were supplied by the animal facility of Kobe RIKEN, LARGE and sacrificed at different days after 9.5−12.5 days of gestation. For RNA-seq, fin buds and limb buds were dissected in cold seawater and phosphate-buffered saline (PBS), respectively, and stored at −80°C. For *in situ* hybridization, embryos were fixed overnight in 4% paraformaldehyde in PBS, dehydrated in a graded methanol series, and stored in 100% methanol at −20°C.

**RNA-seq**

We sampled mouse forelimb buds at E9.5, E10.5, E11.5 and E12.5 and bamboo shark pectoral fin buds at st27, st27.5, st29, st30, st31 and st32 and pooled several individual samples by stage to obtain enough RNA for each time point. We considered this pooled sample to represent one biological replicate (other replicates were generated using different individuals). Total RNAs from these samples were extracted with the RNeasy Micro and Mini plus kit (QIAGEN, Cat. No. 74034 and 74134) and PicoPure RNA Isolation Kit (ThermoFisher, Cat. No. KIT0214). Genomic DNA was removed with gDNA Eliminator columns included with this kit. For quality control, the Agilent 2100 Bioanalyzer system and Agilent RNA 6000 Nano Kit (Agilent, Cat. No. 5067-1511) were used to measure the RNA integrity number for each sample. Using 237 ng of each of the RNA samples, strand-specific single-end RNA-seq libraries were prepared with the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, Cat. No. RS-122-2101 and/or RS-122-2102). For purification, we applied 1.8× (after end repair) and 1.0× (after adapter ligation and PCR) volumes of Agencourt AMPure XP (Beckman Coulter, Cat. No. A63880). The optimal number of PCR cycles for library

amplification was determined by a preliminary quantitative PCR using KAPA HiFi HotStart Real-Time Library Amplification Kit (KAPA, Cat. No. KK2702) and was estimated to be 11 cycles for mouse limb buds and 10 cycles for bamboo shark fin buds. The quality of the libraries was checked by Agilent 4200 TapeStation High Sensitivity D1000. The libraries were sequenced after on-board cluster generation for 80 cycles using 1× HiSeq Rapid SBS Kit v2 (Illumina, Cat. No. FC-402-4022) and HiSeq SR Rapid Cluster Kit v2 (Illumina, Cat. No. GD-402-4002) on a HiSeq 1500 (Illumina) operated by HiSeq Control Software v2.2.58 (Run type: SR80 bp). The output was processed with Illumina RTA v1.18.64 for base-calling and with bcl2fastq v1.8.4 for de-multiplexing. Quality control of the obtained fastq files for individual libraries was performed with FASTQC v0.11.5. RNA-seq was performed with three biological replicates for each stage.

**Transcriptome assembly and orthology assignment**

We used the NCBI RefSeq mouse proteins (GRCm38.p5; only curated proteins were used) and two bamboo shark gene lists: a genome sequence–based gene model (Hara et al., 2018) and transcripts assembled from RNA-seq in this study (see below) for orthology assignment. The amino acid sequences of the published gene model of the bamboo shark are available from https://doi.org/10.6084/m9.fig (Supplementary file 1). For the transcriptome assembly, the short reads from the bamboo shark RNA-seq data were trimmed and filtered with Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and assembled using Trinity v2.4.0 (Grabherr et al., 2011; options: --SS_lib_type RF --normalize_max_read_cov 200 --min_kmer_cov 2). Protein coding sequences were predicted with a program that finds coding regions, TransDecoder v3.0.1 (Haas et al., 2013), according to the guide in TransDecoder (Supplementary data 2 and 3). Using these coding gene lists as queries, orthologous pairs were assigned as illustrated in Figure 1–figure supplement 1. The idea behind this algorithm is the “gar bridge” (Braasch et al., 2016), an empirical observation that a comparison including intermediate and slowly evolving animals yields a better resolution for identifying homologous sequences than a direct comparison between two species. First, BLASTP v2.7.1 was performed between mouse and bamboo shark genes reciprocally, and also against the coding genes of

the elephantfish (or elephant shark; Callorhinchus_milii-6.1.3), spotted gar (LepOcu1), coelacanth (LatCha1), chicken (GRCg6a), alligator (ASM28112v4), and human (GRCh38.p12; options: -outfmt 6 -evalue 1e-30 -window 0). Then, the BLASTP results of bamboo shark queries against the animals listed above (except for the elephantfish) were concatenated, and the best hit across species (cross-species best hit) was identified for each of the bamboo shark genes. If there was no cross-species best hit, then the best hit among the elephantfish genes was retrieved, which may include cartilaginous fish–specific genes. Subsequently, orthologous pairs between mouse and bamboo shark genes were assigned by checking if a cross-species best hit from the bamboo shark BLASTP results also had a best hit in the BLASTP result of mouse genes against the corresponding animal (species-wise best hit; Supplementary files 4–6).

For quality control, the orthology of Fgf family members was independently determined by generating molecular phylogenetic trees (Figure 1–figure supplement 2 and 3). Amino acid sequences were aligned with an alignment tool, MAFFT v7.419-1 (Katoh, 2002; options: --localpair --maxiterate 1000) and trimmed with trimAL v1.2 (Capella-Gutiérrez et al., 2009; options: -gt 0.9 -cons 60). Then, maximum-likelihood trees were constructed with RaxML v8.2.12 (Stamatakis, 2014; options: -x 12345 -p 12345 -m PROTGAMMAWAG -f a -# 100). The orthology of *Hox* genes was confirmed based on genome synteny. These independently confirmed orthologous pairs were compared with the results of the above orthology assignment algorithm. For a comparison, we also used the results from a reciprocal best hit algorithm, proteinOrtho v6.0.4 (Lechner et al., 2011) and the previously generated orthology groups (Hara et al., 2018; Figure 1B).

**Quantification and scaling**

The trimmed RNA-seq short reads were aligned to the transcript contigs for the bamboo shark and curated RefSeq genes (GRCm38.p5) for the mouse using RSEM v1.3.0 (Li and Dewey, 2011) and Perl scripts (align_and_estimate_abundance.pl and abundance_estimates_to_matrix.pl) in the Trinity package. TPM (transcripts per million), but not TMM (trimmed mean of M-values), was used for all analyses, because we found some artificial biases in TMM values (see Figure 1–figure supplement 4). TPM values from the

splicing variants of a single gene were summed up to generate a single value per gene. Then, the means and standard errors of TPM values from three replicates were used for the downstream analyses. Genes with a maximum TPM < 1.0 were considered not expressed. For clustering and distance measures, TPM values were scaled so that the maximum value of each gene of each species was set to '1' (Max 1). Whereas this scaling method loses information with respect to the absolute value of the TPMs, it has a substantial advantage when comparisons are being made between evolutionarily distant species. Indeed, previous comparative transcriptome studies have scaled gene expression values in different ways. Among those approaches, the use of Z-scores (standardization) and log transformations are relatively common strategies (e.g., Kalinka et al., 2010; Leiboff and Hake, 2019; Levin et al., 2016). Some researchers have used the intact RPKM (reads per kilobase per million) values to compare closely related species (Wang et al., 2013), but, because the RPKM is known to be inconsistent between samples even within a species (Wagner et al., 2012). Scaled transcriptional values are commonly used for clustering analyses and visualization of transcriptomic data from different samples within a single species. In this case, scaling is mainly aimed at flattening the dynamic range of transcription levels among genes. For inter-specific comparisons, scaling is also useful for being simultaneously sensitive to differentially regulated genes and also insensitive to conserved housekeeping genes. Here, we examine the effect of several scaling methods and the use of intact TPM values. We define the four relevant scaling methods as follows:

$$M_{g,s,t} = \frac{x_{g,s,t}}{max\left(\{x_{g,s,t}: t = 1..T_s\}\right)}$$

$$Z_{g,s,t} = \frac{\left(x_{g,s,t} - \bar{x}_{g,s}\right)}{\sigma_{g,s}}$$

$$U_{g,s,t} = \frac{x_{g,s,t}}{\left\|\{x_{g,s,t}: t = 1..T_s\}\right\|}$$

$$L_{g,s,t} = log_{10}\left(x_{g,s,t} + 1\right)$$

where $x_{g,s,t}$ is the intact TPM of gene $g$, species $s$, and time point $t$; $T_s$ is the total number of time points in species $s$; $M_{g,s,t}$, $Z_{g,s,t}$, $U_{g,s,t}$ and $L_{g,s,t}$ are scaled values that we refer to as the Max 1, Z-score, Unit vector and

Log10 methods, respectively; and $\bar{x}_{g,s}$ and $\sigma_{g,s}$ are the mean and standard deviation, respectively, of $\{x_{g,s,1}...x_{g,s,Ts}\}$.

First, we take a simple example to develop some intuition as to how these calculations transform TPM values. Let us assume that we compare two species [(species 1 and species 2)], and each species has two genes (gene 1 and gene 2) and three developmental time points (t1, t2, and t3; Figure 1–figure supplement 5A). Gene 1 is a constitutively active gene (i.e., a housekeeping gene), and gene 2 is differentially regulated between species. In this example, we want to identify t2 as the most conserved time point because gene 2 is expressed in both species at this time point. In addition, we want to ignore the subtle expression differences of gene 1 within and between species. As seen in Figure 1–figure supplement 5A, scaling by the Max 1, Unit vector, and Log10 methods effectively conserves the expression dynamics of gene 2 while suppressing the expression noise of gene 1. In contrast, Z-score scaling amplifies the expression dynamics of both genes to the same degree, which suggests that the Z-score method is sensitive to noise. Calculation of the Euclidean distances for each time point between species 1 and 2 ("Distance" in Figure 1–figure supplement 5A) shows that although all scaling methods and the use of intact TPMs indicate that t2 is the most similar time point, Max 1 creates a greater contrast between conserved and non-conserved time points than the other methods. Therefore, Max 1 is likely to be able to sensitively detect inter-specific differences. We also examined a subset of our real transcriptomic data from mouse limb buds and bamboo shark fin buds. As an example, we chose three housekeeping genes conserved in most vertebrates, *Psmd5*, *Mrpl21*, and *Polr1b*—these genes are listed both in a housekeeping gene list https://www.tau.ac.il/~elieis/HKG/HK_genes.txt (Eisenberg and Levanon, 2013) and in the BUSCO data set, a gene list used to assess the completeness of genome assemblies (Simão et al., 2015). As shown in Figure 1–figure supplement 5B and C, the TPM values of these genes were stable throughout developmental time in both species, suggesting that these genes also play a role in the maintenance of basic cellular function in bamboo shark fin development. However, the TPM values of *Mrpl21* and *Polr1b* in mouse limb buds were roughly twice as high as those in bamboo shark fin buds. One explanation for this finding is that the expression of housekeeping genes is low in the bamboo shark because the relatively low temperature of the environment in which it lives slows its metabolic activity. We note, however, that there are many technical

uncertainties when directly interpreting TPM values, particularly when comparing distantly related species. For example, differences in DNA sequences of transcripts (such as variations in GC content) between species probably affects the efficiency of library preparation and sequencing. The TPM values are also likely to be biased because of the incompleteness of the reference transcriptome sequence that we used for the bamboo shark (e.g., some genes lack 3′ untranlated regions). Therefore, the dynamics of TPM values extracted by scaling methods rather than absolute TPM values are likely to contain more biologically relevant information. Of the scaling methods, Max 1, Unit vector, and Log10 conserved the stable expression profile of the housekeeping genes, whereas the Z-score method amplified the subtle variation in TPM values as seen in the above simple example (Figure 1–figure supplement 5B). In particular, the Max 1 and Unit vector methods transformed the TPM values into relatively comparable values between the two species (compare Figure 1–figure supplement 5B with C). For a comparison, we also examined three genes that are heterochronically regulated between bamboo shark fin buds and mouse limb buds (Figure 1–figure supplement 6A and B). In this case, all of the scaling methods seemed to conserve the temporal dynamics of gene expression.

To obtain an objective measure, we calculated the ratio of the interspecific Euclidean distance of the three housekeeping genes to that of the three heterochronic genes with different scaling methods (Figure 1–figure supplement 6C and D). Namely, the Euclidean distance of expression values was close to zero if we used only housekeeping genes (left panel of Figure 1–figure supplement 6C), but it was larger when comparing heterochronic genes (right panel of Figure 1–figure supplement 6C). As a result, the Max1 method resulted in the highest ratio (Figure 1–figure supplement 6D), suggesting that Max1 is most sensitive to interspecific differences in dynamically regulated genes.

**Clustering analyses of transcriptome data**

The scaled values of each orthologous pair were concatenated as a 10-dimensional vector (consisting of four stages for mouse limb buds and six stages for bamboo shark fin buds), and all gene expression vectors were dimensionally reduced with UMAP (hyper parameters: a = 10, b = 1.8) followed by hierarchical clustering

(hyper parameters: method = “ward”, metric = “euclidean”; the code is available at https://github.com/koonimaru/easy_heatmapper). To find genes that have an opposite trend in their expression relative to “Heterochronic2”, a Pearson correlation coefficient (PCC) for TPM values and developmental stages was calculated for each gene for each species, and genes with PCC > 0.5 for bamboo shark fin buds and PCC < −0.5 for mouse limb buds were listed (Figure 2–figure supplement 1B and Supplementary file 8). For the distance measurements, four different distance methods were calculated: Euclidean distance ($\sqrt{\sum(u_i - v_i)^2}$), correlation distance ($1 - \frac{(u-\bar{u})(v-\bar{v})}{\|(u-\bar{u})\|_2\|(v-\bar{v})\|_2}$), Shannon distance ($-\frac{1}{2}\sum u_i \, log \frac{(u_i+v_i)}{2u_i} + v_i log \frac{(u_i+v_i)}{2v_i}$), standardized Euclidean distance ($\sqrt{\sum(u_i - v_i)^2/V_i}$), where $u$ and $v$ are gene expression vectors of two samples and $V_i$ is the variance computed over all the values of gene $i$. For PCA analysis, we used the PCA module in a python package, scikit-learn (https://scikit-learn.org/stable/).

For the stage-associated gene analysis in Figure 3–figure supplement 1B and 1C, we first calculated the z-score of each gene at each stage as $\frac{(u_{k,i}-\bar{u}_i)}{\sigma_i}$, where $u_{k,i}$ is the TPM value of gene $i$ at stage $k$, $\bar{u}_i$ is a mean of TPM over all the stages, and $\sigma_i$ is the standard deviation of the TPM. Genes with TPM ≥ 1.0 and the absolute Z-score ≥1.0 were counted as stage-associated genes. For the tissue-associated gene analysis, the entropy of each gene was calculated using RNA-seq data of 71 tissues downloaded from the ENCODE web site (https://www.encodeproject.org/; see Supplementary Table 4 for all list). Entropy was calculated as follows:

$$p_{k,i} = \frac{TPM_{k,i}}{\sum_k TPM_{k,i}},$$

$$H_i = -\sum_k p_{k,i} \, log(p_{k,i}),$$

where $TPM_{k,i}$ is the TPM value of gene $i$ in tissue $k$, $p_{k,i}$ is a probability distribution and $H_i$ is entropy. Genes with TPM (of mouse limb buds) ≥ 1.0 and 0.65 ≤ entropy were counted as tissue-associated genes.

**Whole-mount *in situ* hybridization**

To clone DNA sequences for RNA probes, we used primers that were based on the nucleotide sequences in the ENSEMBL database (https://www.ensembl.org) for mouse genes and in the transcriptome assembly (Supplementary data 3): bamboo shark *Hand2* (Chipun0004250/g4250.t1/ TRINITY_DN85524_c0_g1_i1), 5′-ACCAGCTACATTGCCTACCTCATGGAC-3′ and 5′-CACTTGTTGAACGGAAGTGCACAAGTC-3′; bamboo shark *Vcan* (Chipun0003941/g3941.t1/ TRINITY_DN95522_c0_g1_i8), 5′-AGCTTGGGAAGATGCAGAGAAGGAATG-3′ and 5′-AGAGCAGCTTCACAATGCAGTCTCTGG-3′; bamboo shark *Hoxd12* (Chipun0005654/g5654.t1/TRINITY_DN85970_c0_g1_i1), 5′-GCCAGTATGCAACAGATCCTCTGATGG-3′ and 5′-CTAATGACCTGTTGTACTTACATTCTC-3′; bamboo shark *Ptch1* (Chipun0003320/g3320.t1/TRINITY_DN92499_c0_g1_i3), 5′-TTCAGCCAGATTGCAGATTACATCAACC-3′ and 5′-TTCTCTGTGTTTCACATTCAACGTCCTG-3′; bamboo shark *Aldh1a2* (Chipun0010503/g10503.t1/TRINITY_DN81423_c0_g1_i1), 5′-TTGAACTTGTACTAAGTGGTATCGCTG-3′ and 5′-AGGATGTGAACATTAGGCTGACCTCAC-3′; mouse *Hand2* (ENSMUST00000040104.4), 5′-ACCAAACTCTCCAAGATCAAGACACTG-3′ and 5′-TTGAATACTTACAATGTTTACACCTTC-3′; mouse *Vcan* (ENSMUST00000109546.8), 5′-TGCAAAGATGGTTTCATTCAGCGACAC-3′ and 5′-ACACGTGCAGAGACCTGCAAGATGCTG-3′; mouse *Hoxd12* (ENSMUST00000109546.8), 5′-TGCAAAGATGGTTTCATTCAGCGACAC-3′ and 5′-ACACGTGCAGAGACCTGCAAGATGCTG-3′; mouse *Aldh1a2* (ENSMUST00000034723.5), 5′-ACCGTGTTCTCCAACGTCACTGATGAC-3′ and 5′-TCTGTCAGTAACAGTATGGAGAGCTTG-3′; mouse *Ptch1* (ENSMUST00000192155.5), 5′-GGGAAGGCAGTTCATTGTTACTGTAACTG-3′ and 5′-TGTAATACGACTCACTATAGGTCAGAAGCTGCCACACACAGGCATGAAGC-3′. Note that although we also tried bamboo shark *Shh* expression analysis using several RNA probes, we did not obtain specific signals. Fixed embryos were processed for *in situ* hybridization as described (Westerfield, 2000) with slight modifications. Briefly, embryos were re-hydrated with 50% MeOH in PBST (0.01% Tween 20 in PBS) and with PBST for 5–30 min each at room temperature (RT). Then, embryos were treated with 20 μg/ml proteinase K (Roche) in PBST (5 sec for mouse E11.5 and E12.5 embryos, 5 min for st. 27 and st. 29 bamboo shark embryos, 10 min for st. 31 and st. 32 bamboo shark embryos). After the proteinase treatment, embryos were fixed in 4% paraformaldehyde/PBS for 1 hour, followed by one or two washes with PBST for 5–10 min each. Optionally, if embryos had some pigmentation, they were immersed in 2% $H_2O_2$ until they

became white. Then, embryos were incubated for 1 hour in preheated hybridization buffer (50 ml formaldehyde; 25 ml 20× SSC, pH 5.0; 100 μl 50 mg/ml yeast torula RNA; 100 μl 50 mg/ml heparin; 1 ml 0.5 M EDTA; 2.5 ml 10% Tween 20; 5 g dextran sulfate; and DEPC-treated MilliQ water to a final volume of 100 ml) at 68°C. Subsequently, embryos were incubated with fresh hybridization buffer containing 0.25–4 μl/ml of RNA probes at 68°C overnight. Embryos were washed twice with preheated Wash buffer 1 (50 ml formaldehyde; 25 ml 20× SSC, pH 5.0; 2.5 ml 10% Tween 20; and DEPC-treated MilliQ water to a final volume of 100 ml) for 1 hour each at 68°C; once with preheated Wash buffer 2, which consisted of equal volumes of Wash buffer 1 and 2× SSCT (10 ml 20× SSC, pH 7.0; 1 ml 10% Tween 20; and MilliQ water to a final volume of 100 ml), for 10 min at 68°C; once with preheated 2× SSCT at 68°C for 10 min; and once with TBST at room temperature for 10 min. Embryos were then incubated with a blocking buffer (20 μl/ml 10% bovine serum albumin, 20 μl/ml heat-inactivated fetal bovine serum in TBST) for 1 hour at room temperature, followed by incubation with 1/4000 anti-digoxigenin (Roche) in fresh blocking buffer at 4°C overnight. Embryos were washed four times with TBST for 10–20 min each and were incubated at 4°C overnight. Finally, embryos were incubated with NTMT (200 μl 5 M NaCl; 1 ml 1 M Tris-HCl, pH 9.8; 500 μl 1 M $MgCl_2$; 100 μl 10% Tween 20; and MilliQ water to a final volume of 10 ml) for 20 min and then with 15 μg/ml nitro-blue tetrazolium chloride (NBT) and 175 μg/ml 5-bromo-4-chloro-3-indolyphosphate *p*-toluidine salt (BCIP) in NTMT for 10 min to 2 hours until signals appeared. Pictures were taken with an Olympus microscope. For bamboo shark embryos, experiments were performed for at least two biological replicates.

