

RESEARCH

Open Access



Comparative genomics analyses reveal sequence determinants underlying interspecies variations in injury-responsive enhancers

Luzhang Ji[†], Yuanyuan Shi[†] and Qian Bian^{*}

Abstract

Background Injury induces profound transcriptional remodeling events, which could lead to only wound healing, partial tissue repair, or perfect regeneration in different species. Injury-responsive enhancers (IREs) are *cis*-regulatory elements activated in response to injury signals, and have been demonstrated to promote tissue regeneration in some organisms such as zebrafish and flies. However, the functional significances of IREs in mammals remain elusive. Moreover, whether the transcriptional responses elicited by IREs upon injury are conserved or specialized in different species, and what sequence features may underlie the functional variations of IREs have not been elucidated.

Results We identified a set of IREs that are activated in both regenerative and non-regenerative neonatal mouse hearts upon myocardial ischemia-induced damage by integrative epigenomic and transcriptomic analyses. Motif enrichment analysis showed that AP-1 and ETS transcription factor binding motifs are significantly enriched in both zebrafish and mouse IREs. However, the IRE-associated genes vary considerably between the two species. We further found that the IRE-related sequences in zebrafish and mice diverge greatly, with the loss of IRE inducibility accompanied by a reduction in AP-1 and ETS motif frequencies. The functional turnover of IREs between zebrafish and mice is correlated with changes in transcriptional responses of the IRE-associated genes upon injury. Using mouse cardiomyocytes as a model, we demonstrated that the reduction in AP-1 or ETS motif frequency attenuates the activation of IREs in response to hypoxia-induced damage.

Conclusions By performing comparative genomics analyses on IREs, we demonstrated that inter-species variations in AP-1 and ETS motifs may play an important role in defining the functions of enhancers during injury response. Our findings provide important insights for understanding the molecular mechanisms of transcriptional remodeling in response to injury across species.

Keywords Injury-responsive enhancer, Transcriptional response, AP-1, ETS, Motif frequency

Background

Proper responses to the injury caused by a variety of potentially deleterious factors, which minimize the tissue damage and promote the healing processes, are essential for living organisms to survive and thrive. The heart is one of the most vital organs of the body, and heart disease is the leading cause of mortality worldwide [1]. In response to cardiac injury, pronounced transcriptional responses are triggered, which could eventually lead to only wound healing, partial myocardial repair, or perfect

[†]Luzhang Ji and Yuanyuan Shi contributed equally to this work.

*Correspondence:

Qian Bian
qianbian@shsmu.edu.cn
Shanghai Institute of Precision Medicine, Ninth People's Hospital,
Shanghai Jiao Tong University School of Medicine, Shanghai 200125,
China



regeneration in different species. Some vertebrates, such as the newt [2] and zebrafish [3], have the remarkable ability to reconstruct lost cardiac tissue. The heart regenerative process consists of a series of intimately linked events including rapid clotting, inflammatory cell infiltration, epicardial activation, and cardiomyocyte proliferation [4]. Any perturbations to these steps could result in the failure of regeneration. In contrast, the regenerative capability of mammalian heart tissue is generally limited. The mechanisms underlying the interspecies differences in cardiac injury responses are not fully understood.

Induction of cell proliferation and cell fate transitions upon injury requires precise regulation and remodeling of gene expression programs. Previous studies have revealed the critical roles of transcription factors, epigenetic regulators and transcriptional enhancers in injury responses [5, 6]. Enhancers are highly abundant non-coding sequences in the genome that promote the expression of target genes in *cis* by recruiting transcription factors and coactivators [7]. Enhancer elements that are activated rapidly in response to injury signals, termed as injury-responsive enhancers (IREs), have been identified across many species. Intriguingly, some IREs have been shown to promote tissue regeneration in several organisms. For instance, a small intergenic element located upstream of the *lepb* locus, named LEN, is required for the dramatic increase in the expression level of *lepb* and the effective regeneration of zebrafish fins and hearts [8]. In *Drosophila*, thousands of enhancers are identified during the recovery process of wing imaginal discs, and some of these elements have been validated as bona fide regeneration enhancers [9]. Importantly, comparative H3K27ac (an active enhancer mark) chromatin immunoprecipitation sequencing (ChIP-seq) and RNA sequencing (RNA-seq) data of zebrafish and African killifish during the early stage of fin regeneration have revealed that activator protein 1 (AP-1)-binding motifs are enriched in a group of evolutionarily conserved enhancers, suggesting that the AP-1 transcription factor may play a critical role in instructing the function of IREs [10].

The mammalian heart exhibits a progressive decline in regenerative abilities after birth. Neonatal mice on the day of birth (P1) can regenerate cardiac tissues efficiently after injury, but this regenerative response is lost at postnatal day 7 (P8) [11]. In line with this, in response to injury, such as myocardial infarction (MI), the adult human heart can only form fibrotic scars to replace the lost cardiomyocytes (CMs), which eventually leads to heart failure [12]. Previous transcriptome analyses have shown that the regenerative and non-regenerative mouse hearts exhibit differential transcriptome changes upon injury [13–15]. Epigenomic profiling revealed profound changes in H3K27ac levels in the regenerative

and non-regenerative mouse hearts [14], suggesting the activation of IREs upon injury. However, the functional significances of mammalian IREs during injury responses remain elusive. Moreover, whether IREs elicit differential transcriptional effects in different species, thereby contributing to the different injury-responsive strategies, namely potent regeneration versus limited regeneration, has not been evaluated. Lastly, how the IREs vary across the evolution spectrum, and what sequence determinants underlie the variations of IREs, are still unclear.

Here, we address these questions by employing a comparative genomic approach. We compared the epigenome and transcriptome changes of P1 and P8 mouse hearts in response to myocardial ischemia-induced injury, and found that the same IREs activated in regenerative P1 mouse hearts are also induced in non-regenerative P8 heart tissues. By comparing the mouse cardiac IREs versus the IREs activated during zebrafish heart regeneration, we showed that AP-1 and ETS transcription factor binding motifs are highly enriched in IREs for both species. However, the genes associated with IREs vary considerably between zebrafish and mice. Cross-species comparison between IREs and their orthologous sequences reveals that the turnover of IREs is accompanied by alterations in AP-1 and ETS binding motif frequencies. For zebrafish and mice, the IREs specific to one species harbor relatively higher frequencies of AP-1 or ETS motifs than their non-injury-responsive orthologous sequences in the other species. Using mouse cardiomyocytes as a model, we showed that sufficient AP-1 or ETS motifs are indeed required for the effective activation of IREs upon hypoxia-induced damage. Together, our results suggest that shuffling of AP-1 and ETS binding motifs within IREs during evolution leads to diverse transcriptional changes in response to similar injury signals, which may contribute to the differential regeneration abilities in different organisms.