**ATAC-seq**

Mouse forelimb buds at E9.5, E10.5, E11.5 and E12.5 were dissected, and samples from several individuals were pooled by stage to obtain enough cells. We considered this pooled sample to represent a biological replicate (other replicates were generated using different individuals). To obtain single-cell suspensions, pooled samples were treated with collagenase for 10 min at room temperature. The tissues were then

dissociated into single-cell suspensions by pipetting the mixture and passing it through a 40-μm mesh filter (Funakoshi, Cat. No. HT-AMS-14002); the cell suspension was frozen in CryoStor medium (STEMCELL Technologies, Cat. No. ST07930) with Mr. Frosty (Thermo Scientific, Cat. No. 5100-0001) at −80°C overnight, according to (Milani et al., 2016). An ATAC-seq library was prepared as described (Buenrostro et al., 2013) with some minor modifications. For library preparation, stored cells were thawed in a 38°C water bath and centrifuged at 500*g* for 5 min at 4°C, which was followed by a wash using 50 μl of cold PBS and a second centrifugation at 500*g* for 5 min at 4°C. Ten thousand cells per sample were collected, without distinguishing dead cells, and were lysed using 50 μl of cold lysis buffer (10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$; and 0.1% IGEPAL CA-630). Immediately after lysis, cells were spun at 1000*g* for 10 min at 4°C, and the supernatant was discarded. For the transposition reaction, cells were re-suspended in the transposase reaction mix (25 μl 2× TD buffer, 2.5 μl Tn5 transposase [ in the Nextera DNA Sample Preparation Kit, Illumina, Cat. No. FC-121-1031], and 22.5 μl nuclease-free water) and incubated for 30 min at 37°C. The reaction mix was purified using DNA Clean & Concentrator-5 (Zymo Research, Cat. No. D4014) by adding 350 μl of DNA binding buffer and eluting in a volume of 10 μl. After a five-cycle pre-PCR amplification, the optimal number of PCR cycles was determined by a preliminary PCR using KAPA HiFi HotStart Real-Time Library Amplification Kit and was estimated to be four cycles. The PCR products were purified using 1.8× volumes of Agencourt AMPure XP. As a control, 50 ng of mouse genomic DNA was also transposed following the standard procedure of the Nextera DNA Sample Preparation Kit. Sequencing with HiSeq X was outsourced to Macrogen, Inc., which was carried out with HiSeq Control Software 3.3.76 (Run type: PE151bp). The output was processed with Illumina RTA 2.7.6 for base-calling and with bcl2fastq 2.15.0 for de-multiplexing. Quality control of the obtained fastq files for individual libraries was performed with FASTQC v0.11.5. ATAC-seq was performed with three biological replicates for each stage.

**ATAC-seq data analysis**

The short-read data from ATAC-seq were trimmed and filtered with Trim-Galore! (v0.5.0; options: --paired --phred33 -e 0.1 -q 30). We also removed reads that originated from mitochondrial genome contamination by mapping reads to the mouse mitochondrial genome using bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012). The rest of the reads were mapped onto the mouse genome (mm10) using bwa v0.7.17 with the "mem" option (Li and Durbin, 2010). Among the mapped reads, we removed reads with length > 320 bp to reduce noise. The rest of the reads were further down-sampled to around 83.2 million reads to equalize the sequence depth of every sample. Peak calls were done with MACS2 v2.1.1 (Zhang et al., 2008; options: --nomodel --shift -100 --extsize 200 -f BAMPE -g mm -B -q 0.01; the genomic reads were used as a control for all samples). For FRiP score calculation, a module, "countReadsPerBin.CountReadsPerBin" in deepTools v3.2.1 (Ramírez et al., 2016), was used to count reads in peaks, and these read counts were then divided by the total number of reads. To evaluate reproducibility among the replicates, we first divided the mouse genome into 500-bp bins. Then, the ATAC-seq peaks were re-distributed into these bins with bedtools (Quinlan and Hall, 2010; options: intersect -F 0.4 -f 0.4 -e -wo). Peaks of >500 bp were subdivided into 500-bp-long regions, and those of <500 bp were extended to fit within the closest 500 bp window. Subsequently, these peaks were converted into one-hot vectors, in which '1' means that a 500-bp-long genomic region harbors an ATAC-seq peak. Genomic regions that lacked ATAC-seq peaks in all data were omitted. Using these one-hot vectors, Euclidean distances between the ATAC-seq data were calculated (Figure 5–figure supplement 1A).

For the conservation analysis, the significant variation in the length of ATAC-seq peaks complicated this evaluation. To deal with such variation, we the ATAC-seq peaks were re-distributed into 100-bp bins with bedtools (Quinlan and Hall, 2010; options: intersect -F 0.4 -f 0.4 -e -wo) as described above. The sequences in these peaks were retrieved with BLASTN v2.7.1 against the genomes of 16 vertebrate species listed in Supplementary file 10 (BLASTN options: -task dc-megablast -max_target_seqs 1). The blast hits that scored ≥ 40 were considered as conserved sequences. In this way, the final figures shown in Figure 5C represent the fraction of the total conserved sequence length in the peaks of each stage rather than the number of conserved peaks. For confirmation, we also used a different alignment algorithm, LAST v961 (Kiełbasa et al., 2011) to find conserved sequences. To generate mouse genome databases for LAST, we first masked repeat sequences with N and split the genome file into multiple files, each of which contained a

single chromosome sequence. Then, databases were generated using lastdb (options: -cR01). Alignments with the bamboo shark genome (Cpunctatum_v1.0; https://transcriptome.riken.jp/squalomix/resources/01.GCA_003427335.1_Cpunctatum_v1.0_genomic.rn.fna.gz) and the alligator genome (ASM28112v3) were carried out by lastal (options: -a1 -m100). Only a unique best alignment was selected using last-split. These alignment results were converted into the bed format, and regions that overlapped with the ATAC-seq peaks that were subdivided into 100-bp bins were counted.

For the clustering analysis, we converted the alignment files of the ATAC-seq reads into mapped reads in bins per million (BPM) coverage values with 200-bp resolution using bamCoverage in deepTools v3.2.1 (Ramírez et al., 2016; options: -of bedgraph --normalizeUsing BPM --effectiveGenomeSize 2652783500 -e -bs 200). Then, BPMs at the summits of ATAC-seq peaks and an additional 600 bp to the left and to the right of each summit (1400 bp in total) were collected and clustered by t-SNE (https://github.com/DmitryUlyanov/Multicore-TSNE; hyper parameters: perplexity = 30.0, n_iter = 5000) followed by hierarchical clustering (hyper parameters: method = “ward”, metric = “euclidean”). Enriched motifs were detected using a Perl script, findMotifsGenome.pl in HOMER v4.10.4 (Heinz et al., 2010; options: -size 100 -mask). To count the number of motif occurrences, “-find” option of findMotifsGenome.pl was used, and sequences that scored ≥ 75% of the highest motif score were counted. For GO analysis, annotatePeaks.pl in HOMER was used. For the tissue-specificity analysis, we downloaded several aligned and unaligned reads of ATAC-seq experiments on 25 different tissues from the ENCODE web site (https://www.encodeproject.org/; see Supplementary file 10 for a complete list), and peaks were called as described above. Then, peaks that did not overlap with other tissues/cells were detected using bedtools.

## References


Ahn D, Ho RK. 2008. Tri-phasic expression of posterior Hox genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. *Dev Biol* **322**:220–233.

Amemiya CT, Alföldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, Organ C, Chalopin D, Smith JJ, Robinson M, Dorrington R a, Gerdol M, Aken B, Biscotti

MA, Barucca M, Baurain D, Berlin AM, Blatch GL, Buonocore F, Burmester T, Campbell MS, Canapa A, Cannon JP, Christoffels A, De Moro G, Edkins AL, Fan L, Fausto AM, Feiner N, Forconi M, Gamieldien J, Gnerre S, Gnirke A, Goldstone J V, Haerty W, Hahn ME, Hesse U, Hoffmann S, Johnson J, Karchner SI, Kuraku S, Lara M, Levin JZ, Litman GW, Mauceli E, Miyake T, Mueller MG, Nelson DR, Nitsche A, Olmo E, Ota T, Pallavicini A, Panji S, Picone B, Ponting CP, Prohaska SJ, Przybylski D, Saha NR, Ravi V, Ribeiro FJ, Sauka-Spengler T, Scapigliati G, Searle SMJ, Sharpe T, Simakov O, Stadler PF, Stegeman JJ, Sumiyama K, Tabbaa D, Tafer H, Turner-Maier J, van Heusden P, White S, Williams L, Yandell M, Brinkmann H, Volff J-N, Tabin CJ, Shubin N, Schartl M, Jaffe DB, Postlethwait JH, Venkatesh B, Di Palma F, Lander ES, Meyer A, Lindblad-Toh K. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**:311–6. doi:10.1038/nature12027

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science (80- )* **340**:1234167 1–7. doi:10.1126/science.1234167

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, Berlin AM, Campbell MS, Barrell D, Martin KJ, Mulley JF, Ravi V, Lee AP, Nakamura T, Chalopin D, Fan S, Wcisel D, Caestro C, Sydes J, Beaudry FEG, Sun Y, Hertel J, Beam MJ, Fasold M, Ishiyama M, Johnson J, Kehr S, Lara M, Letaw JH, Litman GW, Litman RT, Mikami M, Ota T, Saha NR, Williams L, Stadler PF, Wang H, Taylor JS, Fontenot Q, Ferrara A, Searle SMJ, Aken B, Yandell M, Schneider I, Yoder JA, Volff JN, Meyer A, Amemiya CT, Venkatesh B, Holland PWH, Guiguen Y, Bobe J, Shubin NH, Di Palma F, Alföldi J, Lindblad-Toh K, Postlethwait JH. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet* **48**:427–437. doi:10.1038/ng.3526

Brunet LJ, McMahon JA, McMahon AP, Harland RM. 1998. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science (80- )* **280**:1455–1457. doi:10.1126/science.280.5368.1455

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**:1213–1218. doi:10.1038/nmeth.2688

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST + : architecture and applications. *BMC Bioinformatics* **9**:1–9. doi:10.1186/1471-2105-10-421

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**:1972–1973. doi:10.1093/bioinformatics/btp348

Chal J, Pourquié O. 2017. Making muscle: Skeletal myogenesis in vivo and in vitro. *Development*. doi:10.1242/dev.151035

Chen Y, Knezevic V, Ervin V, Hutson R, Ward Y, Mackem S. 2004. Direct interaction with Hoxd protein reverses Gli3-repressor function to promote digit formation downstream of Shh. *Development* **131**:2339–2347. doi:10.1242/dev.01115

Clack JA. 2009. The fin to limb transition: new data, interpretations, and hypotheses from paleontology and developmental biology. *Annu Rev Earth Planet Sci* **37**:163–179. doi:10.1146/annurev.earth.36.031207.124146

Dahn RD, Davis MC, Pappano WN, Shubin NH. 2007. Sonic hedgehog function in chondrichthyan fins and the evolution of appendage patterning. *Nature* **445**:311–314. doi:10.1038/nature05436

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK, Onate KC, Graham K, Miyasato SR, Dreszer TR, Strattan JS, Jolanki O, Tanaka FY, Cherry JM. 2018. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res* **46**:D794–D801. doi:10.1093/nar/gkx1081

Davis MC, Dahn RD, Shubin NH. 2007. An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish. *Nature* **447**:473–476. doi:10.1038/nature05838

Deschamps J, Duboule D. 2017. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes Dev*. doi:10.1101/gad.303123.117

Draper BW, Stock DW, Kimmel CB. 2003. Zebrafish fgf24 functions with fgf8 to promote posterior mesodermal development. *Development* **130**:4639–4654. doi:10.1242/dev.00671

Duboule D. 1994. Temporal colinearity and the phylotypic progression: A basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* **120**:135–142.

Eisenberg E, Levanon EY. 2013. Human housekeeping genes, revisited. *Trends Genet*. doi:10.1016/j.tig.2013.05.010

Fischer S, Draper BW, Neumann CJ. 2003. The zebrafish fgf24 mutant identifies and additional level of Fgf signaling involved in vertebrate forelimb initiation. *Development* **130**:3515–3524. doi:10.1242/dev.00537

Fowler DA, Larsson HCE. 2020. The tissues and regulatory pattern of limb chondrogenesis. *Dev Biol*. doi:10.1016/j.ydbio.2020.04.009

Freitas R, Zhang G, Cohn MJ. 2007. Biphasic Hoxd gene expression in shark paired fins reveals an ancient origin of the distal limb domain. *PLoS One* **2**:e754.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**:644–652. doi:10.1038/nbt.1883

Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandewege MW, St John JA, Capella-Gutiérrez S, Castoe TA, Kern C, Fujita MK, Opazo JC, Jurka J, Kojima KK, Caballero J, Hubley RM, Smit AF, Platt RN, Lavoie CA, Ramakodi MP, Finger JW, Suh A, Isberg SR, Miles L, Chong AY, Jaratlerdsiri W, Gongora J, Moran C, Iriarte A, McCormack J, Burgess SC, Edwards S V., Lyons E, Williams C, Breen M, Howard JT, Gresham CR, Peterson DG, Schmitz J, Pollock DD, Haussler D, Triplett EW, Zhang G, Irie N, Jarvis ED, Brochu CA, Schmidt CJ, McCarthy FM, Faircloth BC, Hoffmann FG, Glenn TC, Gabaldón T, Paten B, Ray DA. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science (80- )* **346**. doi:10.1126/science.1254449

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. 2013. De novo transcript sequence

reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**:1494–1512. doi:10.1038/nprot.2013.084

Hao Y, Lee HJ, Baraboo M, Burch K, Maurer T, Somarelli JA, Conant GC, Hoffmann F. 2020. Baby Genomics: Tracing the Evolutionary Changes That Gave Rise to Placentation. *Genome Biol Evol* **12**:35–47. doi:10.1093/gbe/evaa026

Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, Tatsumi K, Tanaka K, Motone F, Kageyama Y, Nozu R, Adachi N, Nishimura O, Nakagawa R, Tanegashima C, Kiyatake I, Matsumoto R, Murakumo K, Nishida K, Terakita A, Kuratani S, Sato K, Hyodo S, Kuraku S. 2018. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat Ecol Evol* **2**:1761–1771. doi:10.1038/s41559-018-0673-5

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**:576–589. doi:10.1016/j.molcel.2010.05.004

Hu H, Uesaka M, Guo S, Shimai K, Lu TM, Li F, Fujimoto S, Ishikawa M, Liu S, Sasagawa Y, Zhang G, Kuratani S, Yu JK, Kusakabe TG, Khaitovich P, Irie N. 2017. Constrained vertebrate evolution by pleiotropic genes. *Nat Ecol Evol* **1**:1722–1730. doi:10.1038/s41559-017-0318-0

Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* **2**:248.

Johanson Z, Joss J, Boisvert CA, Ericsson R, Sutija M, Ahlberg PE. 2007. Fish fingers: Digit homologues in Sarcopterygian fish fins. *J Exp Zool Part B Mol Dev Evol* **308**:757–768.

Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**:811–814.

Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**:3059–3066. doi:10.1093/nar/gkf436

Kherdjemil Y, Lalonde RL, Sheth R, Dumouchel A, de Martino G, Pineault KM, Wellik DM, Stadler HS, Akimenko M-A, Kmita M. 2016. Evolution of Hoxa11 regulation in vertebrates is linked to the pentadactyl state. *Nature* **539**:89–92. doi:10.1038/nature19813

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res* **21**:487–493. doi:10.1101/gr.113985.110

Kuss P, Villavicencio-Lorini P, Witte F, Klose J, Albrecht AN, Seemann P, Hecht J, Mundlos S. 2009. Mutant Hoxd13 induces extra digits in a mouse model of synpolydactyly directly and by decreasing retinoic acid synthesis. *J Clin Invest* **119**:146–156. doi:10.1172/JCI36851

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**:357–359. doi:10.1038/nmeth.1923

Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**. doi:10.1186/1471-2105-12-124

Leiboff S, Hake S. 2019. Reconstructing the transcriptional ontogeny of maize and sorghum supports an inverse hourglass model of inflorescence development. *Curr Biol* **29**:3410-3419.e3. doi:10.1016/j.cub.2019.08.044

Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, Fernandez-Valverde SL, Nakanishi N, Simmons D, Simakov O, Larsson T, Liu SY, Jerafi-Vider A, Yaniv K, Ryan JF, Martindale MQ, Rink JC, Arendt D, Degnan SM, Degnan BM, Hashimshony T, Yanai I. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature* **531**:637–641. doi:10.1038/nature16994

Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental Milestones Punctuate Gene Expression in the Caenorhabditis Embryo. *Dev Cell* **22**:1101–1108. doi:10.1016/j.devcel.2012.04.004

Lewandoski M, Sun X, Martin GR. 2000. Fgf8 signalling from the AER is essential for normal limb development. *Nat Genet* **26**:460–463. doi:10.1038/82609

Lewis PM, Dunn MP, McMahon JA, Logan M, Martin JF, St-Jacques B, McMahon AP. 2001. Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**:599–612. doi:10.1016/S0092-8674(01)00369-5

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323. doi:10.1186/1471-2105-12-323

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**:589–595. doi:10.1093/bioinformatics/btp698

Maier JA, Rivas-Astroza M, Deng J, Dowling A, Oboikovitz P, Cao X, Behringer RR, Cretekos CJ, Rasweiler JJ, Zhong S, Sears KE. 2017. Transcriptomic insights into the genetic basis of mammalian limb diversity. *BMC Evol Biol* **17**. doi:10.1186/s12862-017-0902-6

Mariani F V, Ahn CP, Martin GR. 2008. Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature* **453**:401–405.

Metscher BD, Takahashi K, Crow K, Amemiya C, Nonaka DF, Wagner GP. 2005. Expression of Hoxa-11 and Hoxa-13 in the pectoral fin of a basal ray-finned fish, Polyodon spathula: Implications for the origin of tetrapod limbs. *Evol Dev* **7**:186–195. doi:10.1111/j.1525-142X.2005.05021.x

Milani P, Escalante-Chong R, Shelley BC, Patel-Murray NL, Xin X, Adam M, Mandefro B, Sareen D, Svendsen CN, Fraenkel E. 2016. Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Sci Rep* **6**. doi:10.1038/srep25474

Montavon T, Le Garrec JF, Kerszberg M, Duboule D. 2008. Modeling Hox gene regulation in digits: Reverse collinearity and the molecular origin of thumbness. *Genes Dev* **22**:346–359. doi:10.1101/gad.1631708

Nakamura T, Gehrke AR, Lemberg J, Szymaszek J, Shubin NH. 2016. Digits and fin rays share common developmental histories. *Nature* **537**:1–16. doi:10.1038/nature19322

Nakamura T, Klomp J, Pieretti J, Schneider I, Gehrke AR, Shubin ANH. 2015. Molecular mechanisms underlying the exceptional adaptations of batoid fins. *Proc Natl Acad Sci U S A* **112**:15940–15945. doi:10.1073/pnas.1521818112

Nomura R, Kamei E, Hotta Y, Konishi M, Miyake A, Itoh N. 2006. Fgf16 is essential for pectoral fin bud formation in zebrafish. *Biochem Biophys Res Commun* **347**:340–346. doi:10.1016/j.bbrc.2006.06.108

Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S, Plajzer-Frick I, Shoukry M, Afzal V, Rubenstein JLR, Rubin EM, Pennacchio LA, Visel A. 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**:1521–1531. doi:10.1016/j.cell.2013.11.033

Onimaru K, Kuraku S, Takagi W, Hyodo S, Sharpe J, Tanaka M. 2015. A shift in anterior–posterior positional information underlies the fin-to-limb evolution. *Elife* **4**:e07048. doi:10.7554/eLife.07048.001

Onimaru K, Motone F, Kiyatake I, Nishida K, Kuraku S. 2018. A staging table for the embryonic development of the brownbanded bamboo shark (Chiloscyllium punctatum). *Dev Dyn* **247**:712–723. doi:10.1002/dvdy.24623

Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. 2013. The Hourglass and the Early Conservation Models-Co-Existing Patterns of Developmental Constraints in Vertebrates. *PLoS Genet* **9**. doi:10.1371/journal.pgen.1003476

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. doi:10.1093/bioinformatics/btq033

Raff RA. 1996. The shape of life: genes, development, and the evolution of animal form. Chicago: University of Chicago Press.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**:W160–W165. doi:10.1093/nar/gkw257

Ravi V, Venkatesh B. 2008. Rapidly evolving fish genomes and teleost diversity. *Curr Opin Genet Dev*. doi:10.1016/j.gde.2008.11.001

Rebeiz M, Tsiantis M. 2017. Enhancer evolution and the origins of morphological novelty. *Curr Opin Genet Dev*. doi:10.1016/j.gde.2017.04.006

Sakamoto K, Onimaru K, Munakata K, Suda N, Tamura M, Ochi H, Tanaka M. 2009. Heterochronic shift in Hox-mediated activation of Sonic hedgehod leads to morphological changes during fin development. *PLoS One* **4**. doi:10.1371/journal.pone.0005121

Schneider I, Aneas I, Gehrke AR, Dahn RD, Nobrega M a, Shubin NH. 2011. Appendage expression driven by the Hoxd global control region is an ancient gnathostome feature. *Proc Natl Acad Sci U S A* **108**:12782–12786. doi:10.1073/pnas.1109993108

Sheth R, Marcon L, Bastida MF, Junco M, Quintana L, Dahn R, Kmita M, Sharpe J, Ros M a. 2012. Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science (80-)* **338**:1476–1480. doi:10.1126/science.1226804

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212. doi:10.1093/bioinformatics/btv351

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. doi:10.1093/bioinformatics/btu033

Tulenko FJ, Massey JL, Holmquist E, Kigundu G, Thomas S, Smith SME, Mazan S, Davis MC. 2017. Finfold development in paddlefish and catshark and implications for the evolution of the autopod. *Proc R Soc B Biol Sci* **284**:20162780. doi:10.1098/rspb.2016.2780

Tzchori I, Day TF, Carolan PJ, Zhao Y, Wassif CA, Li L, Lewandoski M, Gorivodsky M, Love PE, Porter FD, Westphal H, Yang Y. 2009. LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136**:1375–1385. doi:10.1242/dev.026476

Uesaka M, Kuratani S, Takeda H, Irie N. 2019. Recapitulation-like developmental transitions of chromatin accessibility in vertebrates. *Zool Lett* **5**. doi:10.1186/s40851-019-0148-9

Vokes SA, Ji H, Wong WH, McMahon AP. 2008. A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes Dev* **22**:2651–2663. doi:10.1101/gad.1693008

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**:281–285. doi:10.1007/s12064-012-0162-3

Wagner GP, Zhang J. 2011. The pleiotropic structure of the genotype-phenotype map: The evolvability of complex organisms. *Nat Rev Genet*. doi:10.1038/nrg2949

Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D, Wang B, Ming Y, Chen Y, Zheng Y, Kuraku S, Pignatelli M, Herrero J, Beal K, Nozawa M, Li Q, Wang Juan, Zhang H, Yu L, Shigenobu S, Wang Junyi, Liu J, Flicek P, Searle S, Weng J, Kuratani S, Yin Y, Aken B, Zhang G, Irie N. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet* **45**:701–706. doi:10.1038/ng.2615

Westerfield M. 2000. The Zebrafish Book, Eugene.

Woltering JM, Holzem M, Meyer A. 2019. Lissamphibian limbs and the origins of tetrapod hox domains. *Dev Biol* **456**:138–144. doi:10.1016/j.ydbio.2019.08.014

Woltering JM, Irisarri I, Ericsson R, Joss JMP, Sordino P, Meyer A. 2020. Sarcopterygian fin ontogeny elucidates the origin of hands with digits. *Sci Adv* **6**. doi:10.1126/sciadv.abc3510

Woltering JM, Noordermeer D, Leleu M, Duboule D. 2014. Conservation and Divergence of Regulatory Strategies at Hox Loci and the Origin of Tetrapod Digits. *PLoS Biol* **12**:e1001773. doi:10.1371/journal.pbio.1001773

Yakushiji-Kaminatsui N, Lopez-Delisle L, Bolt CC, Andrey G, Beccari L, Duboule D. 2018. Similarities and differences in the regulation of HoxD genes during chick and mouse limb development. *PLoS Biol* **16**. doi:10.1371/journal.pbio.3000004

Yonei-Tamura S, Abe G, Tanaka Y, Anno H, Noro M, Ide H, Aono H, Kuraishi R, Osumi N, Kuratani S, Tamura K. 2008. Competent stripes for diverse positions of limbs/fins in gnathostome embryos. *Evol Dev* **10**:737–745. doi:10.1111/j.1525-142X.2008.00288.x

Zákány J, Kmita M, Duboule D. 2004. A dual role for Hox genes in limb anterior-posterior asymmetry. *Science (80- )* **304**:1669–1672. doi:10.1126/science.1096049

Zeller R, López-Ríos J, Zuniga A, Zúñiga A. 2009. Vertebrate limb bud development: Moving towards integrative analysis of organogenesis. *Nat Rev Genet* **10**:845–858. doi:10.1038/nrg2681

Zhang J, Wagh P, Guay D, Sanchez-Pulido L, Padhi BK, Korzh V, Andrade-Navarro MA, Akimenko M-A. 2010. Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature* **466**:234–237. doi:10.1038/nature09137

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Shirley XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**. doi:10.1186/gb-2008-9-9-r137

**Acknowledgments:** We thank Kenta Sumiyama, James Shape and Masahiro Uesaka for fruitful discussions; Itoshi Nikaido and Laboratory for Bioinformatics Research for providing computational resources; Itsuki Kiyatake, Kiyonori Nishida, and Osaka Aquarium Kaiyukan for kindly providing bamboo shark eggs; Laboratory for Animal Resources and Genetic Engineering for supplying mouse embryos; the ENCODE consortium and the ENCODE production laboratories for generating ATAC-seq data sets.

**Funding:** This work was supported in part by JSPS KAKENHI grant number 17K15132, a Special Postdoctoral Researcher Program of RIKEN, and a research grant from MEXT to the RIKEN Center for Biosystems Dynamics Research.

**Author contributions:** KO and SK designed the project. KO, KT, CT, and MK performed experiments. KO and ON analyzed data. All authors contributed to writing the manuscript.

**Competing interests:** The authors declare no competing interests.

**Data and materials availability:** RNA-seq and ATAC-seq data sets generated during the current study are available in the Gene Expression Omnibus (GEO) repository under accession number GSE136445. Other sequence data and raw data are available in the figshare (DOI: 10.6084/m9.figshare.9928541). Code for clustering analysis is available at https://github.com/koonimaru/easy_heatmapper. Materials related to this paper are available upon request from the corresponding authors.

## Figure legends and tables

### Figure 1–6, Table 1, 2

**Figure 1 | Transcriptome analysis and orthology assignment.** (**A**)**,** The skeletal patterns of a mouse limb (top) and a bamboo shark pectoral fin (bottom). Anterior is to the top; distal is to the right. (**B**) Mouse forelimb buds and bamboo shark pectoral fin buds that were analyzed by RNA-seq. (**C**) Comparison of the accuracy of three orthology assignment methods. Vertical axis, the percentages of correctly assigned *Hoxa* and *Hoxd* paralogs (black bars) and *Fgf* paralogs (white bars). (**D**) Heat map visualization of the transcription profile of *Hoxa* and *Hoxd* genes in mouse limb buds (left) and bamboo shark fin buds (right) with scaled TPMs.

**Figure 2 | Detection of heterochronic gene expression between mouse limb buds and bamboo shark fin buds.** (**A**) Clustering analysis of gene expression dynamics. Each column represents an ortholog pair between the bamboo shark and the mouse. Each row indicates scaled gene expression at a time point indicated to the right of the heat map. Values are scaled TPMs. (**B, C**) Whole-mount *in situ* hybridization of *Hand2* (**B**) and *Vcan* (**C**) as examples of the heterochronic genes detected in (**A**). Asterisks, background signals; scale bars, 200 μm. Error bars: SEM.

**Figure 3 | *Shh* pathway in mouse limb buds and bamboo shark fin buds.**

(**A, B**) Scaled expression of *Shh* and related genes in mouse limb buds (**A**) and bamboo shark fin buds (**B**), respectively. The rectangles indicate the expression peaks of *Shh, Hoxd9*, and *Hoxd10* (magenta), Shh target genes (yellow) and *Hoxd11* and *Hoxd12* (green). (**C, D**) Whole-mount *in situ* hybridization of *Ptch1* and *Hoxd12* in mouse limb buds (**C**) and bamboo shark fin buds (**D**); scale bars, 200 μm. White arrowheads in **D** indicate restricted expression of *Ptch1* in bamboo shark fin buds. Black arrowheads in **C** and **D** indicate anteriorly extended expression of *Hoxd12*.

**Figure 4 | Hourglass-shaped conservation of the transcriptome profile between fins and limbs.** (**A**) Euclidean distances of the transcriptome profiles. Every combination of time points of bamboo shark fin buds and mouse limb buds is shown. The darker colors indicate a greater similarity between gene expression profiles. (**B**) A line plot of the Euclidean distances shown in (**A**). The *x* axis indicates the mouse limb stages, and the *y* axis is the Euclidean distance. (**C**) The same as (**A**) except that only *Hoxd* genes are included. (**D, E**) Scatter plots of the first and second principal components (**D**) and of the second and third components (**e**). Arrows in (**E**) indicate the time-order of transcriptome data. (**F**) Count of tissue-associated genes expressed in mouse forelimb buds. Genes with 0.65 ≤ entropy were counted.

**Figure 5 | Hourglass-shaped conservation of OCRs in mouse limb development.** (**A**) ATAC-seq signals in the enhancer regions of the HoxA cluster. e1 to e4, known limb enhancers. Green vertical lines below the signals, peak regions determined by MACS2. (**B**) Comparison of a quality index, FRiP, for ATAC-seq data. Blue bars are samples with a FRiP score > 0.2. The number in the end of the label name indicates the replicate number. (**C**) Conservation analysis of sequences in ATAC-seq peaks with BLASTN. The values to the right of each graph indicate the fraction of conserved sequences in the total peak regions. The common name of each genome sequence is indicated above the graph. The not-conserved heatmap indicates the fraction of sequences that were not aligned to any genome sequences and thus serve as a negative control. (**D**) Temporal changes of sequence conservation frequency in ATAC-seq peaks with LAST. Error bars: SEM.

**Figure 6 | Temporal dynamics of OCRs during mouse limb development.** (**A**) The heatmap (left) shows whole-genome clustering of ATAC-seq peaks. Each row indicates a particular genome region with a length of 1400 bp. Columns indicate developmental stages. C1−C8 are cluster numbers. The motifs (right) show the rank of enriched motifs in the sequences of each cluster. (**B**) Top, volcano plots of ATAC-seq signals between indicated stages (p-values, two-sided Student's t-test). Bottom, the counts of differential signals (black dots in the top panel). + and − are genomic regions with increased or decreased signals, respectively. (**C**) The fraction of limb-specific OCRs for each cluster.

**Table 1 Assembly statistics of bamboo shark transcriptome**

| Characteristic | Bamboo shark transcriptome | Bamboo shark gene model (Hara et al., 2018) |
|---|---|---|
| Total number of sequences | 222015 | 34038 |
| Total sequence length (bp) | 195541367 | 36633751 |
| Average length (bp) | 880 | 1076 |
| Maximum length (bp) | 18451 | 108594 |
| N count | 0 | 10208 |
| L50 | 24765 | 5666 |
| N50 length (bp) | 2075 | 1749 |
| Protein coding | 63898 | 34038 |
| Orthology detected | 41633 | 18180 |
| Unique orthologs | 14139 | 14907 |
| Unique orthologs without gene symbols | 1821 | 1780 |
| Unique orthologs only in elephantfish | 826 | 552 |
| Sequences with no orthology | 20892 | 15254 |
| Orthologs with mouse genes | 12326 | 13005 |

**Table 2 PCA loadings**

| Loading axis: PC2 | | |
|---|---|---|
| Gene symbol | Cluster name | Loading |
| TRHDE | C8 | 0.31 |
| PAX9 | C11 | 0.3 |
| COL9A2 | C8 | 0.3 |
| RTN4R | C8 | 0.3 |
| APC2 | C9 | 0.3 |

| | | |
|---|---|---|
| CNMD | C8 | 0.29 |
| HOXD13 | C8 | 0.29 |
| FAM69C | C8 | 0.29 |
| WFIKKN2 | C8 | 0.29 |
| HOXA13 | C8 | 0.29 |
| LRRN3 | C12 | 0.29 |
| HPSE2 | C9 | 0.29 |
| SERPINB1A | C11 | 0.29 |
| CDKN2B | C8 | 0.28 |
| LTBP1 | C8 | 0.28 |
| CDH19 | C8 | 0.28 |
| PDZD2 | C8 | 0.28 |
| NLGN3 | C9 | 0.28 |
| MATN1 | C8 | 0.28 |
| MYOD1 | C8 | 0.28 |
| TSPAN11 | C12 | 0.28 |
| SERINC2 | C9 | 0.28 |
| FYB | C8 | 0.28 |
| KIF1A | C8 | 0.28 |
| COL9A3 | C8 | 0.28 |

**Supplementary Materials**

**legends for figure supplements and supplementary files (total 28, found below)**

**Supplementary file 1 to 11****

**Supplementary data*****

**found in separate files that accompany this manuscript.

***found in https://figshare.com/articles/Onimaru_et_al_Supplementary_Data/9928541 (DOI: 10.6084/m9.figshare.9928541).

****Other NGS-related data are available at GSE136445 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136445).

**Figure 1–figure supplement 1. Schematic representation of the orthology assignment algorithm.** Red arrows, the main flow of the algorithm. Black arrows, orthology assignment for cartilaginous fish-specific genes. Gray arrows, parallel retrieving of orthologs of mouse genes from other animals. Red rectangles, best hits across other animal genes or in elephantfish genes. Green rectangles, best hits among each animal genome. Note that this schematic explains how the orthology of abstract genes "bamboo shark gene X" and "mouse gene Y" are assigned. First, using BLASTP, putative orthologs of bamboo shark genes are retrieved from other animal genomes, such as human, mouse, alligator, and elephantfish. Then, all BLASTP results except those from elephantfish are concatenated to find a best scored gene across species (cross-species best hit). In this schematic, the alligator gene XP001 is the best hit. In parallel, putative orthologs of mouse genes are also retrieved from the same set of animal genomes. If there is a mouse gene Y that has a best hit with alligator XP001, this mouse gene Y and bamboo shark gene X are considered to be an orthologous pair.

**Figure 1–figure supplement 2. Molecular phylogenetic tree for Fgf family.** The tree was inferred with the maximum-likelihood method. The support values at nodes indicate bootstrap probabilities. Genes highlighted in red are bamboo shark genes (can be converted into the original gene ID by replacing "g" with "Chipu" and fill the digit with 0 to be 7 figure number in total).

**Figure 1–figure supplement 3. Additional molecular phylogenetic trees for Fgf8, Fgf11, Fgf12, and Fgf13.** These trees are shown because alignment sequences used in Figure 1–figure supplement 2 are truncated or absent in these genes. The tree was inferred with the maximum-likelihood method. The support values at nodes indicate bootstrap probabilities. Genes highlighted in red are bamboo shark genes.

**Figure 1–figure supplement 4. Comparison between the TPM and TMM.** (**A**) Visualization of the effect of normalization by showing a housekeeping gene family, *Ndufa*. Left panels show TMM (trimmed mean of M) and TPM (transcripts per million) calculated by RSEM. Right panels show these values with additional normalization using several other housekeeping genes (*Atp5j, Atp5h, Atp5g3, Psmc3, Psmc5, Psmd7, Mrpl54, Mrpl46, Polr2e, Polr1b, Mrpl2*). Housekeeping genes are selected from a previously published list (https://www.tau.ac.il/~elieis/HKG/HK_genes.txt; Eisenberg and Levanon, 2013). All expression values are standardized by setting the maximum expression value of each gene as '1'. Note that because housekeeping genes do not change their expression amount over time, these expression values should be close to '1' (i.e., all colors should be dark blue) with some exceptions. However, the intact TMM (top, left panel) is apparently biased, in that the majority of *Ndufa* genes show their strongest expression at E9.5, with sharp decreases at other stages. This bias can be corrected by normalization with other housekeeping genes (top right panel). In contrast, the intact TPM (bottom, left panel) has a weaker bias than TMM. Additional normalization (bottom, right panel) has less of an effect. Therefore, this study used the intact TPM. (**B**) Euclidean distances of transcriptome data between mouse samples (left) and between bamboo shark samples (right). Whereas the close relation of the replicates of mouse samples can be seen from this heat map, the replicates of bamboo shark samples show less similarity. This noisy data may be attributed to the fact that

there is no established strain of the bamboo shark and/or that bamboo shark embryos were staged by morphology but not physical time. However, the average of replicates seems to mitigate the noise of the bamboo shark samples, because *Hox* gene expression showed a smooth temporal collinearity as seen in Figure 1D.

**Figure 1–figure supplement 5. The effect of scaling methods to housekeeping genes.** (**A**) A simple example for comparing expression distances between two species. Species 1 and 2 are imaginery simple species that have two genes (gene 1 and 2) and three developmental time points (t1, t2 and t3). Distances in the bottom are the Euclidean distance between two species at each stage. (**B, C**) Housekeeping gene expressions with intact TPM values and different scaling methods in mouse limb buds (**B**) and bambooshark fin buds (**C**). Intact TPM, TPM values without any scaling methods; Max 1, TPM values were scaled by setting the highest TPM in each gene of each species to '1'; Z-score, the mean expression value was subtracted from each expression value and each result was then divided by the standard deviation; Unit vector, expression values were divided by the norm; Log10, $\log_{10}$ transformation. These housekeeping genes are listed in both a human housekeeping gene list (https://www.tau.ac.il/~elieis/HKG/HK_genes.txt) and the BUSCO data set (thus these genes are likely conserved in most vertebrates). Note that whereas the expression values of the housekeeping genes were almost constant during development, Z-score scaling amplifies subtle differences between stages. In addition, intact TPM values were not readily comparable between limb buds and bamboo shark fin buds (e.g., the maximum value of POLR1B in mouse limb buds was roughly twice as high as that of bamboo shark fin buds). Error bars are not displayed.

**Figure 1–figure supplement 6. The effect of scaling methods to heterochronic genes.** (**A, B**) Heterochronic gene expressions with intact TPM values and different scaling methods in mouse limb buds (**A**) and bambooshark fin buds (**B**)**.** See the legend of Figure 1–figure supplement 5 for scaling methods. Error bars are omitted. (**C**) The total Euclidean distance with respect to gene expression for the three housekeeping and the three heterochronic genes between mouse limb buds and bamboo shark fin buds. Using the housekeeping genes shown in (A, B) and the different scaling methods, the graph shows the summation of Euclidean distances between all combinations of mouse limb and bamboo shark fin stages. (D) The ratios of the Euclidean distances for housekeeping genes to those for heterochronic genes as shown in C.

**Figure 1–figure supplement 7. Examination of quantitative collinearity of *Hoxd* genes.** TPM values of 5′ *Hoxd* genes in mouse limb buds at E12.5 (left) and bamboo shark fin buds at st. 31 (right, orange bars) and st. 32 (right, blue bars). Error bars, SEM. Note that the genomic locus of *Hoxd* genes was positively correlated with their expression amount in the mouse limb bud, whereas no such correlation was found in the bamboo shark fin bud.

**Figure 1–figure supplement 8. Expression profile of genes related to cellular differentiation.** (**A, B**) Scaled TPM values of indicated genes related to chondrogenesis (**A**) and myogenesis (**B**). Error bars, SEM.

**Figure 2-figure supplement 1. Other heterochronic genes.** (**A**) Left panels, whole mount *in situ* hybridization of *Aldh1a2* (one of the genes from the cluster Heterochronic1) in bamboo shark fin buds and mouse limb buds. Right panels, TPM values of *Aldh1a2*. Arrowheads indicate the late-stage expression of *Aldh1a2* in bamboo shark fin buds. Scale bars, 200 μm. Error bars, SEM. (**B**) Heatmap of genes that exhibit an inverse relation to Heterochronic2 genes in Figure 2A. Yellow empty box, genes that exhibit relatively sharp upregulation in bamboo shark fin buds and downregulation in mouse limb buds over time. (**C**) Comparison of *Fgf* gene expression. Vertical axis, TPM values; error bars, SEM; N/A, not applicable because of the absence of *Fgf24* in the mouse genome.

**Figure 3-figure supplement 1. The temporal dynamics of the Shh pathway based on intact TPM values.** Red filled rectangles, the expression peak of *Shh*, *Hoxd9*, and *Hoxd10*; yellow filled rectangles, the expression peak of *Shh* target genes; green filled rectangles, the expression peak of *Hoxd11* and *Hoxd12*.

**Figure 4–figure supplement 1. Confirmation analyses of the transcriptome comparison.** Cross-species comparisons of transcriptome data between the two species with indicated distance methods. Note that these methods consistently show the closest distance around E10.5 and st. 27.5–30.

**Figure 4–figure supplement 2. Additional PCA data and counts for stage- and tissue-associated genes.** (**A**) The ratio of explained variable for each of the principal components from Figure 4D and E. (**B**) Euclidean distance measures using the indicated principal components. Note that individual principal components do not reproduce the hourglass-shaped conservation shown in Figure 4A, but PC1, PC2, and PC3 are sufficient for the most part to reproduce Figure 4A. (**C**) Percentage of stage-associated genes with [z ≤ 1.0] for mouse limb buds (left) and bamboo shark fin buds (right). Note that both species showed a low percentage of stage-associated genes during the middle stages of development. (**D**) Number (left) and fraction (right) of tissue-associated genes expressed in mouse limb buds. Tissue specificity was evaluated by entropy using RNA-seq data from 71 mouse tissues. A gene with entropy ≥ 0.65 was considered a tissue-specific gene. In the right panel, gene counts were normalized based on the number of total expressed genes. Note that the number of tissue-associated genes was lowest at E10.5.

**Figure 5–figure supplement 1. ATAC-seq quality control.** (**A**) Correlation distance between samples. The numbers in the end of the sample names indicate the replicates of indicated stages. Darker color means more similar gene expressions. (**B**) Percentage of peak regions in the genome sequence. (**C**) ATAC-seq signals in BPM (blue signals), peak regions (blue rectangles) and the known limb enhancers of HoxA cluster (red rectangles, e1–e19). Note that only e5 is not covered by ATAC-seq data.