Results

Dynamic changes of enhancer landscape upon injury in mouse heart

To investigate enhancer dynamics during tissue injury and regeneration in mammals, we first analyzed previously published H3K27ac ChIP-seq data of neonatal mouse hearts subjected to permanent ligation of the left anterior descending (LAD) artery or sham surgery at P1 [14]. By comparing samples at 1.5 days post-injury (dpi) with the control of 1.5 days post-sham (dps), we identified a total of 1,852 IREs that are at least 2 kb away from the nearest transcriptional start sites (TSSs) of genes and exhibited significantly higher H3K27ac signals in P1 + 1.5dpi hearts than P1 + 1.5dps (Fig. 1A, Table S1). As a control, we defined 16,893 non-IREs that are active

enhancers in both uninjured and regenerating samples, and exhibited no increase in H3K27ac signals (Fig. S1A, Table S1). By integrating corresponding RNA-seq data [14], we found that IRE-associated genes (defined as the nearest genes from the IREs) exhibited a significant increase in expression level upon injury compared to the genes not linked to IREs (Fig. 1B). Moreover, the genes located within 100 kb from IREs also exhibited an overall upregulation upon cardiac injury (Fig. S1B), consistent with the notion that enhancers may regulate some genes over a distance. These results collectively suggest a functional role of these IREs in transcriptional regulation. Meanwhile, the IRE-associated genes exhibited an increase in H3K27ac signals at the TSSs in P1 + 1.5dpi than P1 + 1.5dps (Fig. 1C-D, Fig. S1C), which were consistent with their transcriptional upregulation.

To better understand the nature of IREs, we classified the states of enhancers in uninjured P1 mouse hearts based on their H3K4me1, H3K27ac, and H3K27me3 levels [14, 16], and defined 1,159 poised enhancers (H3K4me1 and H3K27me3), 6,823 active enhancers (H3K4me1 and H3K27ac), and 25,894 primed enhancers (H3K4me1 only) (Fig. S1D). Interestingly, the majority of IRE regions (72.68%) did not exhibit any of these epigenetic marks in the absence of injury, and hence did not belong to any of the three enhancer classes (Fig. S1E, upper). By integrating cCRE (candidate *cis*-regulatory element) annotations derived from ENCODE [17], we further found that 84.77% of inactive IREs are likely to be potential enhancers (high DNase-seq signal and high H3K27ac signal) in at least one other cell types or tissues across different developmental stages (Fig. S1E, bottom). Thus, cardiac IREs are bona fide enhancer elements that could be utilized in other developmental contexts.

Interestingly, the P8 mouse hearts, which have lost regenerative capability, exhibit similar changes in enhancer landscape in response to injury signals. By analyzing H3K27ac ChIP-seq data of P8 mouse hearts [14],

we found that H3K27ac signals at the 1,852 IREs identified in P1 heart also exhibited an overall increase in P8 hearts after injury (Fig. 1E). In line with this, the genes nearest to the IREs or the genes located within 100 kb from IREs were activated significantly in the P8 mouse hearts upon injury (Fig. 1F, Fig. S1F) with an increase in H3K27ac signals at gene promoters, which were similar to the P1 mouse hearts (Fig. 1G-H, Fig. S1G). For example, an IRE located upstream of *Ctss*, which can function as a regulator in antigen processing and T cell-mediated immune responses [18], was activated in both P1 and P8 mouse hearts upon injury with an increase in *Ctss* expression level (Fig. 1I). Another example is the genomic region around the *Cxcl1* locus, and this gene plays an important role in inflammation and wound healing [19]. Three IREs are in the intergenic regions upstream of *Cxcl1*, and H3K27ac signals of these IREs were increased upon injury in both P1 and P8 mouse hearts, accompanied by up-regulation of *Cxcl1* expression level (Fig. S1H). These findings suggest the activation of IREs is not sufficient for cardiac regeneration in P8 mice.

Notably, we found that the extent of IRE activation, which was indicated by the fold changes of H3K27ac signals, was markedly higher in P1 than in P8 mouse hearts (Fig. 1J). Consistent with this, the IRE-associated genes exhibited a modestly but significantly greater upregulation in the regenerative P1 than in the non-regenerative P8 mouse hearts (Fig. 1K). A similar pattern of transcriptional changes was also observed when using the genes that are located within 100 kb from IREs (Fig. S1I). Among these genes, some are characterized as pivotal heart regeneration regulators, such as the Notch signaling modulator *lfn3* (P1: 2.5 fold, P8: 1.33 fold) [20], the immune cell infiltration regulator *Cxcl1* (P1: 9.05 fold, P8: 3.17 fold) (Fig. S1H) [19, 21], as well as secretory leukocyte peptidase inhibitor *Slpi* (P1: 7.80 fold, P8: 3.46 fold) [22], that function to promote cardiac remodeling or cardiomyocyte proliferation. Thus, the correlation between

(See figure on next page.)

Fig. 1 Identification of injury-responsive enhancer elements in neonatal mouse heart during regeneration. **A** Heatmaps of H3K27ac ChIP-seq signals (normalized by corresponding input samples) at genomic regions around the center of IREs (from -5 kb to +5 kb) in the P1 + 1.5 dpi and P1 + 1.5dps samples. Each row represents a single enhancer ($N = 1,852$). Red stands for high density, and blue for low signal. **B** Boxplots showing expression level changes of IRE closest genes in P1 mouse hearts upon injury, and other genes are used as background. *P*-value: Wilcoxon rank-sum test. **C, D** Metagene plots (**C**) or heatmaps (**D**) of H3K27ac ChIP-seq signals (normalized by corresponding input samples) in the genomic regions around the TSSs (from -5 kb to +5 kb) of IREs nearest genes in P1 + 1.5dps and P1 + 1.5dpi mouse hearts. **E** Heatmaps of H3K27ac ChIP-seq signals (normalized by corresponding input samples) around the centers of IREs (identified in P1 mouse hearts) (from -5 kb to +5 kb) in the P8 + 1.5dpi and P8 + 1.5dps groups. Each row represents a single enhancer ($N = 1,852$). Red stands for high density, and blue for low signal. **F** Boxplots showing expression level changes of IRE (identified in P1 mouse hearts) closest genes in P8 mouse hearts upon injury compared to the other genes. *P*-value: Wilcoxon rank-sum test. **G, H** Metagene plots (**G**) or heatmaps (**H**) of H3K27ac ChIP-seq signals (normalized by corresponding input samples) around the TSSs (from -5 kb to +5 kb) of IRE (identified in P1 mouse hearts) nearest genes in P8 + 1.5dps and P8 + 1.5dpi mouse hearts. **I** Genome browser view showing the increase in H3K27ac ChIP-seq signal at an IRE (with brown shadow) and expression level of nearby gene (*Ctss*) in P1 and P8 mouse hearts upon injury. Red stands for 1.5 dps, and blue is for 1.5 dpi. Arrow under the gene stands for transcription direction. **J** Boxplot showing difference of IRE induction level quantified by log₂ fold change of H3K27ac ChIP-seq signals (1.5 dpi versus 1.5 dps) between P1 and P8 mouse hearts. *P*-value: Wilcoxon signed-rank test. **K** Boxplot displaying difference of gene activation level measured by log₂ fold change of expression level (1.5 dpi versus 1.5 dps) between P1 and P8 mouse hearts. *P*-value: Wilcoxon signed-rank test

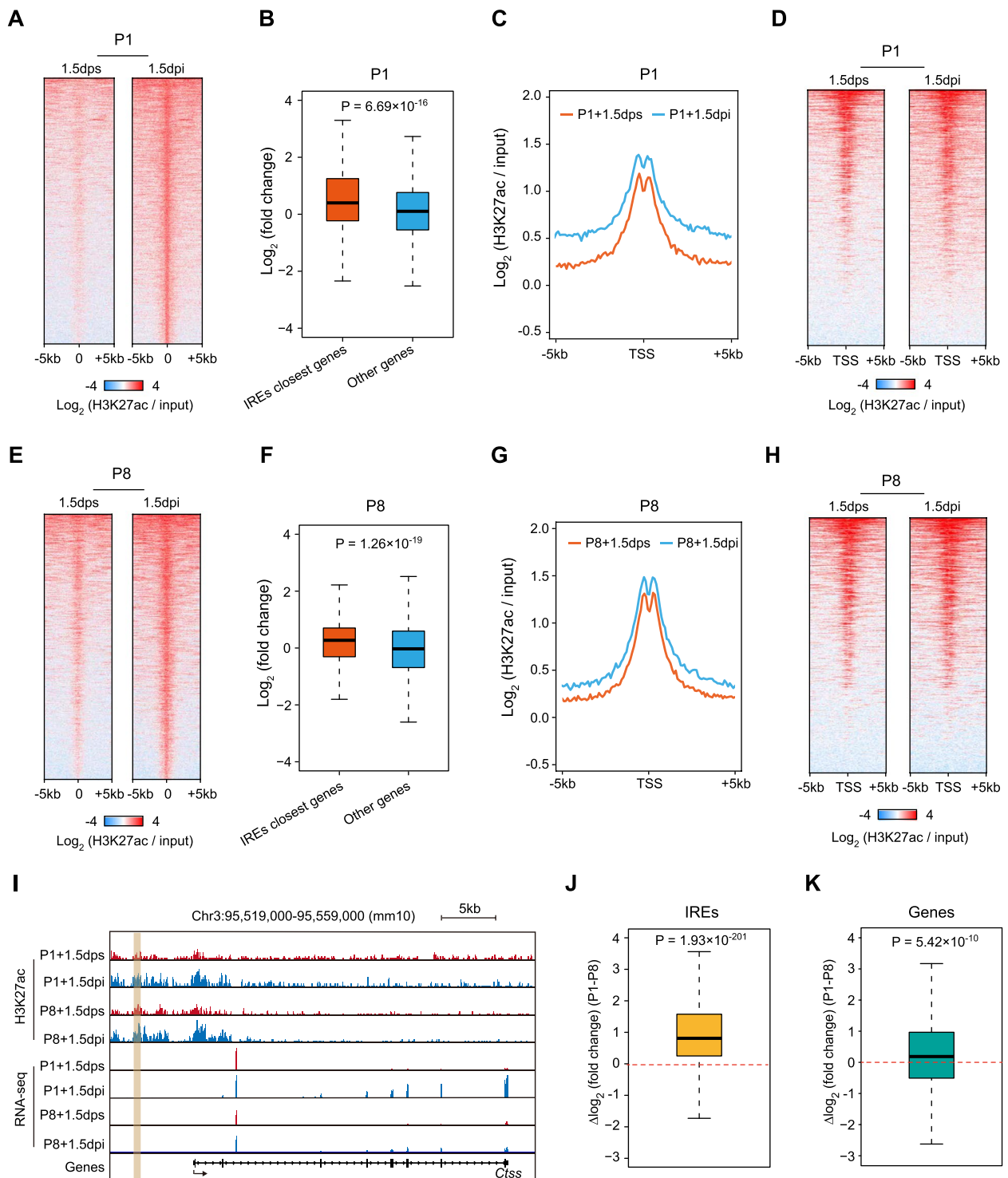


Fig. 1 (See legend on previous page.)

the attenuation of IRE activation and the decline in regenerative capacity implies that IREs might partly contribute to heart regeneration in mice, although this needs further experimental confirmation.

ETS and AP-1 binding motifs are enriched in IREs in both zebrafish and mice

Compared to mice, zebrafish retain a robust ability of heart regeneration upon injury throughout their lifespan

[23]. Previous genetic studies have demonstrated that many zebrafish enhancers activated by injury were able to direct regeneration-related gene expression and promote tissue regeneration [10, 24]. To understand the molecular underpinnings for the functions of zebrafish and mouse IREs during regeneration, we analyzed the previously published H3K27ac ChIP-seq data of zebrafish heart regeneration [8], and identified 646 IREs that exhibited significantly higher H3K27ac signals after injury (Fig. 2A, Table S2). Similar to the mouse IREs, the nearest genes to these zebrafish IREs were significantly up-regulated following injury (Fig. 2B, Fig. S2A), and exhibited increased H3K27ac signals at their promoters (Fig. 2C). These data enable us to compare the sequences between mouse and zebrafish cardiac IREs, and assess their differential functions in the context of cardiac regeneration. It should be noted that we use the genomic data from the P1 mouse heart from now on, as the IREs exert a more prominent regulatory function at this stage.

We performed motif enrichment analysis (MEA) to compare the sequence composition of zebrafish and mouse cardiac IREs. By using a relatively stringent cutoff (p -value $< 1 \times 10^{-10}$), we found that AP-1 and ETS families of transcription factors binding motifs were preferentially enriched in IREs for both zebrafish and mouse heart (Fig. 2D, Fig. S2B, Table S3). AP-1 transcription factor family, composed of many members such as JUN, FOS, and ATF, can regulate a variety of biological processes such as cell proliferation and differentiation [25]. ETS family proteins have been shown to activate the MAPK signaling pathway and regulate the expression level of early response genes to affect cell proliferation [26]. By analyzing previously published ChIP-seq datasets for several different cell lines, we showed that genomic sequences containing more AP-1 or ETS binding motifs were indeed the high-probability binding sites for the corresponding transcription factors (Fig. S2C-D). Interestingly, we found that the expression level of nearly all AP-1 transcription factors whose motifs are enriched in IREs were significantly

up-regulated upon injury (Fig. 2E). On the other hand, four IRE-enriched ETS transcription factors in zebrafish (*etv4*, *spi1a*, *spi1b*, and *elf1*) and their mouse orthologs (*Etv4*, *Spi1*, and *Elf4*) exhibited transcriptional upregulation following cardiac injury in zebrafish and mice, respectively (Fig. 2F). These results suggest that AP-1 and ETS families of transcription factors are likely to respond to cardiac injury signals in a similar fashion.