**Figure 5–figure supplement 2. Conservation measures of OCRs.** (**A, B**) The absolute count of OCRs that overlap with sequences conserved between the mouse and the alligator (**A**) and the bamboo shark (**B**). Error bars, SEM. (**C, D**) The fraction of conserved OCRs sorted by the identified clusters in Figure 6A. Sequence conservation was estimated by pairwise alignment using LAST (**A–D**).

**Figure 6–figure supplement 1. Clustering analyses of ATAC-seq peaks with different replicates.** DDifferent replicates were used for the same analysis as shown in Figure 6A. The number after the stage name indicates the replication number. Note that the clustering analyses with different replicates identified clusters similar to those in Figure 6A (compare the left-most panel with the second and third panels from the left). Including replicates with a low-quality score resulted in a relatively small fraction of early stage–specific peaks and a large fraction of late stage–specific peaks (right-most panel).

**Figure 6–figure supplement 2. *De novo* motif discoveries of ATAC-seq peaks.** The top five motifs from each cluster. See Supplementary data for the full list of motifs. C1–C8 correspond to the clusters in Figure 6A.

**Figure 6–figure supplement 3. Analysis of enrichment for known motifs in ATAC-seq peaks.** The top five motifs from clusters C5 and C6 determined while using all other peaks as the background sequence.

**Figure 6–figure supplement 4. *De novo* motif discoveries and known motif enrichment analysis of ATAC-seq peaks with an alternative background.** The top five motifs from clusters C5 and C6 determined while using all other peaks as the background sequence.

**Figure 6–figure supplement 5. Counts of accessible motifs at each stage.** The average number of top-ranked motifs identified by *de novo* motif discovery in Figure 6–figure supplement 2 are plotted against mouse embryonic stages. The average numbers were calculated using all three replicates of ATAC-seq peaks at each stage. Rows indicate clusters identified in Figure 6A; columns indicate motif rank. Error bars, SEM. C1–C8 correspond to the clusters in Figure **[6]**A. Note that the number of CTCF motifs (top-ranked in C5) was relatively stable over time, which is consistent with the clustering analysis shown in Figure 6A. In addition, the number of motifs enriched for C3, such as BHLHA15, HOX13, TEAD, and Tlx?, increased over time. In contrast, COUP-TFII and TCF7L2 motifs decreased over time. Interestingly, LHX and HOX9 motifs were transiently increased at E10.5.

**Supplementary file 1. Summary of short-read sequencing data.**

**Supplementary file 2. Orthology asignment for the transcriptome of the brown-banded bamboo shark.** Column 1–4: transcriptome assembly ID, NCBI gene ID, gene symbol, blast score.

**Supplementary file 3. Orthology asignment for the gene model of the brown-banded bamboo shark.** Column 1–4: gene model ID, NCBI gene ID, gene symbol, blast score.

**Supplementary file 4. Quality control of orthology assignment.** Source data to create Figure 1C**.**

**Supplementary file 5. The mean and SEM of TPM values of mouse limb RNA-seq data.** Source data for Figure 1D and other plots related to gene expression amount.

**Supplementary file 6. The mean and SEM of TPM values of bamboo shark fin RNA-seq data.** Source data for Figure 1D and other plots related to gene expression amount.

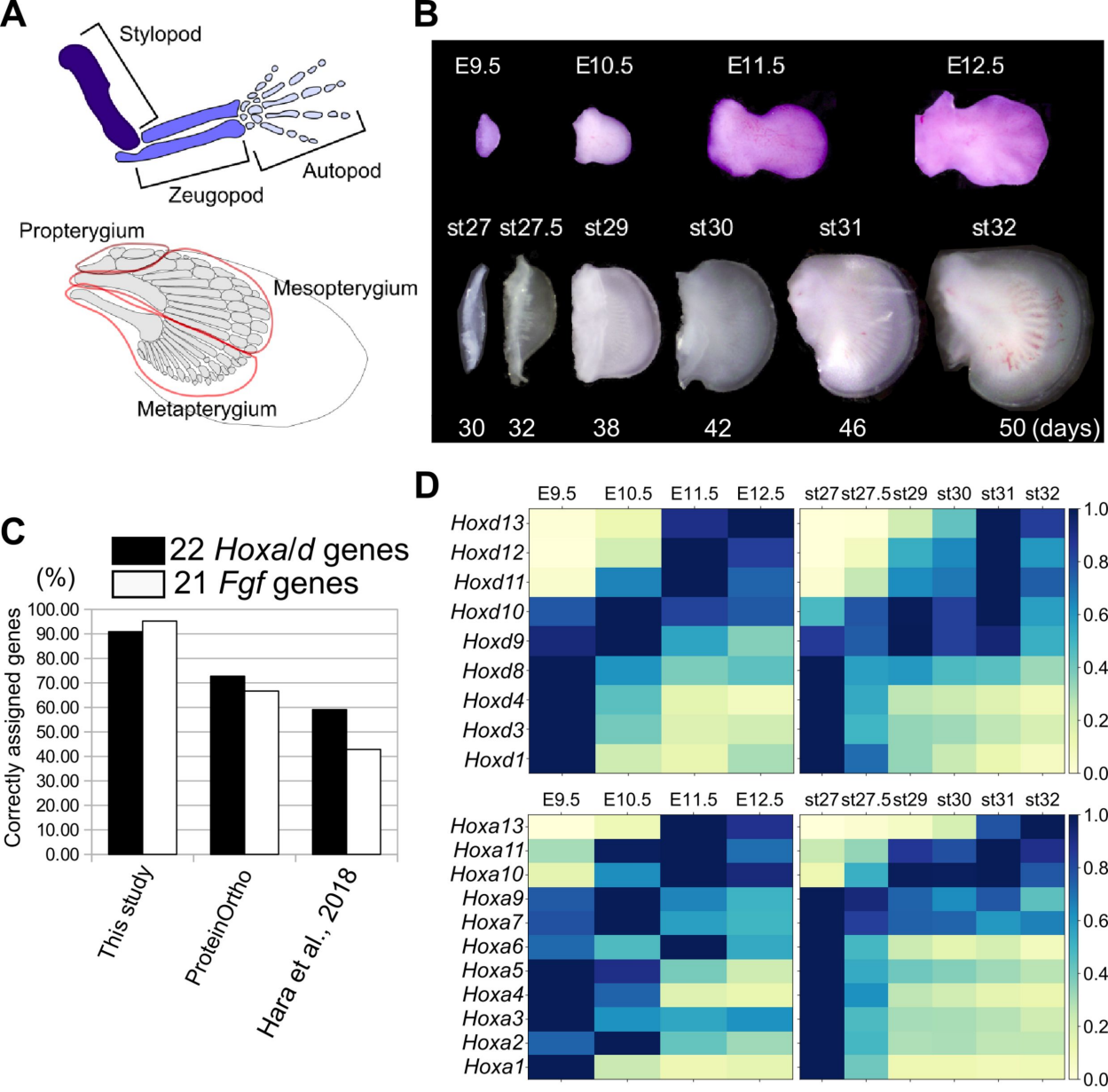
**Supplementary file 7. Clustered gene expression table with phenotype annotation.** The details of Figure 2A.

**Supplementary file 8. The list of genes downregulated over time in mouse limb buds being upregulated in bamboo shark fin buds over time (related to Figure 2-figure supplement 1B).**

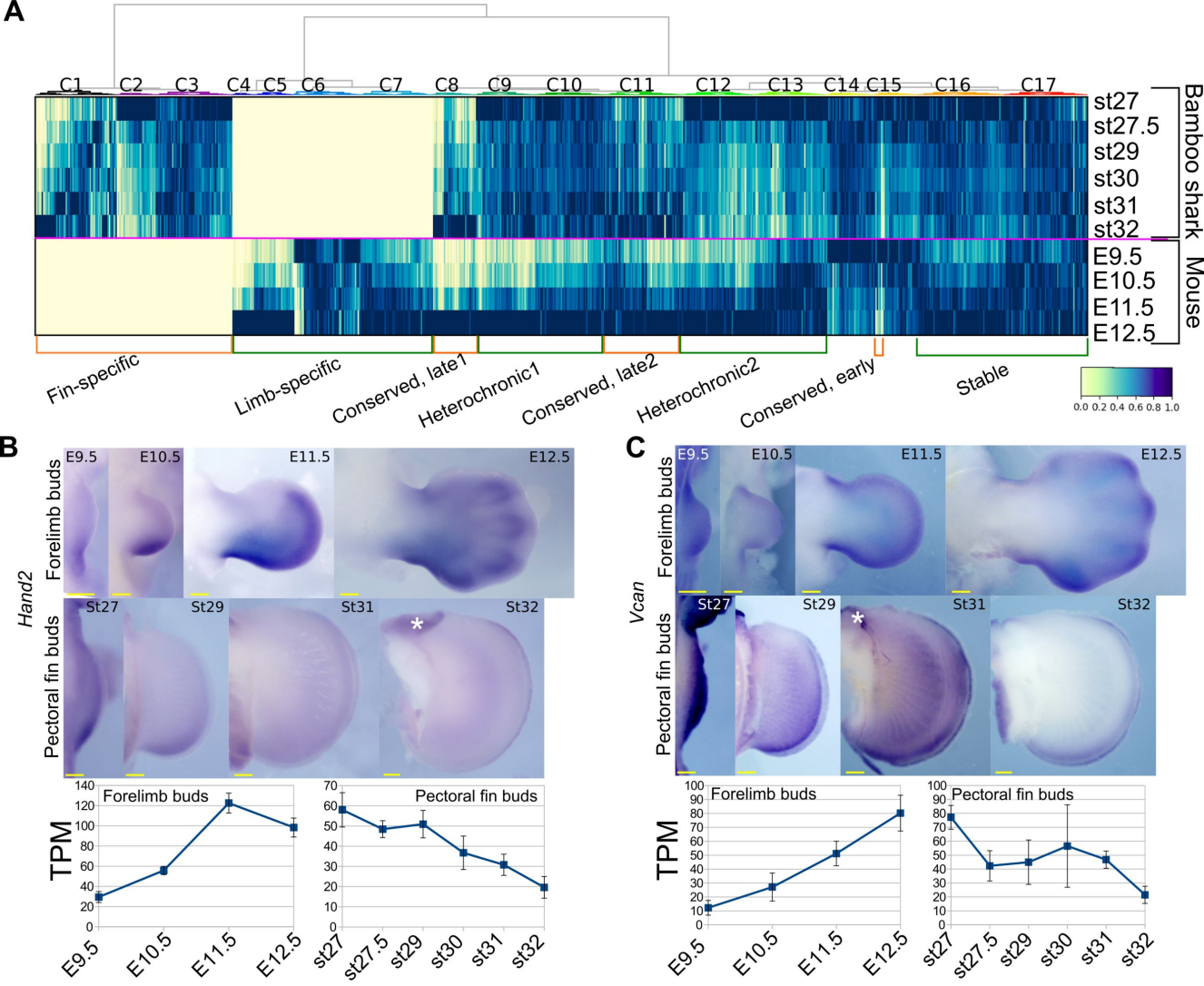
**Supplementary file 9. PCA loadings of Figure 4D and E.**

**Supplementary file 10. List of public data used in Figures. 4, 5 and 6.**

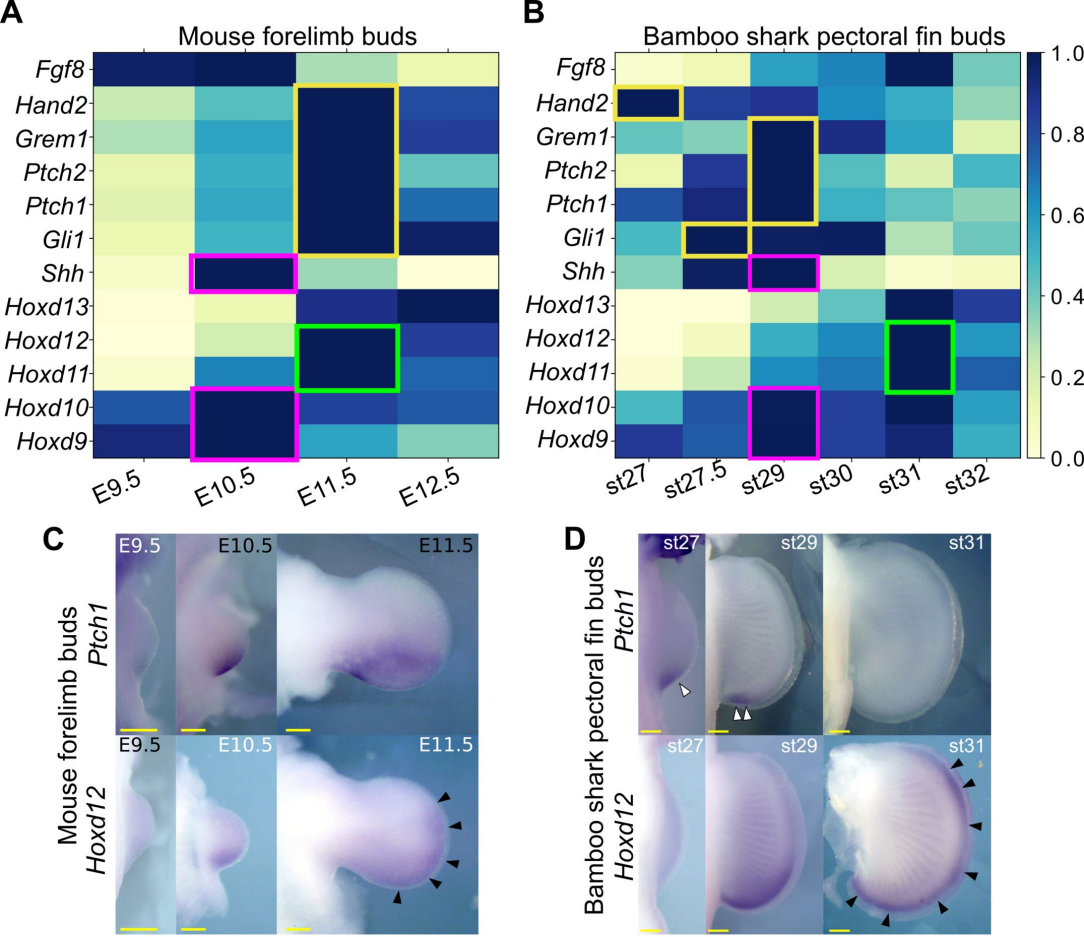
**Supplementary file 11. GO analysis of ATAC-seq peaks. c1 to c8 correspond to the clusters in Figure 5A.**