We further compared the motif composition of mouse and zebrafish IREs. For AP-1 transcription factors, 314 out of 646 zebrafish IREs and 853 out of 1,852 mouse IREs contain at least one motif. In contrast, the ETS motifs are more prevalent within the IREs, as 563 zebrafish IREs and 1,645 mouse IREs contain at least one ETS motif (Fig. 2G). We further classified IREs into four groups based on transcription factor binding motifs: AP-1 only, ETS only, both AP-1 and ETS, and neither AP-1 nor ETS (Fig. 2G). We found that the percentages of the four groups of IREs are largely similar between zebrafish and mice, with the zebrafish containing a slightly higher fraction of AP-1-only IREs (4.5% in zebrafish versus 2.5% in mice) (Fig. 2G). Therefore, the IREs in mice and zebrafish share similar motif compositions.

We further assessed the activation patterns for IREs associated with different classes of transcription factors. In both species, the AP-1-motif-containing IREs exhibited significantly higher H3K27ac signals before and after injury than the IREs without the AP-1 motifs (Fig. 2H). However, due to the relatively higher signals in the uninjured heart, the AP-1-motif-containing IREs exhibited smaller fold changes of H3K27ac signals after injury (Fig. 2H). A similar pattern of H3K27ac changes was observed for the ETS-motif-containing IREs (Fig. 2I). Collectively, these results suggest that the AP-1 and ETS binding motifs designate a class of enhancers that are already active in the uninjured heart but are poised for further induction upon injury. The similar behaviors for the IREs in zebrafish and mice imply that these regulatory elements are regulated by similar injury-induced signals and regulators.

(See figure on next page.)

Fig. 2 AP-1 and ETS motifs are enriched in cardiac IREs in both zebrafish and mice. **A** Heatmaps of H3K27ac ChIP-seq signals (normalized read counts) at genomic regions around the center of zebrafish IREs (from -5 kb to +5 kb) in the regenerating and uninjured groups. Each row represents a single enhancer ($N=646$). Red stands for high density, and white for low signal. **B** Boxplots showing expression level changes of IRE closest genes and other genes (as background) in zebrafish hearts upon injury. P -value: Wilcoxon rank-sum test. **C** Metagene plots (left) or heatmaps (right) showing H3K27ac ChIP-seq signals (normalized read counts) at the genomic regions around the TSSs (from -5 kb to +5 kb) of IRE nearest genes in uninjured and regenerating zebrafish hearts. **D** Scatter plots of motif enrichment analyses for zebrafish (left) and mouse (right) IREs during heart regeneration. AP-1 motifs are highlighted in blue, and ETS motifs are in red. Each dot stands for a single transcription factor binding motif. Red dotted line represents a p -value cutoff of 1×10^{-10} . **E, F** Line plots showing expression levels of genes encoding AP-1 (**E**) or ETS (**F**) transcription factors upon cardiac injury in zebrafish and mice. P -values: paired Student's t -test. Mouse ETS genes up-regulated notably and their orthologs in zebrafish are highlighted in (**F**). **G** Donut plots showing the counts and percentages of different types of IREs relative to AP-1 or ETS motif (orange: only with AP-1 motif, blue: only with ETS motif, red: with both AP-1 and ETS motif, grey: with neither AP-1 nor ETS motif) in zebrafish (left) and mouse (right) hearts. **H, I** Boxplots of H3K27ac ChIP-seq signals (log₂ transformed) for zebrafish and mouse IREs with or without AP-1 (**H**) or ETS (**I**) motif in uninjured (left) and regenerating (middle) samples, as well as log₂ fold changes (regenerating versus uninjured) of H3K27ac signals (right). P -values: Wilcoxon rank-sum test. The median values are shown at the top of graphs