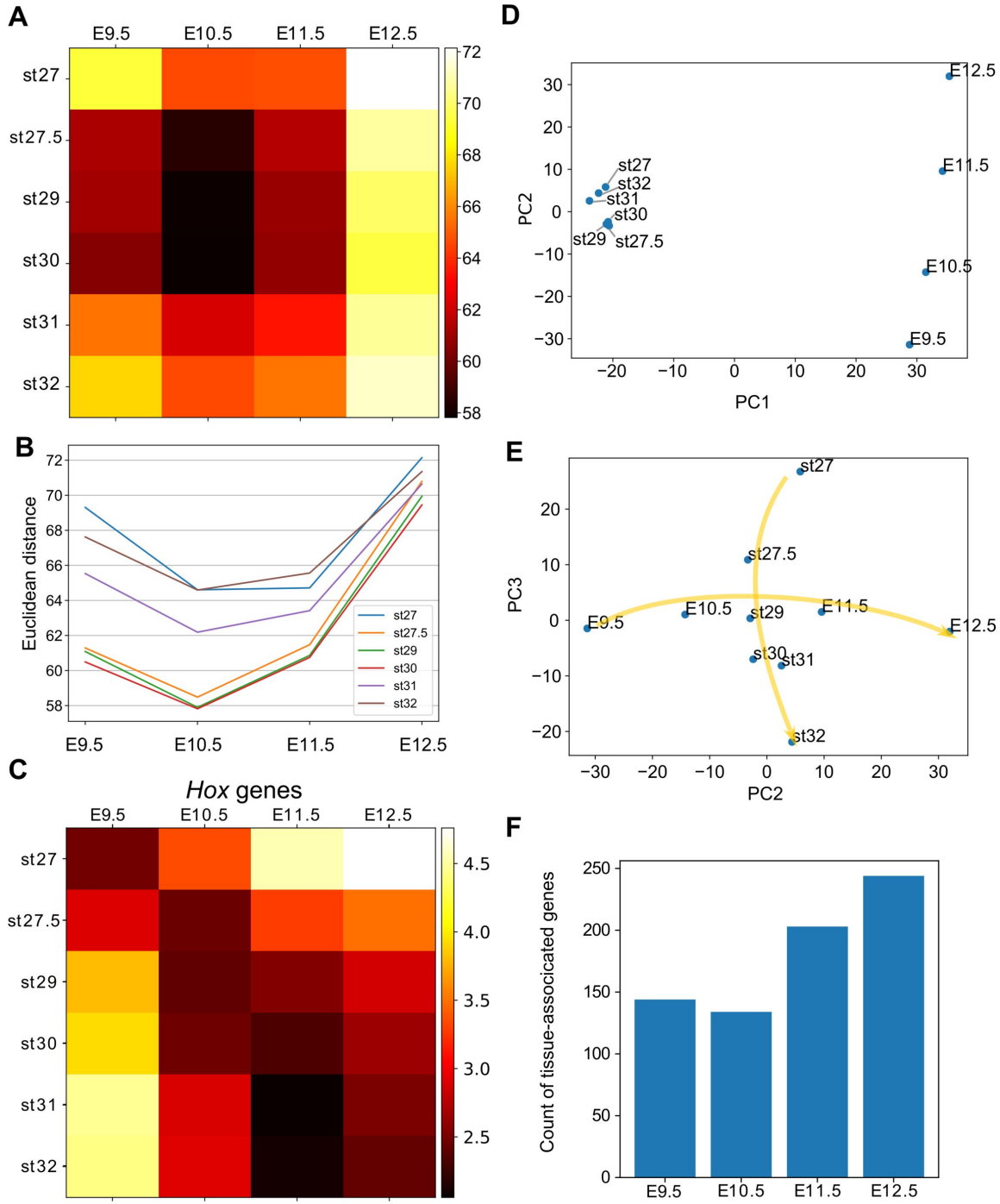

Onimaru et al., Fig.1



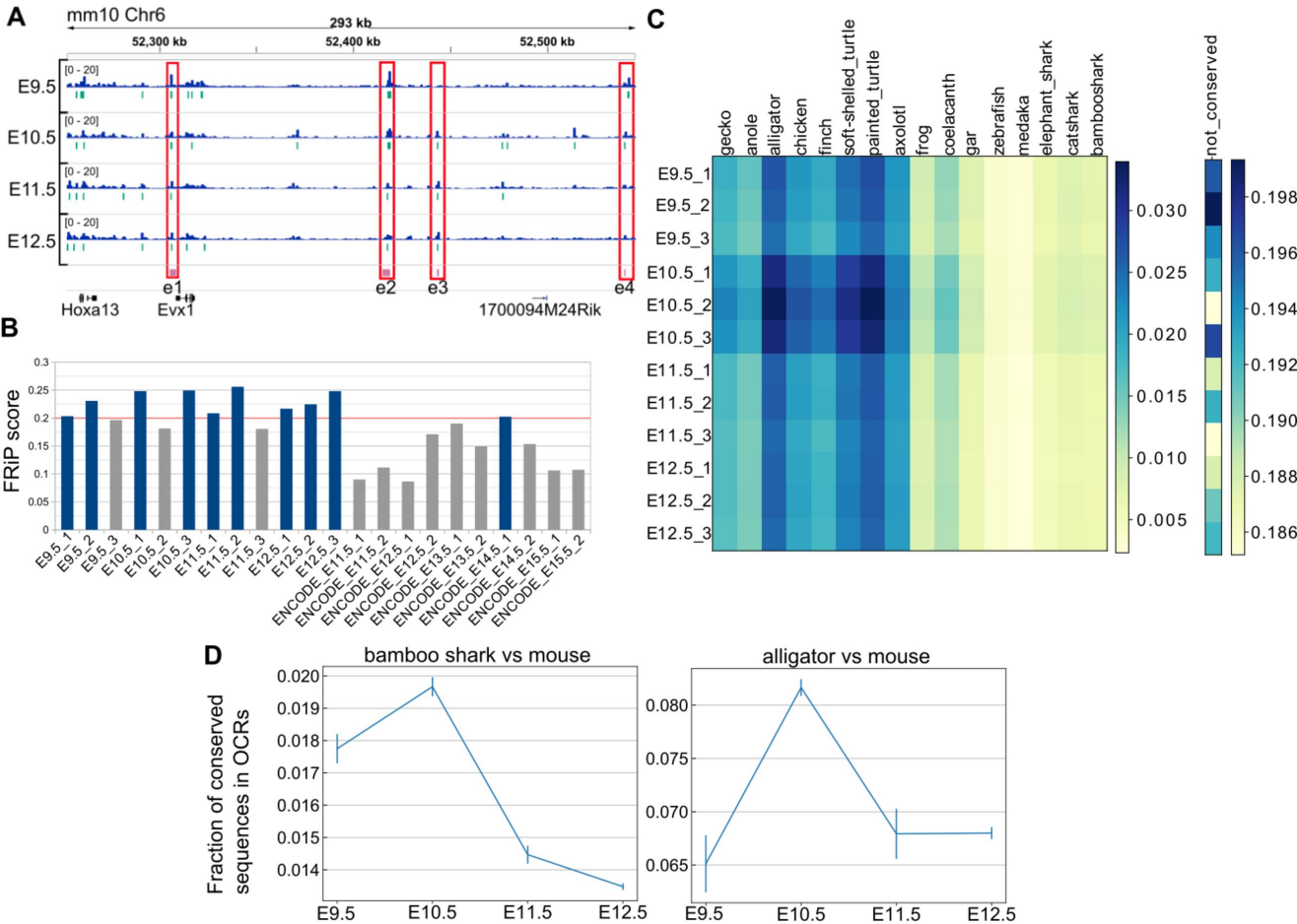

Onimaru et al Figure 2



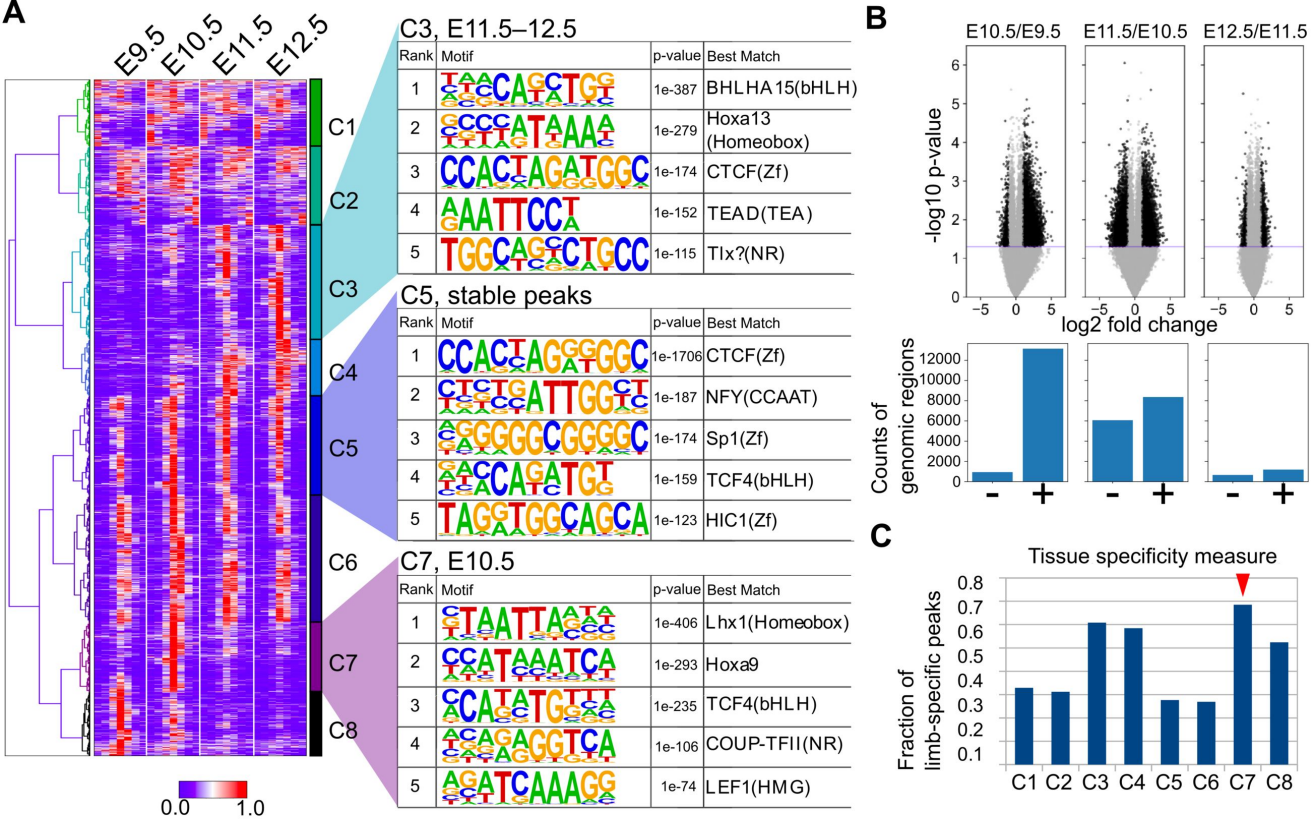

Onimaru et al Figure 3



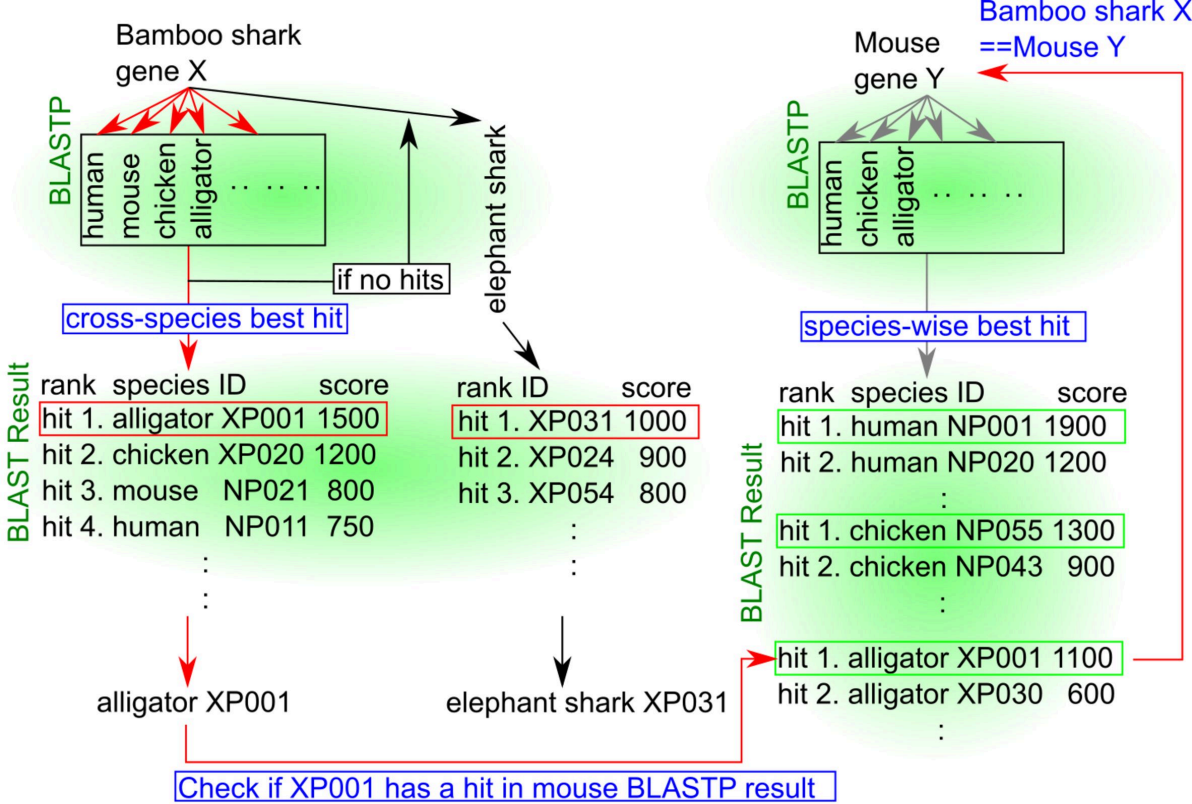

Onimaru et al Figure 4



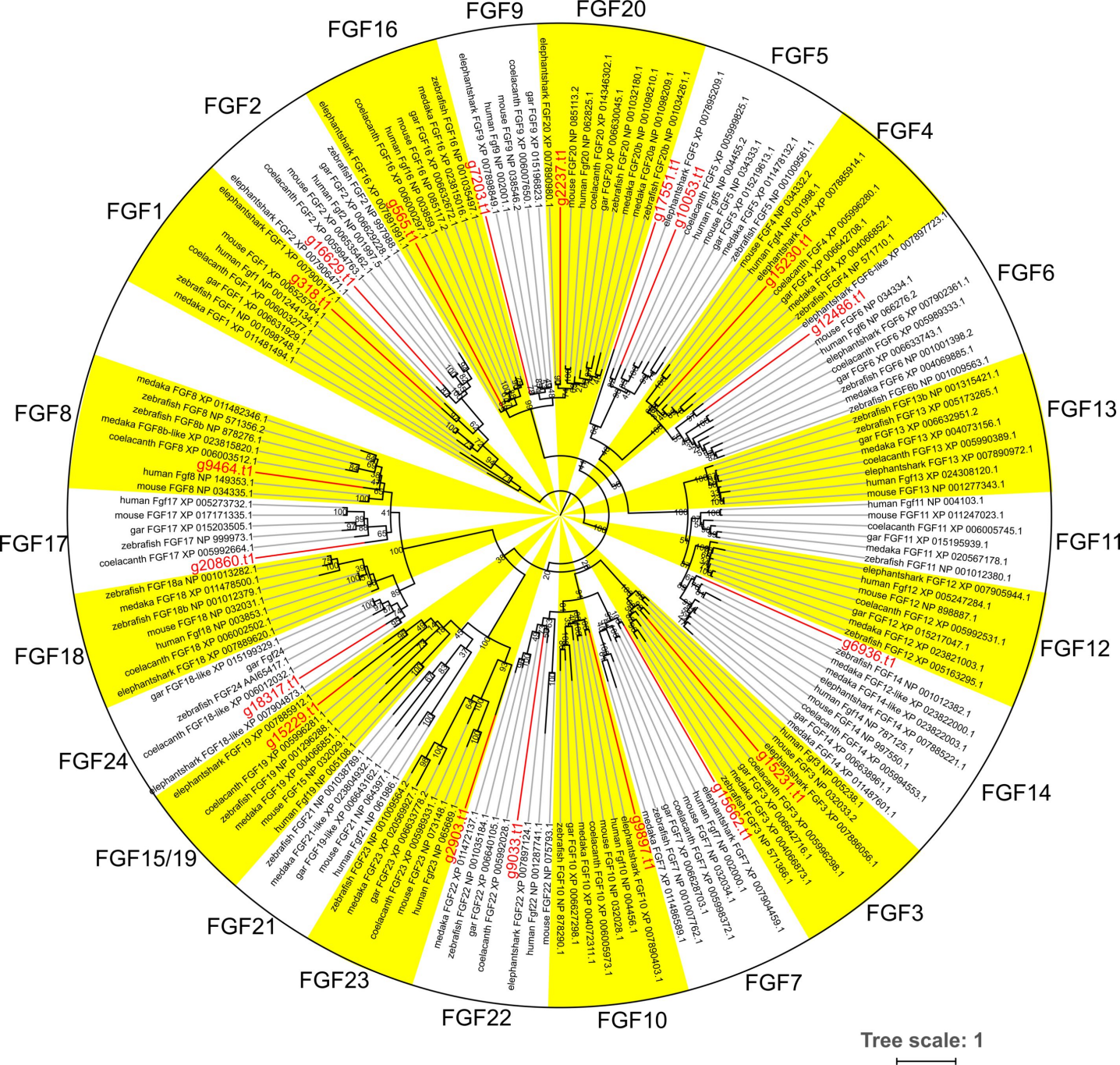

Onimaru et al Figure 5

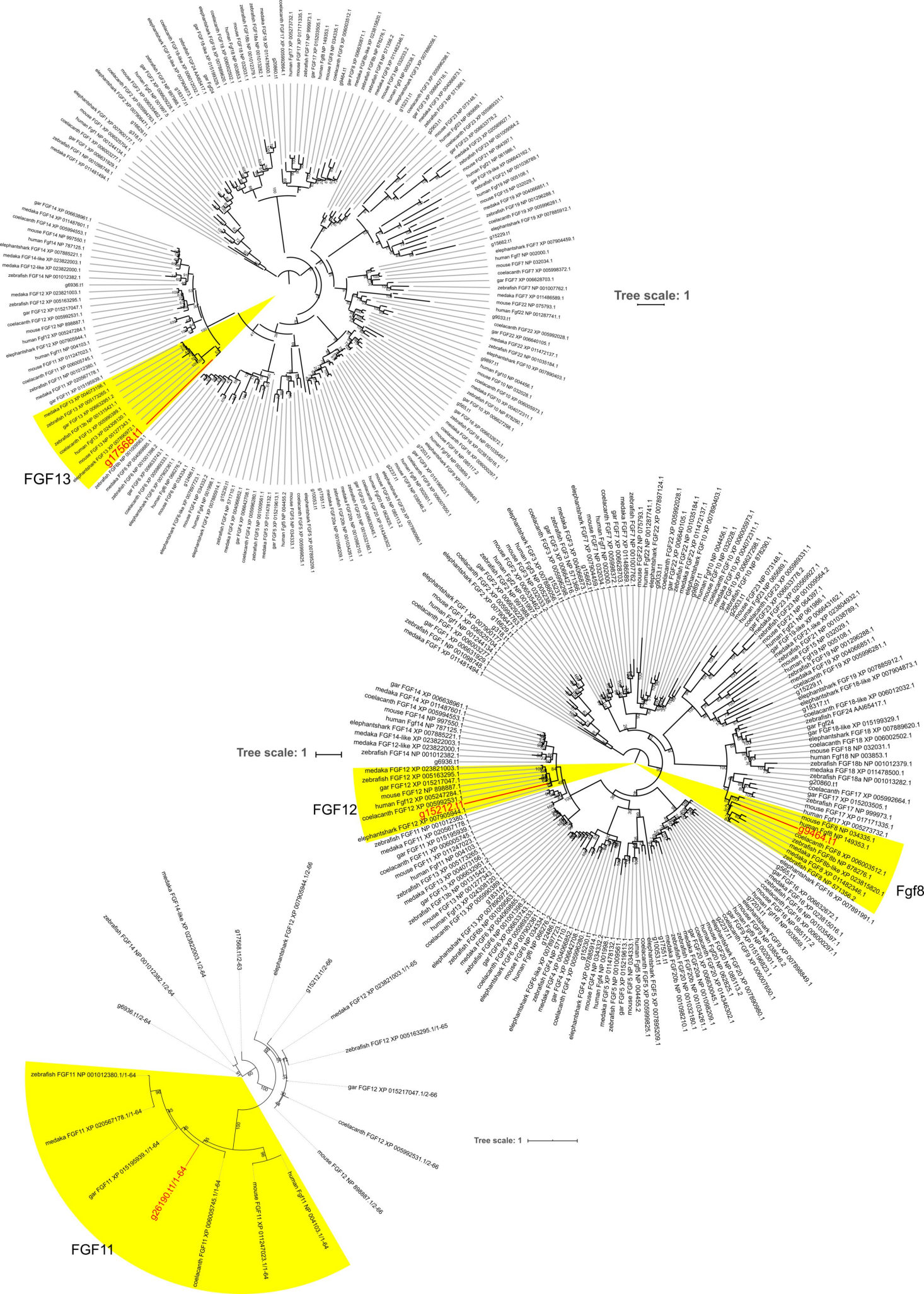


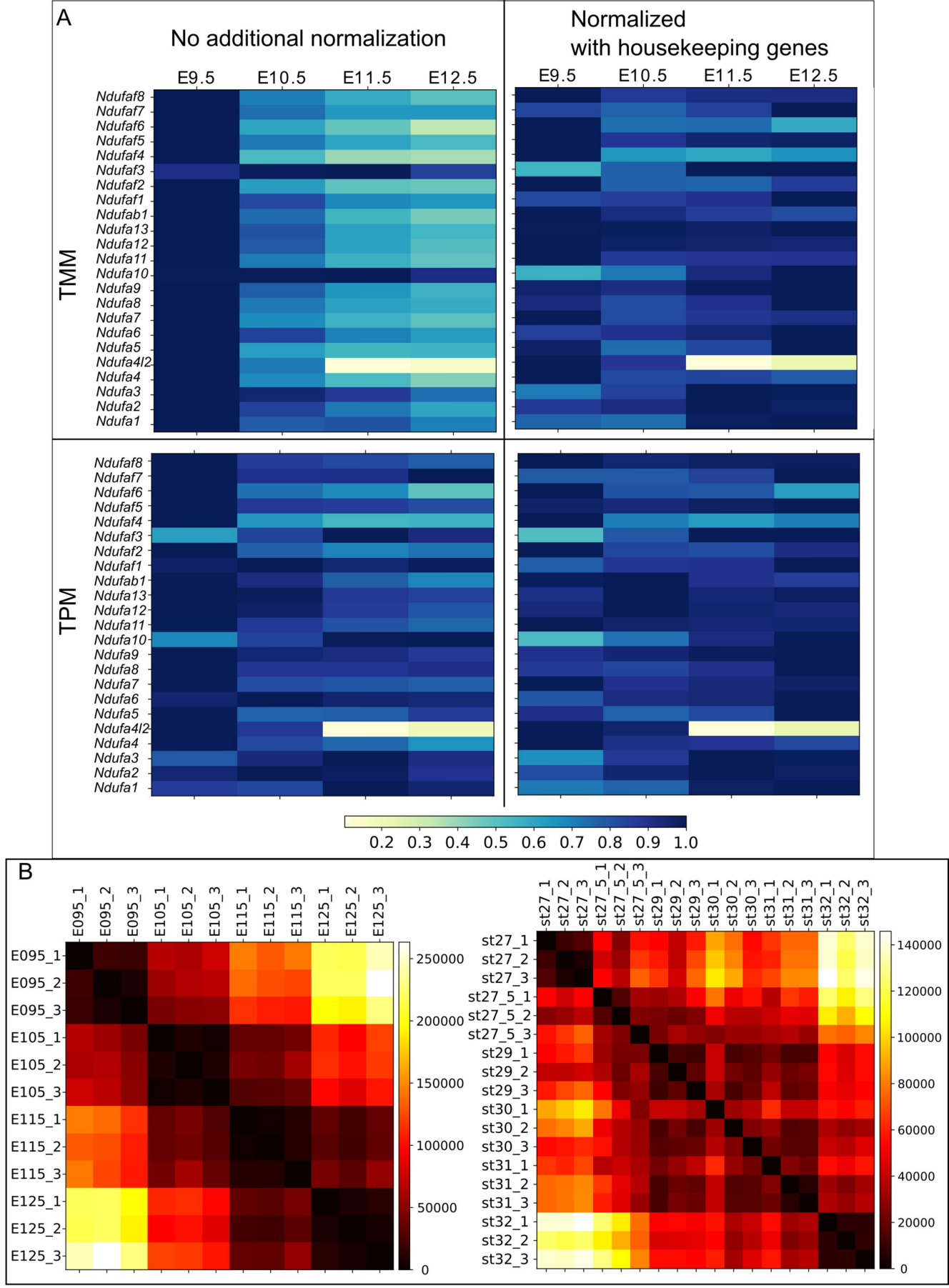

Onimaru et al Figure 6



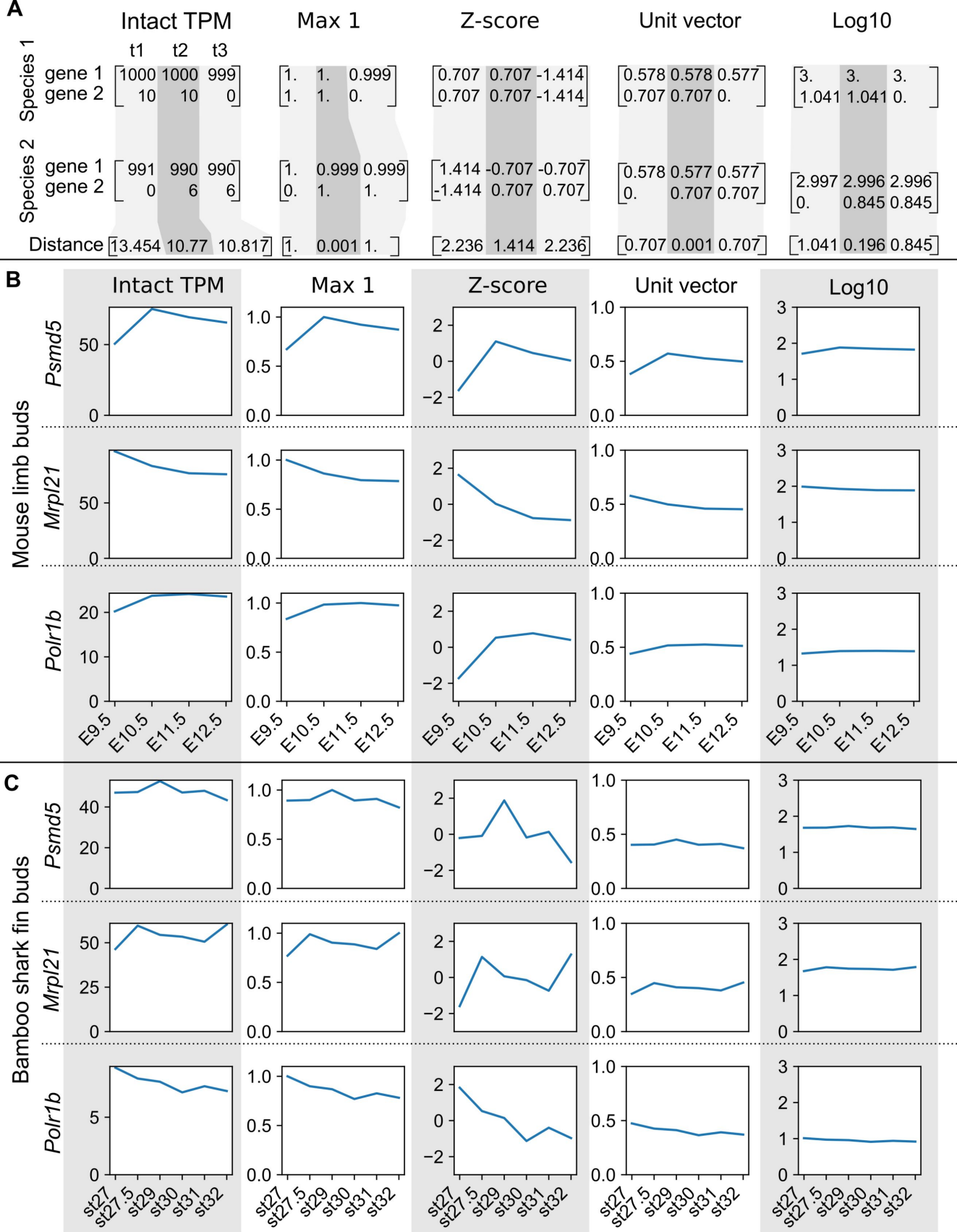

Onimaru et al Figure 1-figure supplement 1


FGF9
FGF20
FGF16
FGF5
FGF2
FGF4
FGF1
FGF6
FGF8
FGF13
FGF17
FGF11
FGF18
FGF12
FGF24
FGF14
FGF15/19
FGF3
FGF21
FGF7
FGF23
FGF22
FGF10
Tree scale: 1