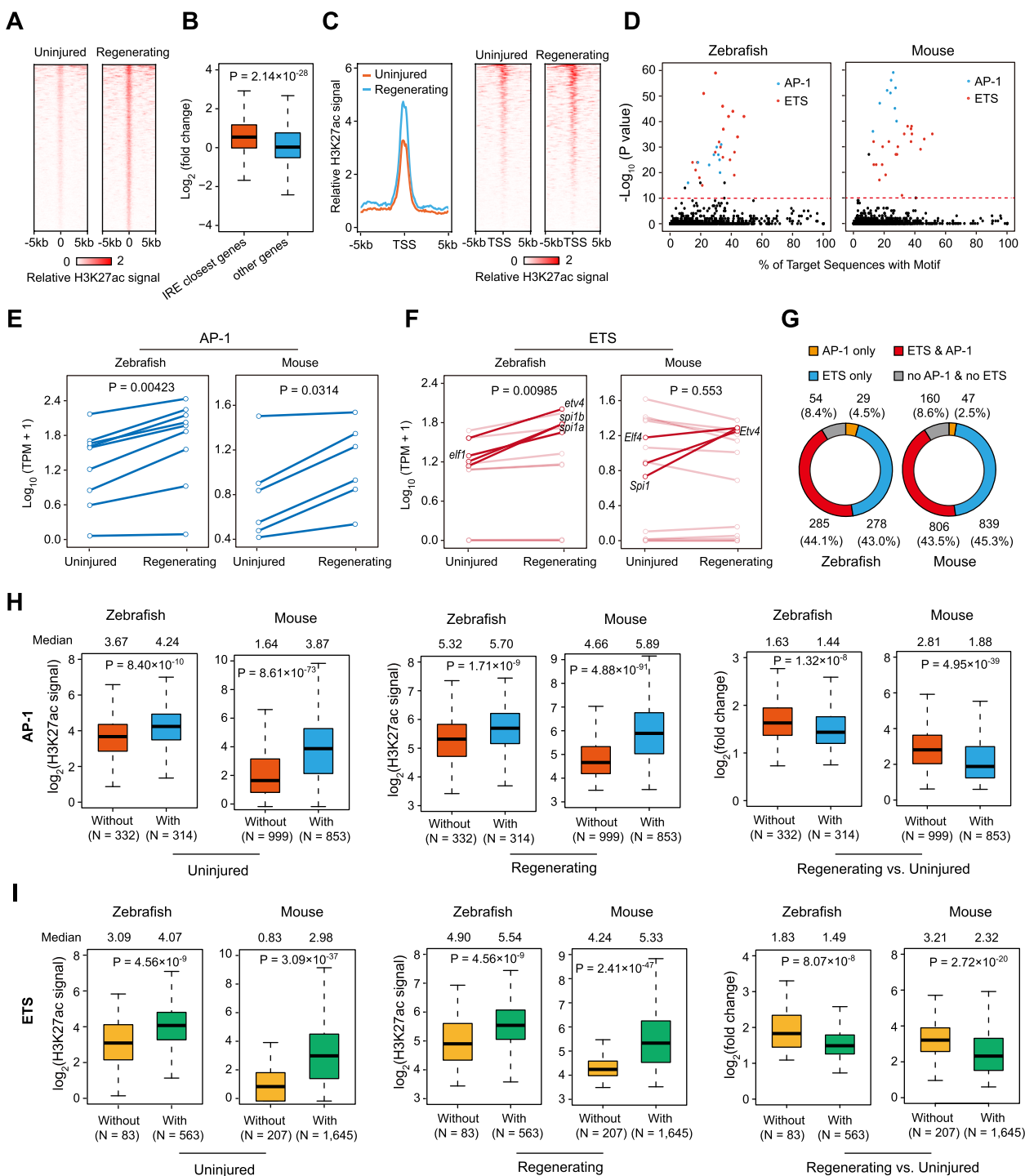


Fig. 2 (See legend on previous page.)

Cardiac IREs in zebrafish and mice are associated with genes of different biological functions

To further understand the regulatory roles of IREs in the regeneration process, we examined whether the IREs regulate different sets of genes in different species. We found

that the genomic distances between the IREs and their associated genes are similar in zebrafish and mice (Fig. S3A). Moreover, almost all (more than 97%) zebrafish or mouse IREs locate within 100 kb from their associated genes (Fig. S3B).

Using the scRNA-seq (single-cell RNA-seq) data of cardiac tissues in zebrafish [27] and mice [28] as the reference, we assigned every significantly up-regulated IRE-associated gene identified in bulk RNA-seq data with a corresponding cell type. The cell type compositions associated with the IRE-regulated genes are notably different between zebrafish and mice (Fig. S3C-D). In zebrafish, the significantly up-regulated IRE-associated genes upon heart regeneration are mainly induced in B cells, monocytes, and epicardial cells (Fig. S3C). In contrast, the up-regulated IRE-associated genes in mice are mainly associated with T cells, B cells, and cardiomyocytes (Fig. S3D). These findings suggest the differential functions of cardiac IRE-associated genes in zebrafish and mice.

To examine the epigenome signatures of IREs in the cell types mentioned above, we analyzed mouse histone modification CHIP-seq data (zebrafish-related data are not available) that are currently available in the Cistrome Data Browser [29]. We found that the IREs are generally associated with active epigenetic marks (H3K4me1, H3K4me3, H3K9ac, H3K27ac, except H3K36me3), and depleted in inactive marks (H3K9me3 and H3K27me3) in these cell types (Fig. S3E). However, the chromatin signatures of IRE regions differ across various cell types. Notably, the cardiomyocytes and fibroblasts exhibit lower H3K4me3 and H3K27ac signals than the immune cell types (Fig. S3E), suggesting that IREs may respond differently to injury signals in cell types that are relevant to cardiac regeneration.

The differences in IRE-gene regulatory networks between zebrafish and mice can be partly attributed to the loss or gain of genes during evolution. Among the 559 zebrafish IRE-associated genes, 376 (67.3%) have orthologs in mice (Fig. S3F, left). However, only 461 of 1,221 mouse IRE-associated genes (37.8%) have orthologs in zebrafish (Fig. S3F, right). We next performed GO (Gene Ontology) [30] (Table S4) and KEGG (Kyoto Encyclopedia of Genes and Genomes) [31] (Table S5) enrichment analyses on IRE-associated genes that have orthologs in both zebrafish and mice (376 zebrafish genes, 461 mouse genes). To overcome species-biased information in the database, we converted zebrafish genes to their orthologs in mice. For simplicity, we still used “IRE-associated genes” to refer to these genes in this part. Based on the extent of enrichment for each GO term or KEGG pathway within the IRE-associated genes in zebrafish versus mice, we classified these terms into four categories: enriched only in zebrafish, enriched only in mice, enriched in both species, and not enriched in either species (Fig. 3A). Whereas 1.4%, 2.9%, and 2.1% of GO terms are enriched in IRE-associated genes only in zebrafish, only in mice, or in both species, respectively, most of GO terms are not enriched in either zebrafish

or mice (3,853 GOs, 91.5%). Similarly, the ratios of the KEGG pathways belonging to these four categories are 13.8%, 8.2%, 9.0%, and 69.0%, respectively (Fig. 3B). These results indicate that the functions of cardiac IRE-associated genes diverge between zebrafish and mice.

We then took a closer examination of the most enriched biological functions for IRE-associated genes in zebrafish and mice. Among the ten most significantly enriched GO terms in zebrafish, seven are also significantly enriched in mouse IRE-associated genes (q -value < 0.05), including “cell activation”, “positive regulation of cell death”, and “positive regulation of locomotion” (Fig. 3C). However, the GO terms “dendritic cell differentiation” and “collagen catabolic process”, both of which have important implications for heart repair and regeneration processes [32, 33], are only significantly enriched in zebrafish IRE-associated genes, but not in mouse IRE-associated genes. Similarly, six out of the top ten representative KEGG pathways enriched in zebrafish, such as “TNF and thyroid hormone signaling pathways”, are also enriched in mice (Fig. 3D). However, genes associated with “MAPK signaling pathway”, which promotes cell proliferation, growth, and survival, and may play important roles during heart regeneration [34], are only enriched in zebrafish IRE-associated genes. In line with this, we also found that a portion of the top ten enriched GO terms or KEGG pathways in mouse IRE-associated genes are not enriched in zebrafish (Fig. 3E-F). Concordantly, the up-regulated genes (fold change > 1) located within 100 kb from IREs in zebrafish and mice also exhibit similar differences in the enriched GO terms (Fig. S3G-I). Taken together, our analyses demonstrate that the genes associated with IREs vary considerably between zebrafish and mice, indicative of different regulatory functions of IREs.

Loss of zebrafish-specific cardiac IREs in mice correlates with the changes in transcriptional injury responses

Next, we checked IRE-gene regulatory relationship for the genes that have orthologs in both zebrafish and mice and are associated with the IREs in at least one species. For the 376 mouse orthologs of the zebrafish IRE-associated genes, only 10.1% are associated with mouse cardiac IREs (Fig. S4A, left). Similarly, only 8.2% of the 461 zebrafish orthologs of the mouse IRE-associated genes are also linked to the IREs in zebrafish (Fig. S4A, right), suggesting that the regulatory relationships between genes and IREs are altered during evolution. To directly assess how these differences influence transcriptional responses upon cardiac injury, we classified these genes into three categories: genes associated with IREs in both zebrafish and mice (common), genes associated with IREs only in zebrafish (zebrafish-specific), and genes associated with IREs only in mice (mouse-specific) (Table

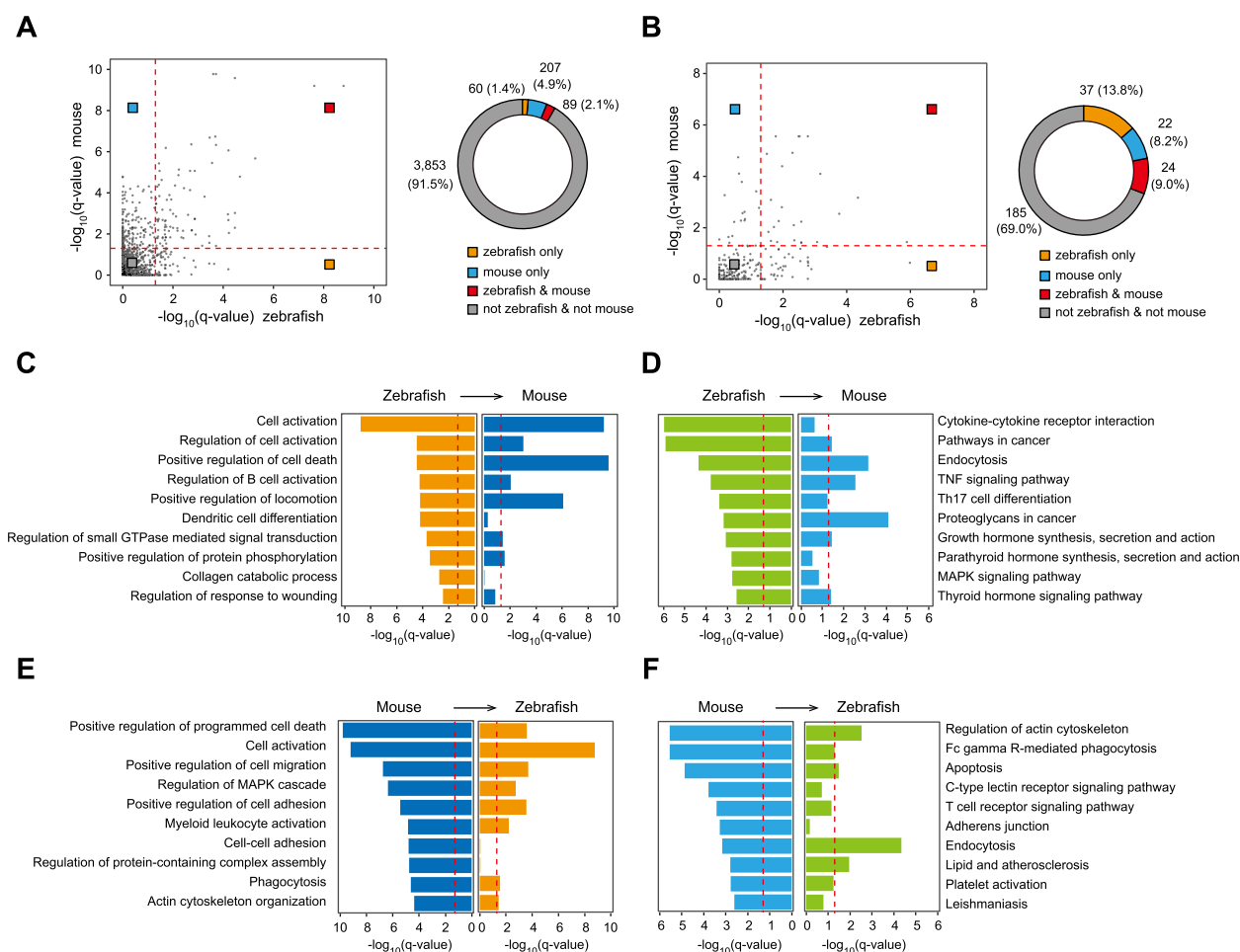


Fig. 3 Genes associated with cardiac IREs are functionally different between zebrafish and mice. **A** Left, scatter plot comparing enrichment of biological process GO terms ($-\log_{10}$ transformed q-value) between zebrafish (x-axis) and mice (y-axis). Red dotted lines correspond to a q-value of 0.05. Right, donut plot showing the counts and percentages of different types of GO terms relative to zebrafish or mice (orange: only enriched in zebrafish, blue: only enriched in mice, red: enriched in both zebrafish and mice, grey: enriched in neither zebrafish nor mice). **B** Left, scatter plot showing enrichment of KEGG pathways ($-\log_{10}$ transformed q-value) between zebrafish (x-axis) and mice (y-axis). Red dotted lines stand for a q-value of 0.05. Right, donut plot displaying the counts and percentages of different types of KEGG pathways relative to zebrafish or mice as shown in (A). **C, D** Top 10 enriched GO terms (C) or KEGG pathways (D) in zebrafish and corresponding q-values ($-\log_{10}$ transformed) in mice. Red dotted lines stand for a q-value of 0.05. **E, F** Bar plots of top 10 enriched GO terms (E) or KEGG pathways (F) in mice and corresponding q-values ($-\log_{10}$ transformed) in zebrafish. Red dotted lines represent a q-value of 0.05

S6). Consistent with the notion that enhancers promote the expression of their associated genes, genes associated with the IREs in both zebrafish and mice indeed exhibited significant upregulation following injury in both species (Fig. 4A, upper). Concordantly, the H3K27ac signals at the promoters of these genes were also elevated (Fig. 4A, middle and bottom). These common genes are associated with biological processes such as “positive regulation of apoptotic process”, “myeloid leukocyte activation”, and “positive regulation of immune response” (Fig. S4B, Table S7).

We further assessed how the species-specific association with IREs influences injury-induced gene expression.

We found that the 338 genes only associated with IREs in zebrafish exhibited increased gene expression as well as promoter H3K27ac signals only in zebrafish, but not in mice (Fig. 4B). Thus, the loss of the IRE association of these genes in mice may dampen their transcriptional responses upon injury. These zebrafish-specific IRE-associated genes are involved in “cell activation”, “regulation of tissue remodeling” (Fig. S4C, Table S7), “chemokine signaling pathway”, and “cAMP signaling pathway” (Fig. S4D, Table S8).