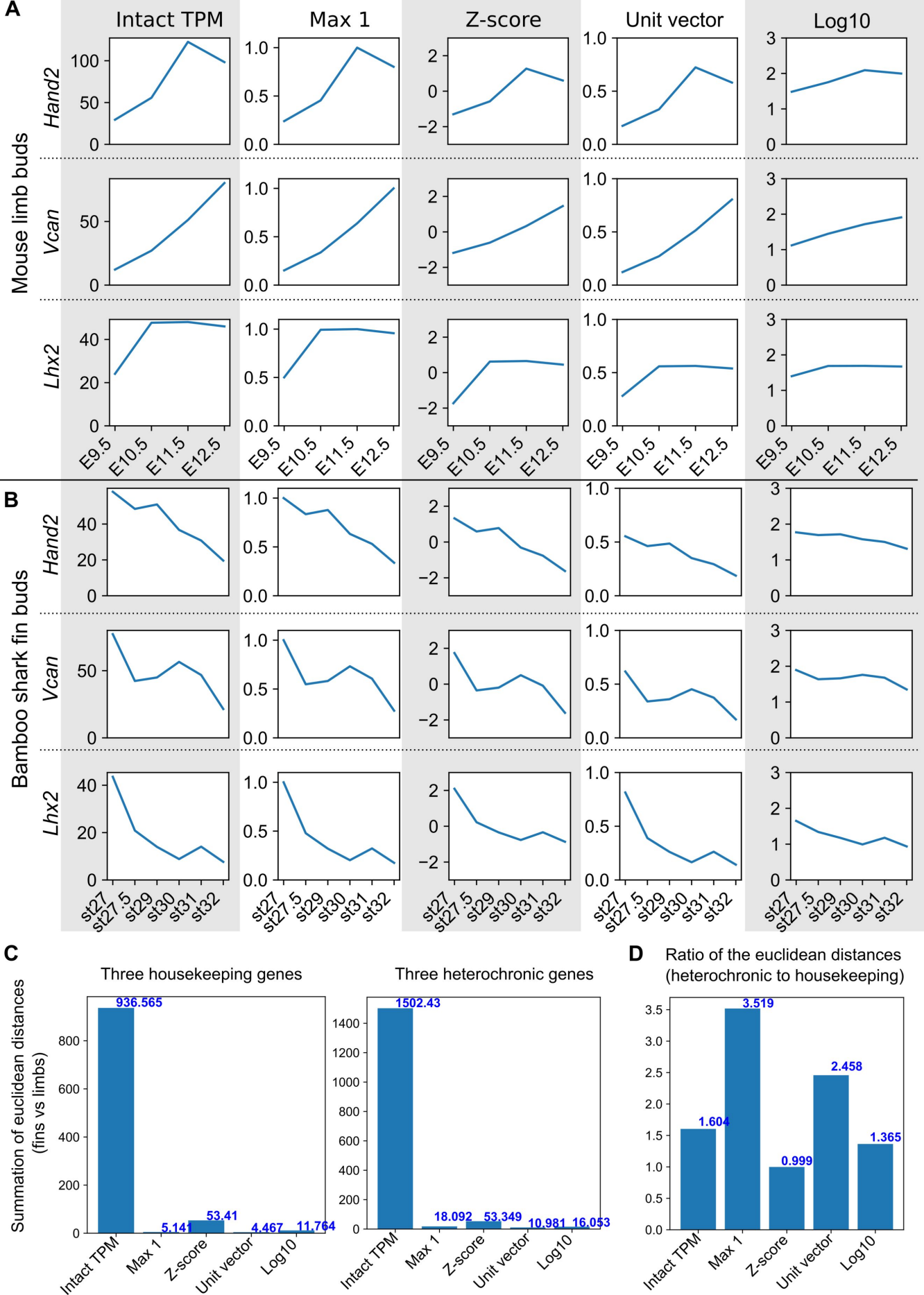

FGF13
g17568.t1
Tree scale: 1
FGF12
g15217214
Tree scale: 1
Fgf8
g94641.t1
FGF11
g26190.t1/1-64
Tree scale: 1

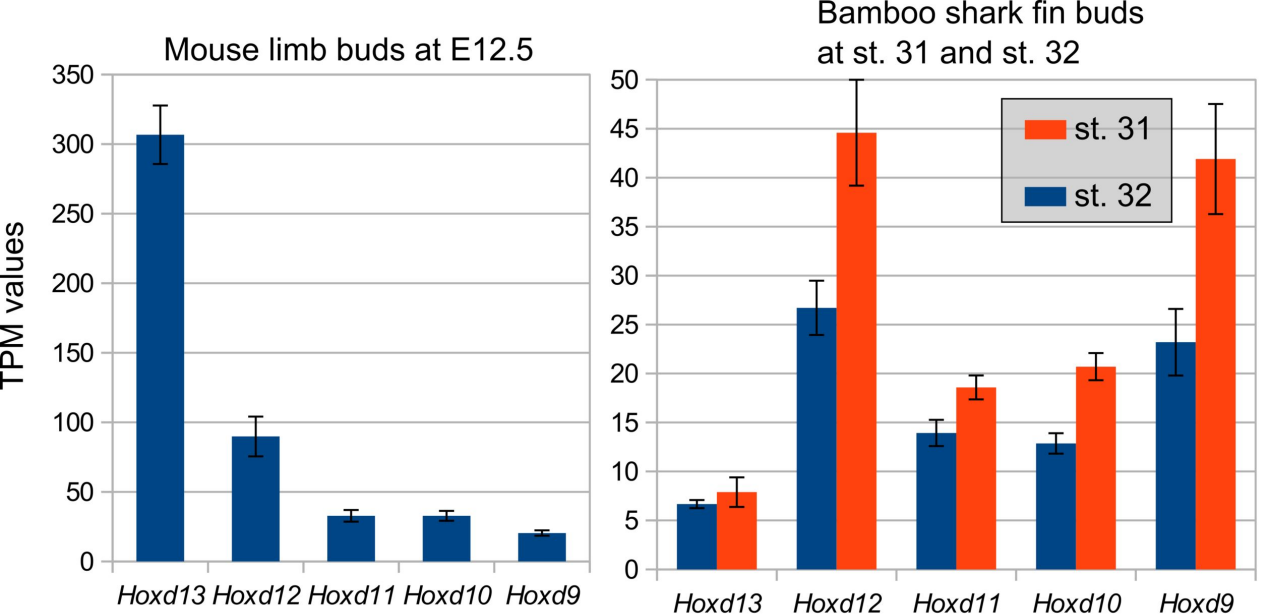


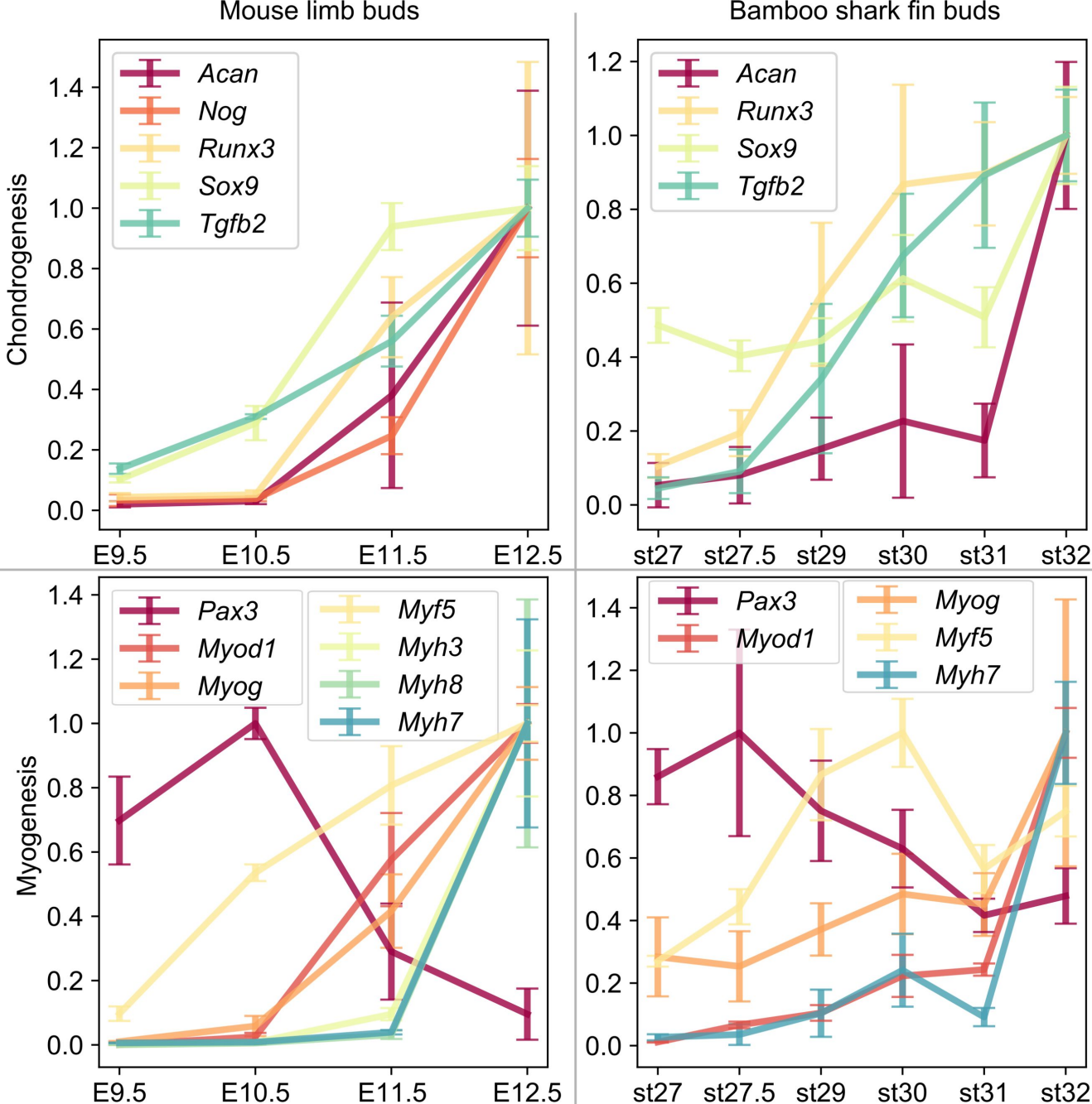

Onimaru et al Figure 1-figure supplement 4

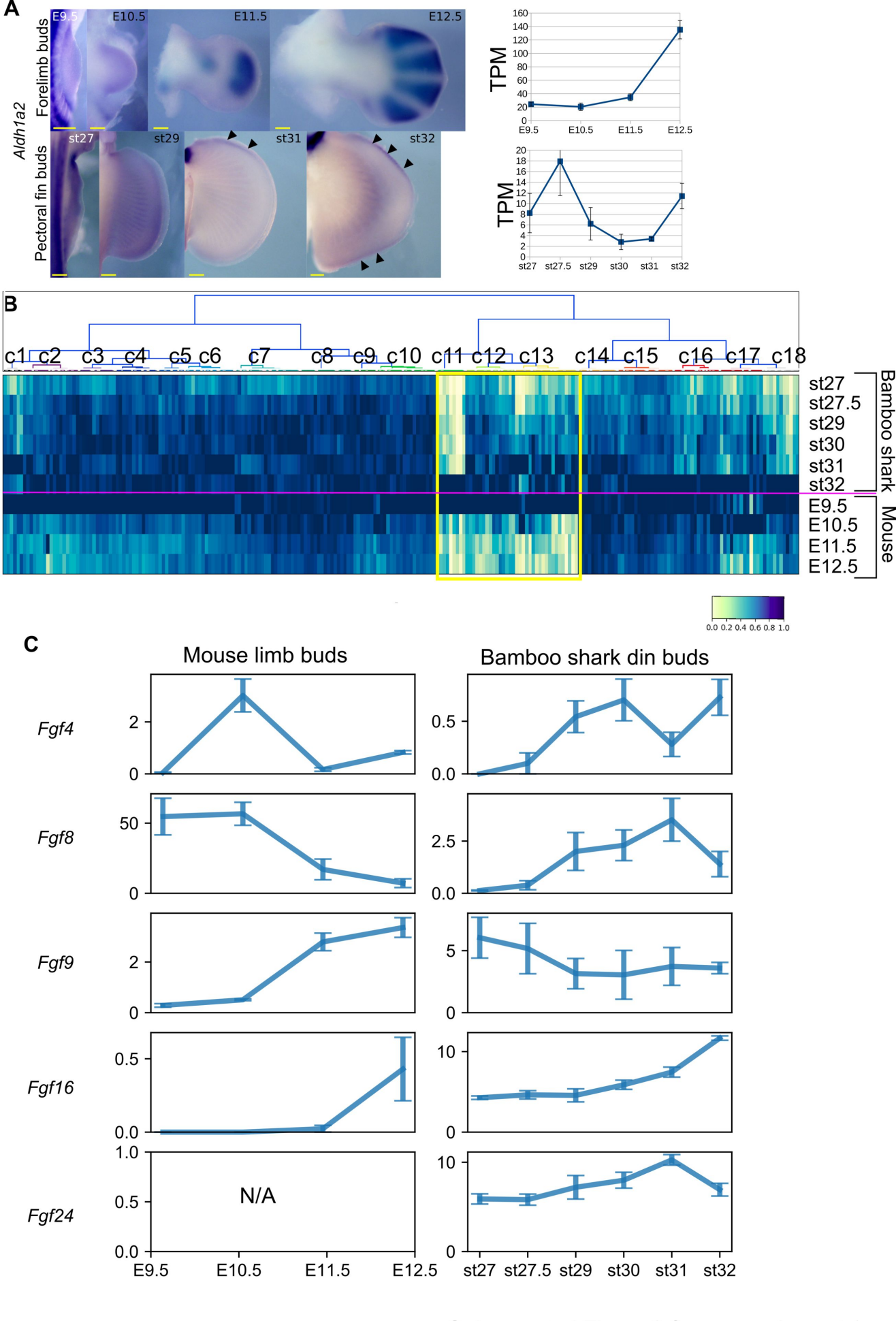



Onimaru et al Figure 1-figure supplement 5

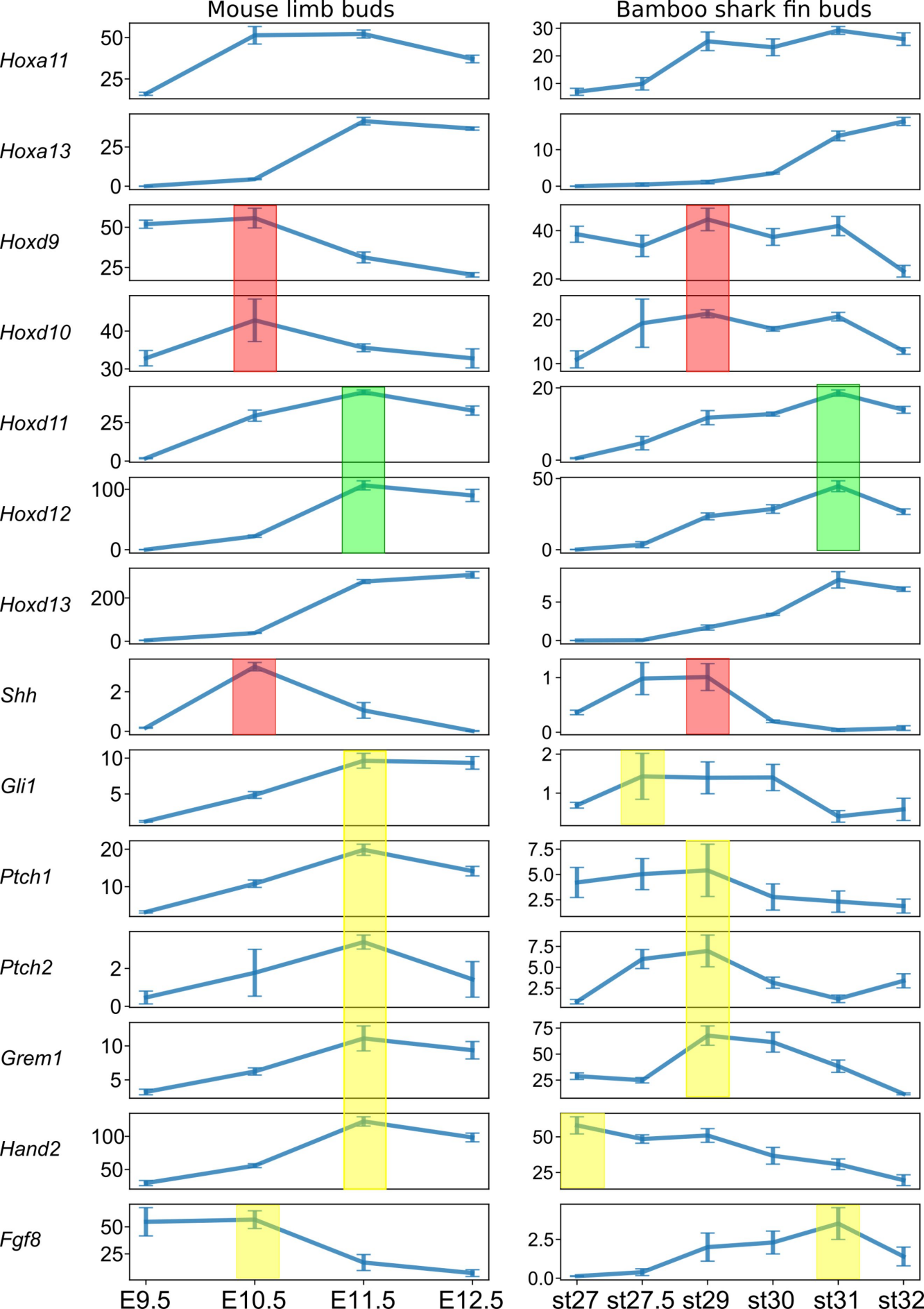

A
Intact TPM
Max 1
Z-score
Unit vector
Log10
Mouse limb buds
Hand2
Vcan
Lhx2
E9.5
E10.5
E11.5
E12.5
B
Bamboo shark fin buds
st27
st27.5
st29
st30
st31
st32
C
Three housekeeping genes
Three heterochronic genes
Summation of euclidean distances
(fins vs limbs)
936.565
5.141
53.41
4.467
11.764
1502.43
18.092
53.349
10.981
16.053
D
Ratio of the euclidean distances
(heterochronic to housekeeping)
1.604
3.519
0.999
2.458
1.365