Unexpectedly, the 423 genes that are associated with IREs only in mice exhibited comparable transcription upregulation upon cardiac injury in zebrafish despite the

lack of association with the IREs (Fig. 4C, upper). Functional enrichment analyses showed that these mouse-specific IRE-associated genes are related to the GO terms including “cell activation” and “positive regulation of programmed cell death” (Fig. S4E, Table S7), and KEGG pathways such as “C-type lectin receptor signaling pathway” and “Rap1 signaling pathway” (Fig. S4F, Table S8). Interestingly, the H3K27ac signals at the promoters of these mouse-specific IRE-associated genes were elevated upon injury in zebrafish (Fig. 4C, middle and bottom), suggesting that these genes may be regulated by promoter-proximal regulatory elements. Given the significant enrichment of AP-1 and ETS binding motifs in IREs, we further assessed the motif frequency within these promoters in zebrafish. Indeed, we found that the promoters of the zebrafish orthologs of the mouse-specific IRE-associated genes exhibited higher AP-1 motif frequency than the randomly selected zebrafish promoters (Fig. S4G). Thus, the mouse-specific IREs are preferentially linked to genes that can be induced efficiently in zebrafish even without the need for IREs.

Our findings together paint a picture that while the emergence of mouse-specific IREs does not necessarily increase the induction of their associated genes upon injury, the loss of zebrafish-specific IREs in mice may abrogate the activation of genes involved in certain biological functions. Indeed, we found that the IRE-associated genes in the GO terms or KEGG pathways that are enriched for both zebrafish and mice, such as “cell activation” or “TNF signaling pathway”, exhibited significant upregulation after injury in both species, as well as increased H3K27ac signals at the promoters (Fig. 4D-E). Moreover, genes in the GO terms or pathways that are only enriched for zebrafish, such as “dendritic cell differentiation” and “parathyroid hormone synthesis, secretion and action”, could only be activated in zebrafish, but not in mice (Fig. 4F-G). However, IRE-associated genes in the terms only enriched for mouse, such as “cell–cell adhesion” and “adherens junction”, could still be activated upon injury in zebrafish, accompanied by an elevation in promoter H2K27ac signals (Fig. 4H-I). Taken together,

our analyses suggest that the differences in IRE-gene regulatory landscapes could profoundly affect the transcriptional responses upon injury, which may contribute to the differential regeneration capacities across species.

Inter-species variations in IRE inducibility are associated with alterations of AP-1 and ETS motif frequencies

To further understand how the lost or gained IREs between zebrafish and mice lead to different outcomes in injury-induced transcriptional responses, we performed a cross-species comparison of the functional properties and sequence features of IREs. We first searched the orthologous sequences for IREs in the different species. Among the 646 zebrafish IREs, 437 (67.6%) are associated with genes that have mouse orthologs (Fig. 5A, left). Since most cardiac IREs are located within 100 kb from their nearest genes (Fig. S3B), we arbitrarily constrained the orthology search within 100 kb from the mouse orthologs of the zebrafish IRE-associated genes. Among the 437 mouse sequences that are orthologous to the zebrafish IREs, only 3.4% were identified as mouse IREs (Fig. 5A, right). Similarly, 706 of 1,852 mouse IREs (38.1%) have orthologous sequences in zebrafish, and only 2.3% of these orthologous sequences were activated in zebrafish upon injury (Fig. 5B). We noted that the percentages of IRE-orthologous sequences that retain the IRE functions in a different species (3.4% and 2.3%) (Fig. 5A-B) are significantly lower than the percentages of IRE-associated gene orthologs that still associate with IREs in a different species (10.1% and 8.2%) (Fig. S4A). Thus, most of the IREs that regulate the orthologous genes in different species do not arise from the same ancestral sequences. These results are consistent with the more rapid enhancer sequence evolution compared to the coding regions and the promoters [35]. Interestingly, we found that mouse IREs are less evolutionarily conserved compared to the non-IREs, which are enhancers exhibit similar high activities before and after injury, according to the phastCons60way (Fig. S5A) or phyloP-60way scores (Fig. S5B), suggesting that the regulation of

(See figure on next page.)

Fig. 4 Differences in IRE-gene regulatory landscapes affect transcriptional responses upon injury. **A–C** Upper, boxplots showing expression level changes of common (**A**), zebrafish-specific (**B**), or mouse-specific (**C**) IRE-associated genes (orange) compared to the other genes (used as background) (blue) in zebrafish and mice. *P*-values: Wilcoxon rank-sum test. Middle, metagene plots displaying smoothed H3K27ac ChIP-seq signals (zebrafish: normalized read counts, mouse: log₂ transformed read counts over corresponding input samples) at the regions around TSSs (from -5 kb to +5 kb) of common (**A**), zebrafish-specific (**B**), or mouse-specific (**C**) genes in zebrafish and mice (orange: uninjured, blue: regenerating). Bottom, heatmaps showing H3K27ac ChIP-seq signals around TSSs (from -5 kb to +5 kb) of common (**A**), zebrafish-specific (**B**), or mouse-specific (**C**) genes upon injury in zebrafish and mice. Each row represents a single gene. Red stands for high density, and white or blue for low signal. **D–I** Upper, boxplots displaying expression level changes of a representative common (**D**, “cell activation”; **E**, “TNF signaling pathway”), zebrafish-specific (**F**, “dendritic cell differentiation”; **G**, “parathyroid hormone synthesis, secretion and action”), or mouse-specific (**H**, “cell–cell adhesion”; **I**, “adherens junction”) GO term or KEGG pathway associated genes (orange) compared to the other genes (as background) (blue) upon cardiac injury. *P*-values: Wilcoxon rank-sum test. Middle and lower, metagene plots (middle) or heatmaps (lower) showing H3K27ac signals at the regions around TSSs (from -5 kb to +5 kb) of common (**A**), zebrafish-specific (**B**), or mouse-specific (**C**) genes in zebrafish and mice. The meaning of labels is the same as (**A–C**)