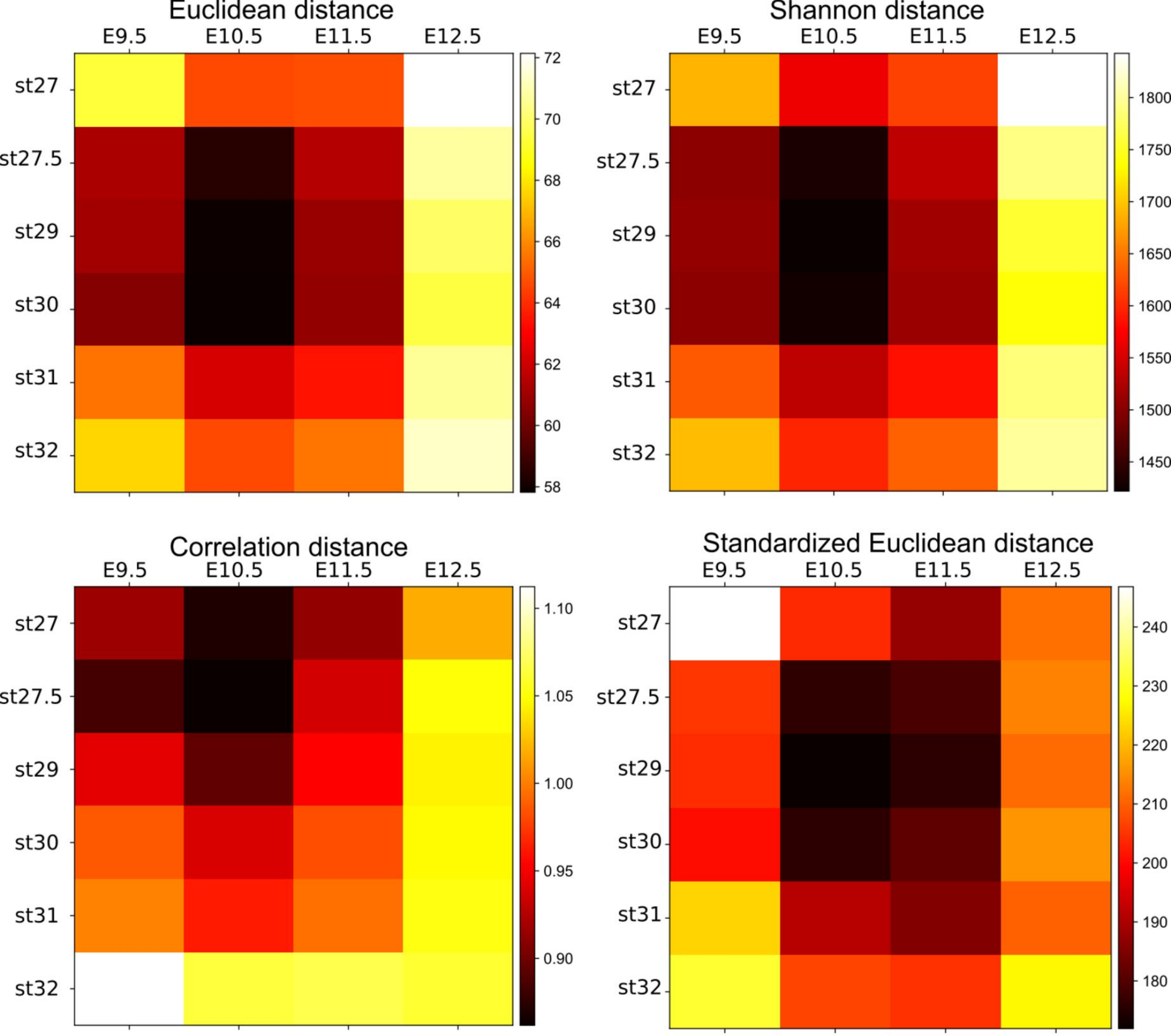

Mouse limb buds at E12.5
TPM values
350
300
250
200
150
100
50
0
Hoxd13
Hoxd12
Hoxd11
Hoxd10
Hoxd9
Bamboo shark fin buds
at st. 31 and st. 32
st. 31
st. 32
50
45
40
35
30
25
20
15
10
5
0
Hoxd13
Hoxd12
Hoxd11
Hoxd10
Hoxd9


Mouse limb buds
Bamboo shark fin buds
Chondrogenesis
Myogenesis
Acan
Nog
Runx3
Sox9
Tgfb2
Pax3
Myod1
Myog
Myf5
Myh3
Myh8
Myh7
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
E9.5
E10.5
E11.5
E12.5
st27
st27.5
st29
st30
st31
st32

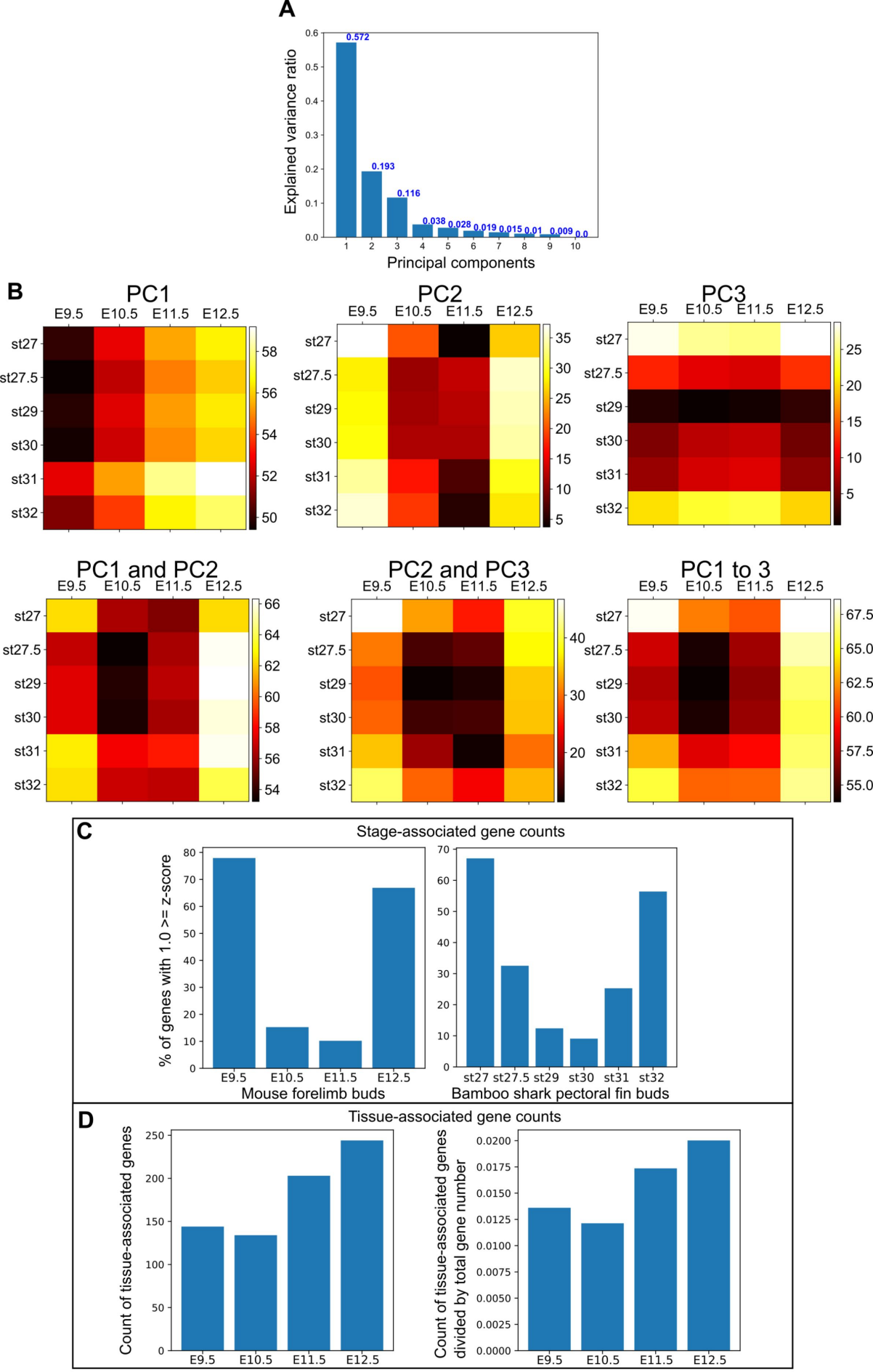



Onimaru et al Figure 2-figure supplement 1

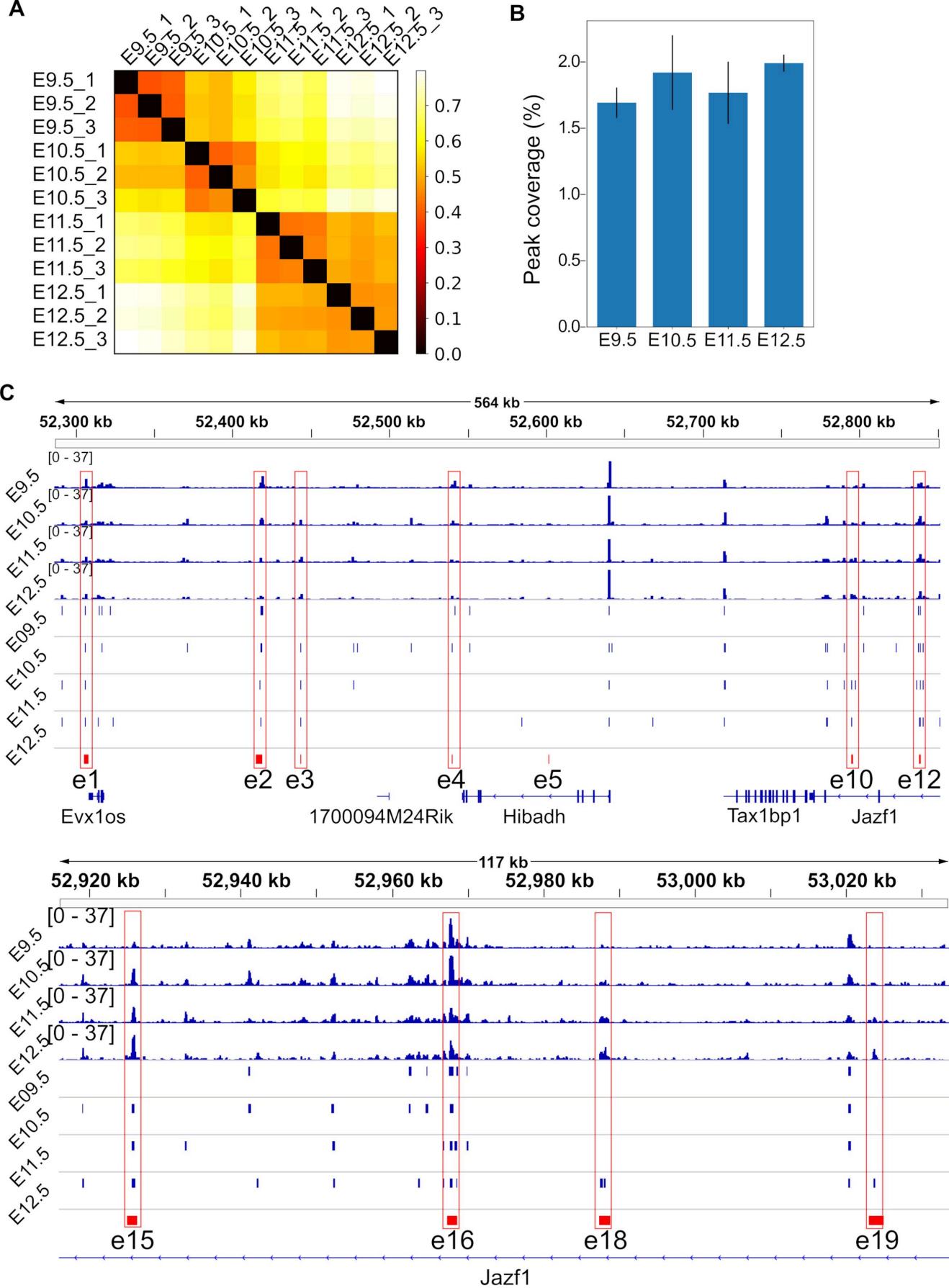

Mouse limb buds
Bamboo shark fin buds
Hoxa11
Hoxa13
Hoxd9
Hoxd10
Hoxd11
Hoxd12
Hoxd13
Shh
Gli1
Ptch1
Ptch2
Grem1
Hand2
Fgf8
E9.5
E10.5
E11.5
E12.5
st27
st27.5
st29
st30
st31
st32



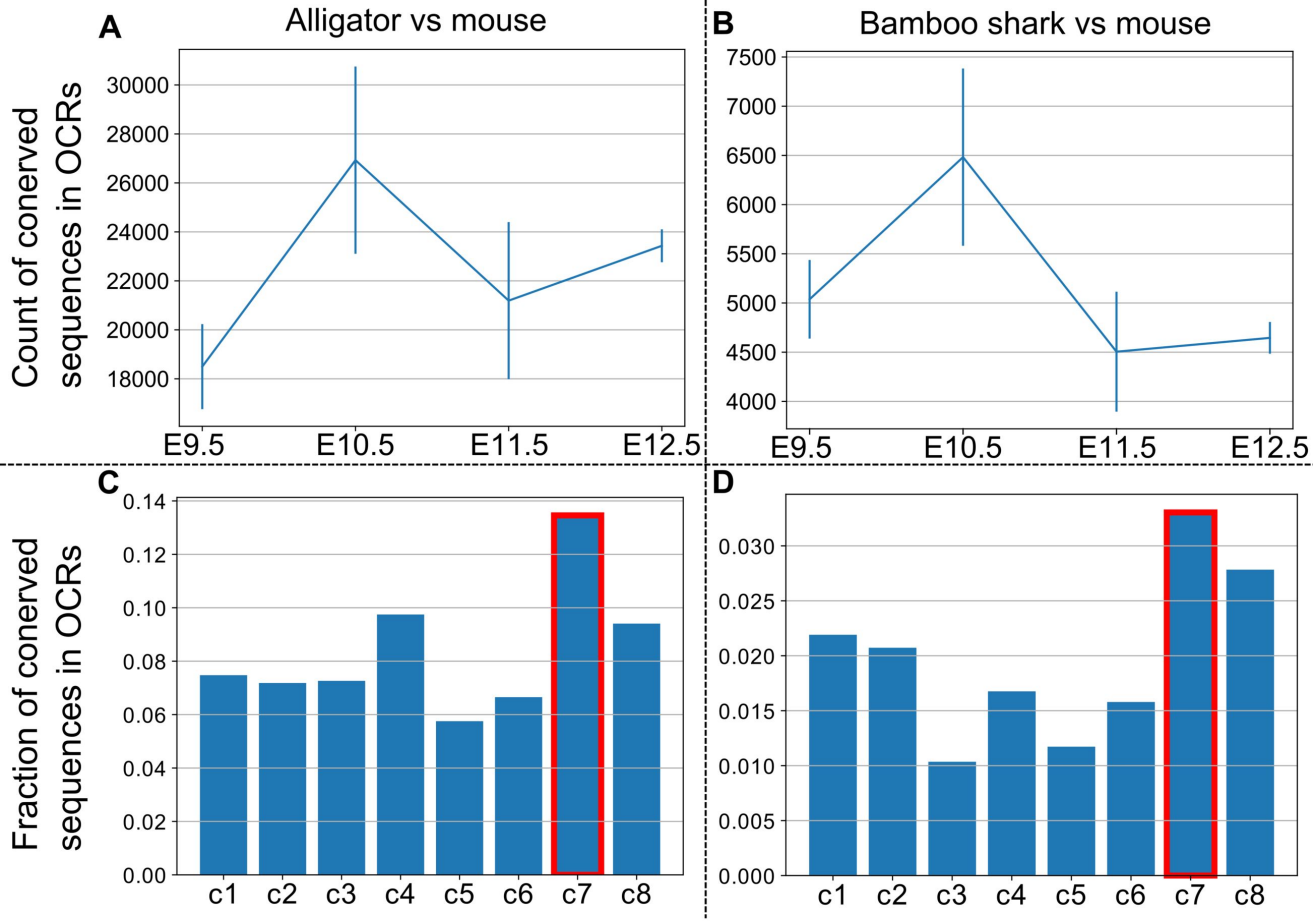

Onimaru et al Figure 4-figure supplement 1



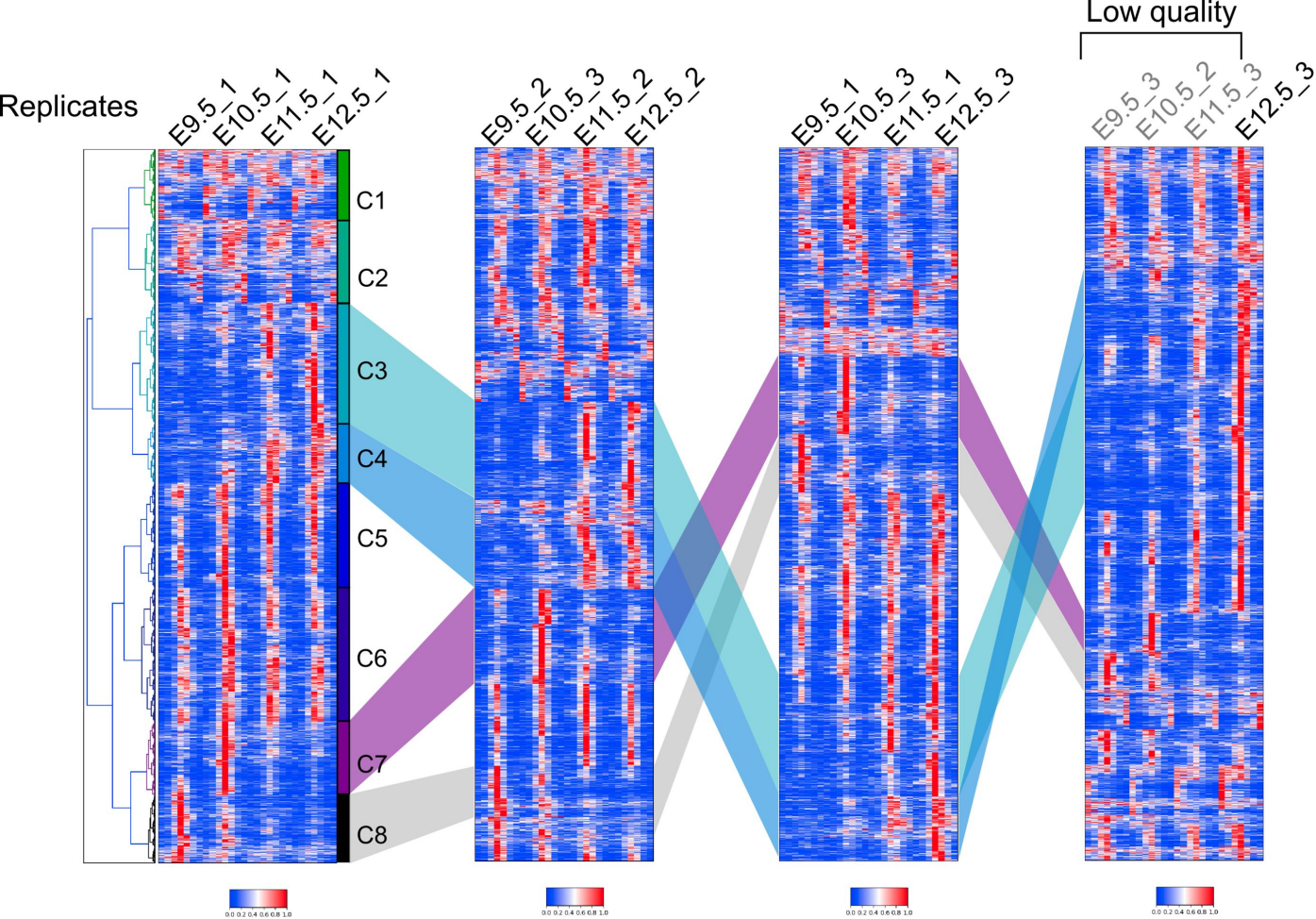

Onimaru et al Figure 4-figure supplement 2




Onimaru et al Figure 5 figure supplement 1




Onimaru et al Figure 5-figure supplement 2




(from Figure 5A)

Onimaru et al Figure 6-figure supplement 1

C1

| Rank | Motif | P-value | log P-pvalue | % of Targets | % of Background | STD(Bg STD) | Best Match/Details |
|---|---|---|---|---|---|---|---|
| 1 | AACCATATG | 1e-37 | -8.675e+01 | 13.69% | 9.79% | 27.3bp (27.3bp) | TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer(0.938) |
| 2 | TAATTAGATC | 1e-36 | -8.399e+01 | 10.68% | 7.29% | 26.6bp (28.3bp) | Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer(0.876) |
| 3 | CCACTAGGTGGC | 1e-32 | -7.379e+01 | 2.01% | 0.79% | 25.6bp (26.2bp) | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer(0.896) |
| 4 | TGTAAACG | 1e-31 | -7.351e+01 | 24.98% | 20.26% | 27.1bp (29.4bp) | Foxo3(Forkhead)/U2OS-Foxo3-ChIP-Seq(E-MTAB-2701)/Homer(0.708) |
| 5 | GCATTCCG | 1e-29 | -6.802e+01 | 5.67% | 3.48% | 27.1bp (27.5bp) | TEAD3/MA0808.1/Jaspar(0.912) |

C2

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | CAATTAAT | 1e-61 | -1.411e+02 | 22.84% | 17.07% | 27.3bp (30.0bp) | NKX6-2/MA0675.1/Jaspar(0.916) |
| 2 | CCACCAGGGGGC | 1e-53 | -1.237e+02 | 3.19% | 1.33% | 26.5bp (28.0bp) | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer(0.935) |
| 3 | TACCATCTGG | 1e-37 | -8.610e+01 | 7.12% | 4.55% | 26.5bp (26.4bp) | BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer(0.971) |
| 4 | TATAGCGC | 1e-26 | -6.036e+01 | 16.80% | 13.45% | 28.0bp (28.6bp) | PB0008.1_E2F2_1/Jaspar(0.659) |
| 5 | CTTGAGTGGATA | 1e-26 | -6.006e+01 | 0.15% | 0.00% | 24.6bp (0.0bp) | Nkx2-5(var.2)/MA0503.1/Jaspar(0.834) |

C3

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | TAACCAGCTGG | 1e-387 | -8.926e+02 | 18.08% | 8.10% | 26.5bp (32.6bp) | BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer(0.980) |
| 2 | GCCCCATAAAA | 1e-279 | -6.443e+02 | 22.91% | 12.98% | 26.2bp (30.6bp) | Hoxa13(Homeobox)/ChickenMSG-Hoxa13.Flag-ChIP-Seq(GSE86088)/Homer(0.981) |
| 3 | CCACTAGATGGC | 1e-174 | -4.016e+02 | 1.84% | 0.22% | 25.6bp (28.2bp) | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.903) |
| 4 | AAATTCCT | 1e-152 | -3.507e+02 | 8.03% | 3.70% | 26.3bp (30.7bp) | TEAD(TEA)/Fibroblast-PU.1-ChIP-Seq(Unpublished)/Homer(0.911) |
| 5 | TGGCAGCCTGCC | 1e-115 | -2.656e+02 | 5.30% | 2.28% | 26.0bp (31.2bp) | Tlx?(NR)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer(0.933) |

C4

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | CACAGCTGGTC | 1e-115 | -2.650e+02 | 17.42% | 9.58% | 26.4bp (31.1bp) | BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer(0.973) |
| 2 | CCTATAAA | 1e-86 | -1.992e+02 | 12.26% | 6.51% | 27.1bp (31.1bp) | CDX4(Homeobox)/ZebrafishEmbryos-Cdx4.Myc-ChIP-Seq(GSE48254)/Homer(0.950) |
| 3 | GAAATCAAAT | 1e-56 | -1.293e+02 | 28.15% | 21.04% | 27.0bp (29.2bp) | Six2(Homeobox)/NephronProgenitor-Six2-ChIP-Seq(GSE39837)/Homer(0.758) |
| 4 | CTGTGGAATGCA | 1e-38 | -8.898e+01 | 49.84% | 42.94% | 26.0bp (28.9bp) | TEAD3(TEA)/HepG2-TEAD3-ChIP-Seq(Encode)/Homer(0.816) |
| 5 | CAAGTCATTA | 1e-24 | -5.717e+01 | 0.76% | 0.16% | 24.7bp (25.8bp) | MEOX2/MA0706.1/Jaspar(0.715) |

C5

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | CCACTAGGGGGC | 1e-1706 | -3.930e+03 | 13.25% | 0.85% | 24.5bp (28.0bp) | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.921) |
| 2 | CTCTGATTGGCT | 1e-187 | -4.322e+02 | 5.11% | 1.51% | 25.4bp (28.9bp) | NFY(CCAAT)/Promoter/Homer(0.943) |
| 3 | AGGGGGCGGGGC | 1e-174 | -4.014e+02 | 10.60% | 5.02% | 25.6bp (27.8bp) | Sp1(Zf)/Promoter/Homer(0.984) |
| 4 | GAACCAGATGT | 1e-159 | -3.673e+02 | 10.18% | 4.92% | 26.1bp (27.8bp) | TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer(0.969) |
| 5 | TAGGTGGCAGCA | 1e-123 | -2.835e+02 | 0.67% | 0.02% | 30.5bp (5.4bp) | HIC1(Zf)/Treg-ZBTB29-ChIP-Seq(GSE99889)/Homer(0.650) |

C6

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | CCACTAGGGGGC | 1e-544 | -1.254e+03 | 5.73% | 0.87% | 24.7bp (26.4bp) | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.919) |
| 2 | GGGGCGGGGC | 1e-170 | -3.927e+02 | 12.29% | 6.88% | 26.0bp (28.4bp) | Sp2(Zf)/HEK293-Sp2.eGFP-ChIP-Seq(Encode)/Homer(0.928) |
| 3 | ACCAGATG | 1e-92 | -2.121e+02 | 4.78% | 2.34% | 27.1bp (26.4bp) | TWIST1/MA1123.1/Jaspar(0.936) |
| 4 | ACCGGAAGTG | 1e-72 | -1.664e+02 | 4.11% | 2.08% | 26.5bp (27.5bp) | ETV4(ETS)/HepG2-ETV4-ChIP-Seq(ENCODE)/Homer(0.973) |
| 5 | ACTAATTA | 1e-67 | -1.555e+02 | 21.71% | 17.00% | 27.1bp (30.1bp) | Nobox/MA0125.1/Jaspar(0.944) |

C7

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | CTAATTAATA | 1e-406 | -9.353e+02 | 33.78% | 16.63% | 26.1bp (33.4bp) | Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer(0.988) |
| 2 | CCATAAATCA | 1e-293 | -6.767e+02 | 14.30% | 4.95% | 26.5bp (31.9bp) | Hoxa9/MA0594.1/Jaspar(0.960) |
| 3 | CCATATGTTT | 1e-235 | -5.412e+02 | 22.84% | 11.57% | 27.0bp (29.9bp) | TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer(0.952) |
| 4 | TCAGAGGTCA | 1e-106 | -2.461e+02 | 16.44% | 9.61% | 26.2bp (28.7bp) | COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer(0.960) |
| 5 | AGATCAAAGG | 1e-74 | -1.715e+02 | 8.00% | 4.06% | 26.9bp (28.4bp) | LEF1(HMG)/H1-LEF1-ChIP-Seq(GSE64758)/Homer(0.986) |

C8

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | ACTAATTA | 1e-233 | -5.387e+02 | 17.94% | 7.66% | 27.0bp (33.9bp) | VSX2/MA0726.1/Jaspar(0.986) |
| 2 | TGATTTATGG | 1e-227 | -5.233e+02 | 9.61% | 2.75% | 26.3bp (28.1bp) | PBX2(Homeobox)/K562-PBX2-ChIP-Seq(Encode)/Homer(0.957) |
| 3 | TGACCTCT | 1e-189 | -4.364e+02 | 16.43% | 7.41% | 26.6bp (30.6bp) | COUP-TFII(NR)/Artia-Nr2f2-ChIP-Seq(GSE46497)/Homer(0.954) |
| 4 | AGATCAAAGGGA | 1e-155 | -3.591e+02 | 7.75% | 2.48% | 26.1bp (28.7bp) | TCF7L2/MA0523.1/Jaspar(0.958) |
| 5 | AACATCTGGT | 1e-81 | -1.883e+02 | 8.47% | 4.06% | 25.7bp (28.8bp) | TWIST1/MA1123.1/Jaspar(0.946) |

Onimaru et al Figure 6-figure supplement 2

C1

| Rank | Motif | Name | P-value | log P-pvalue | q-value (Benjamini) | # Target Sequences with Motif | % of Targets Sequences with Motif | # Background Sequences with Motif | % of Background Sequences with Motif |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | Pitx1:Ebox(Homeobox,bHLH)/Hindlimb-Pitx1-ChIP-Seq(GSE41591)/Homer | 1e-43 | -9.972e+01 | 0.0000 | 172.0 | 1.61% | 174.7 | 0.45% |
| 2 |  | Hoxd11(Homeobox)/ChickenMSG-Hoxd11.Flag-ChIP-Seq(GSE86088)/Homer | 1e-27 | -6.342e+01 | 0.0000 | 1377.0 | 12.85% | 3732.4 | 9.58% |
| 3 |  | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer | 1e-27 | -6.301e+01 | 0.0000 | 240.0 | 2.24% | 395.0 | 1.01% |
| 4 |  | BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer | 1e-27 | -6.237e+01 | 0.0000 | 1058.0 | 9.87% | 2739.7 | 7.04% |
| 5 |  | Hoxa11(Homeobox)/ChickenMSG-Hoxa11.Flag-ChIP-Seq(GSE86088)/Homer | 1e-27 | -6.221e+01 | 0.0000 | 1312.0 | 12.24% | 3537.6 | 9.08% |

C2

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | Pitx1:Ebox(Homeobox,bHLH)/Hindlimb-Pitx1-ChIP-Seq(GSE41591)/Homer | 1e-67 | -1.552e+02 | 0.0000 | 233.0 | 1.84% | 168.1 | 0.45% |
| 2 |  | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer | 1e-52 | -1.202e+02 | 0.0000 | 238.0 | 1.88% | 214.2 | 0.58% |
| 3 |  | Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer | 1e-43 | -9.975e+01 | 0.0000 | 1138.0 | 8.97% | 2174.9 | 5.87% |
| 4 |  | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer | 1e-37 | -8.744e+01 | 0.0000 | 319.0 | 2.52% | 411.8 | 1.11% |
| 5 |  | LXH9(Homeobox)/Hct116-LXH9.V5-ChIP-Seq(GSE116822)/Homer | 1e-37 | -8.723e+01 | 0.0000 | 1025.0 | 8.08% | 1970.8 | 5.32% |

C3

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer | 1e-394 | -9.079e+02 | 0.0000 | 3056.0 | 17.54% | 2407.7 | 7.67% |
| 2 |  | Twist2(bHLH)/Myoblast-Twist2.Ty1-ChIP-Seq(GSE127998)/Homer | 1e-355 | -8.195e+02 | 0.0000 | 3486.0 | 20.01% | 3074.9 | 9.80% |
| 3 |  | TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer | 1e-329 | -7.590e+02 | 0.0000 | 3069.0 | 17.62% | 2630.8 | 8.39% |
| 4 |  | Tcf21(bHLH)/ArterySmoothMuscle-Tcf21-ChIP-Seq(GSE61369)/Homer | 1e-311 | -7.173e+02 | 0.0000 | 2256.0 | 12.95% | 1693.9 | 5.40% |
| 5 |  | Atoh1(bHLH)/Cerebellum-Atoh1-ChIP-Seq(GSE22111)/Homer | 1e-286 | -6.586e+02 | 0.0000 | 2354.0 | 13.51% | 1885.7 | 6.01% |

C4

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | BHLHA15(bHLH)/NIH3T3-BHLHB8.HA-ChIP-Seq(GSE119782)/Homer | 1e-100 | -2.307e+02 | 0.0000 | 1244.0 | 13.97% | 2957.1 | 7.40% |
| 2 |  | Twist2(bHLH)/Myoblast-Twist2.Ty1-ChIP-Seq(GSE127998)/Homer | 1e-97 | -2.246e+02 | 0.0000 | 1477.0 | 16.58% | 3780.0 | 9.45% |
| 3 |  | Hoxa13(Homeobox)/ChickenMSG-Hoxa13.Flag-ChIP-Seq(GSE86088)/Homer | 1e-94 | -2.166e+02 | 0.0000 | 2032.0 | 22.81% | 5830.6 | 14.58% |
| 4 |  | TCF4(bHLH)/SHSY5Y-TCF4-ChIP-Seq(GSE96915)/Homer | 1e-90 | -2.089e+02 | 0.0000 | 1277.0 | 14.34% | 3175.0 | 7.94% |
| 5 |  | HOXB13(Homeobox)/ProstateTumor-HOXB13-ChIP-Seq(GSE56288)/Homer | 1e-79 | -1.840e+02 | 0.0000 | 946.0 | 10.62% | 2193.3 | 5.49% |

C5

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer | 1e-1977 | -4.554e+03 | 0.0000 | 2200.0 | 13.90% | 243.0 | 0.73% |
| 2 |  | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer | 1e-1511 | -3.480e+03 | 0.0000 | 2175.0 | 13.74% | 394.5 | 1.19% |
| 3 |  | Pitx1:Ebox(Homeobox,bHLH)/Hindlimb-Pitx1-ChIP-Seq(GSE41591)/Homer | 1e-331 | -7.626e+02 | 0.0000 | 623.0 | 3.94% | 164.5 | 0.49% |
| 4 |  | REST-NRSF(Zf)/Jurkat-NRSF-ChIP-Seq/Homer | 1e-156 | -3.598e+02 | 0.0000 | 173.0 | 1.09% | 19.0 | 0.06% |
| 5 |  | NFY(CCAAT)/Promoter/Homer | 1e-144 | -3.318e+02 | 0.0000 | 1351.0 | 8.54% | 1322.8 | 3.98% |

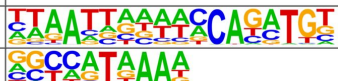
C6

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer | 1e-675 | -1.555e+03 | 0.0000 | 1160.0 | 5.61% | 179.7 | 0.62% |
| 2 |  | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer | 1e-478 | -1.103e+03 | 0.0000 | 1281.0 | 6.20% | 346.1 | 1.20% |
| 3 |  | Pitx1:Ebox(Homeobox,bHLH)/Hindlimb-Pitx1-ChIP-Seq(GSE41591)/Homer | 1e-141 | -3.261e+02 | 0.0000 | 405.0 | 1.96% | 116.5 | 0.40% |
| 4 |  | Sp5(Zf)/mES-Sp5.Flag-ChIP-Seq(GSE72989)/Homer | 1e-134 | -3.100e+02 | 0.0000 | 2768.0 | 13.39% | 2388.9 | 8.27% |
| 5 |  | Sp1(Zf)/Promoter/Homer | 1e-125 | -2.896e+02 | 0.0000 | 1273.0 | 6.16% | 851.9 | 2.95% |

C7

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | Pitx1:Ebox(Homeobox,bHLH)/Hindlimb-Pitx1-ChIP-Seq(GSE41591)/Homer | 1e-574 | -1.323e+03 | 0.0000 | 884.0 | 8.26% | 296.9 | 0.78% |
| 2 |  | Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer | 1e-353 | -8.134e+02 | 0.0000 | 2690.0 | 25.13% | 4258.4 | 11.18% |
| 3 |  | Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer | 1e-333 | -7.683e+02 | 0.0000 | 3691.0 | 34.49% | 7056.1 | 18.53% |
| 4 |  | Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer | 1e-315 | -7.273e+02 | 0.0000 | 1982.0 | 18.52% | 2773.4 | 7.28% |
| 5 |  | Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer | 1e-313 | -7.218e+02 | 0.0000 | 2048.0 | 19.13% | 2936.5 | 7.71% |

C8

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | Lhx2(Homeobox)/HFSC-Lhx2-ChIP-Seq(GSE48068)/Homer | 1e-211 | -4.880e+02 | 0.0000 | 1183.0 | 12.36% | 1766.1 | 4.45% |
| 2 |  | Lhx3(Homeobox)/Neuron-Lhx3-ChIP-Seq(GSE31456)/Homer | 1e-206 | -4.765e+02 | 0.0000 | 1615.0 | 16.87% | 2938.2 | 7.40% |
| 3 |  | Lhx1(Homeobox)/EmbryoCarcinoma-Lhx1-ChIP-Seq(GSE70957)/Homer | 1e-206 | -4.756e+02 | 0.0000 | 1203.0 | 12.57% | 1844.9 | 4.65% |
| 4 |  | LXH9(Homeobox)/Hct116-LXH9.V5-ChIP-Seq(GSE116822)/Homer | 1e-197 | -4.550e+02 | 0.0000 | 1450.0 | 15.15% | 2545.6 | 6.41% |
| 5 |  | Nkx6.1(Homeobox)/Islet-Nkx6.1-ChIP-Seq(GSE40975)/Homer | 1e-189 | -4.353e+02 | 0.0000 | 2210.0 | 23.09% | 4859.2 | 12.24% |

Onimaru et al Figure 6-figure supplement 3

# C5+C6 vs other peak regions

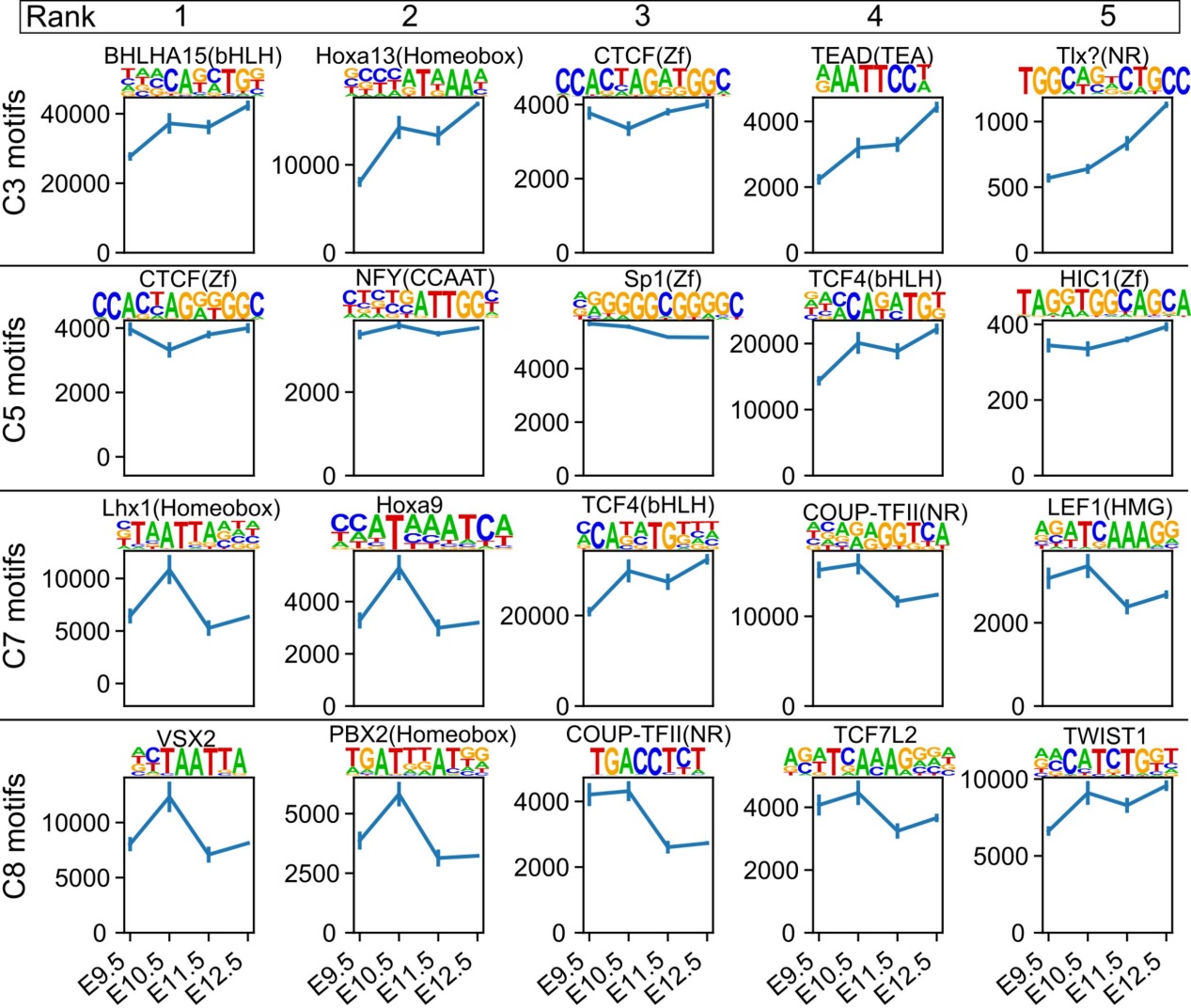
## De novo motif discovery

| Rank | Motif | P-value | log P-pvalue | % of Targets | % of Background | STD(Bg STD) | Best Match/Details |
|---|---|---|---|---|---|---|---|
| 1 |  | 1e-937 | -2.159e+03 | 10.44% | 2.99% | 25.2bp (31.1bp) | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer(0.915) |
| 2 |  | 1e-246 | -5.675e+02 | 5.51% | 2.38% | 26.2bp (29.5bp) | NFYA/MA0060.3/Jaspar(0.937) |
| 3 |  | 1e-243 | -5.609e+02 | 8.97% | 4.81% | 26.2bp (40.6bp) | Sp1(Zf)/Promoter/Homer(0.968) |
| 4 |  | 1e-184 | -4.256e+02 | 2.50% | 0.81% | 26.8bp (37.7bp) | ELK3/MA0759.1/Jaspar(0.981) |
| 5 |  | 1e-137 | -3.170e+02 | 2.78% | 1.13% | 27.8bp (36.3bp) | CTCFL/MA1102.2/Jaspar(0.778) |

## Known motif enrichment analysis

| Rank | Motif | Name | P-value | log P-pvalue | q-value (Benjamini) | # Target Sequences with Motif | % of Targets Sequences with Motif | # Background Sequences with Motif | % of Background Sequences with Motif |
|---|---|---|---|---|---|---|---|---|---|
| 1 |  | CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et_al.)/Homer | 1e-1043 | -2.403e+03 | 0.0000 | 3360.0 | 9.21% | 1359.9 | 2.16% |
| 2 |  | BORIS(Zf)/K562-CTCFL-ChIP-Seq(GSE32465)/Homer | 1e-793 | -1.827e+03 | 0.0000 | 3456.0 | 9.47% | 1793.5 | 2.85% |
| 3 |  | Sp1(Zf)/Promoter/Homer | 1e-203 | -4.697e+02 | 0.0000 | 2259.0 | 6.19% | 1932.3 | 3.07% |
| 4 |  | NFY(CCAAT)/Promoter/Homer | 1e-190 | -4.380e+02 | 0.0000 | 2599.0 | 7.12% | 2409.7 | 3.82% |
| 5 |  | KLF1(Zf)/HUDEP2-KLF1-CutnRun(GSE136251)/Homer | 1e-180 | -4.162e+02 | 0.0000 | 4192.0 | 11.48% | 4587.6 | 7.28% |

Onimaru et al Figure 6-figure supplement 4




Onimaru et al Figure 6-figure supplement 5