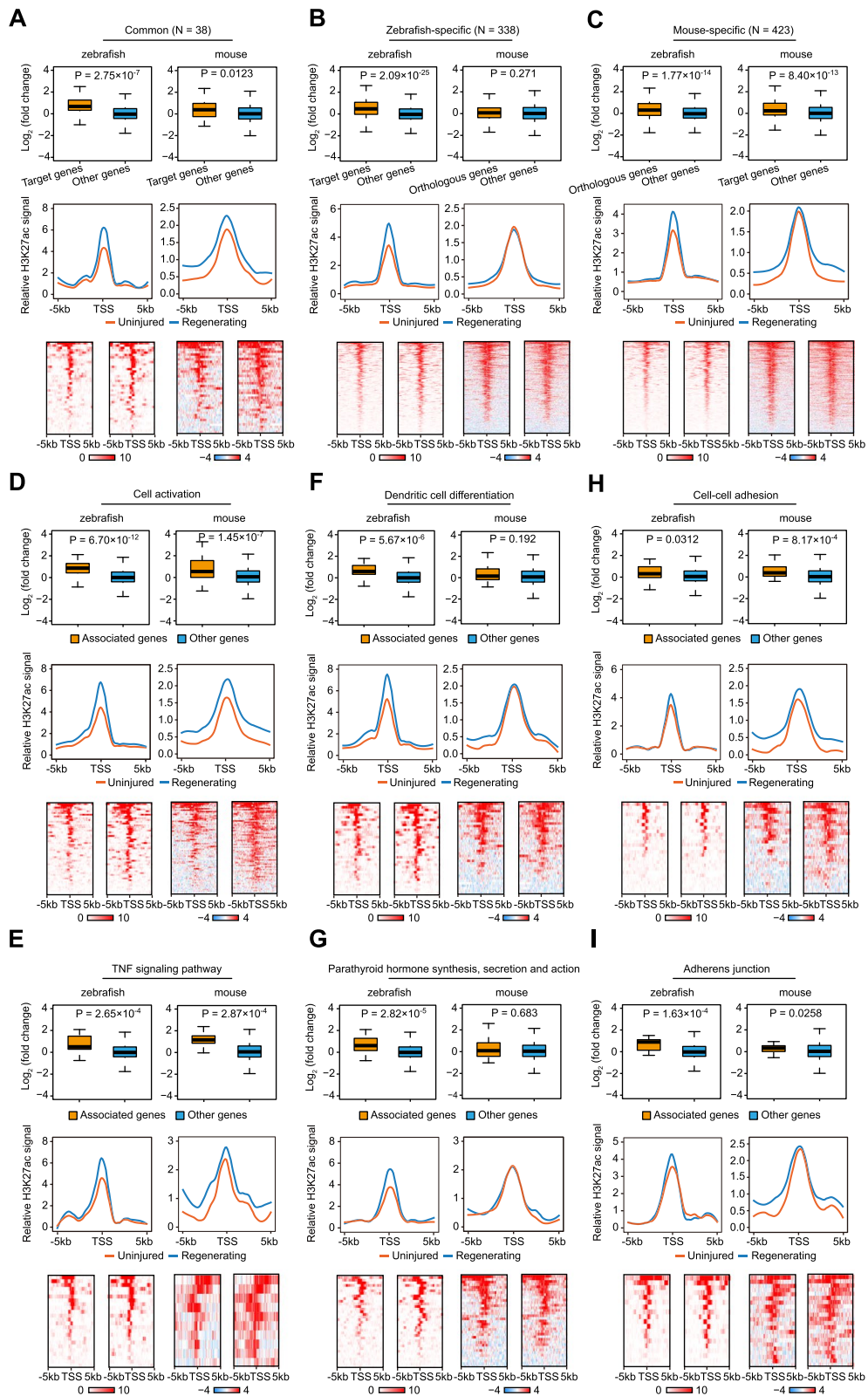


Fig. 4 (See legend on previous page.)

transcriptional responses upon injury does not contribute a strong evolution constraint.

To better characterize the functional divergence of the IRE-related sequences in different species, we selected all the cardiac IREs in zebrafish and mice that can be assigned with orthologous sequences in the other species, and classified these IREs into three categories: shared IREs, zebrafish-specific IREs, and mouse-specific IREs (Table S9). As expected, the 15 shared cardiac IREs showed an increase in H3K27ac ChIP-seq signals in both zebrafish and mice in response to cardiac injury (Fig. 5C). In contrast, while the 422 zebrafish-specific IREs exhibited significantly increased H3K27ac signals in zebrafish upon injury, their orthologous sequences in mice displayed no detectable H3K27ac signals both before and after injury (Fig. 5D). For the 690 mouse-specific IREs, these enhancer elements exhibited increased activities during mouse heart regeneration, but the H3K27ac signals at the center of their orthologous sequences in zebrafish showed no elevation upon cardiac injury (Fig. 5E). These findings together suggest that a species possesses a distinctive set of enhancer elements in response to injury.

The differential injury-responsiveness between the species-specific IREs and their orthologous sequences are accompanied by changes in their sequence composition, particularly the AP-1 and ETS transcription factor binding motifs. We quantified the frequencies of AP-1 and ETS motifs for each category of IREs. For shared IREs, their AP-1 or ETS motif frequencies showed no significant differences between zebrafish and mice (Fig. 5F). For instance, in zebrafish, *zeb2a* is located near an intronic IRE which contains only ETS motifs. The ortholog of *zeb2a* in mouse is *Zeb2*, which encodes a transcription factor that regulates nervous system development [36]

and angiogenesis [37]. The orthologous sequence of the *zeb2a* IRE in mice is located ~80 kb upstream of the *Zeb2* promoter, contains a similar frequency of ETS motifs, and exhibits a significant increase in H3K27ac signal upon cardiac injury, suggesting this sequence also functions as an IRE in mice (Fig. 5G). Consistent with this, *zeb2a* and *Zeb2* are up-regulated following injury in both organisms (Fig. 5G).

In contrast, the orthologous sequences of the zebrafish-specific IREs in mice, which lose injury-responsiveness, exhibited substantially lower frequencies of both AP-1 and ETS motifs (Fig. 5H). One prominent example is zebrafish *lepb* IRE, whose critical role in zebrafish heart and fin regeneration has been well established [8, 38]. The mouse orthologous sequence of the zebrafish *lepb* IRE contains a much lower frequency of both AP-1 and ETS motifs, and showed no elevated H3K27ac signals upon injury. Concomitant to the loss of IRE, the expression level of *Lep*, the mouse ortholog of *lepb*, is no longer upregulated in response to injury (Fig. 5I). Similar losses of zebrafish-specific IREs accompanying the decreased AP-1 and ETS motif frequency occur at *melk* (Fig. S5C) and *llgl2* genes (Fig. S5D), both of which are implicated in cell cycle and cell proliferation regulation [39, 40].

Interestingly, the zebrafish orthologs of the mouse-specific IREs contained significantly fewer ETS motifs but a similar amount of AP-1 motifs (Fig. 5J). Consistent with this finding, all 14 transcription factors from the ETS family exhibited lower frequencies in the zebrafish orthologs of the mouse-specific IREs (Fig. S5E-G), whereas only 2 out of 8 AP-1 family transcription factors (JUN and JUNB) exhibited this trend (Fig. S5H-J). Thus, the increases in ETS family transcription factors binding alone may be sufficient to increase the injury-responsiveness of enhancers. For instance, a mouse-specific IRE

(See figure on next page.)

Fig. 5 Variations in IRE inducibility are associated with alterations of AP-1 and ETS motif frequencies. **A** Pie charts showing the counts and percentages of IREs linked to orthologous (orange) and non-orthologous (grey) genes in zebrafish (left), and IREs activated (blue) or not activated (grey) in mice (right). **B** Pie charts displaying the counts and percentages of IREs linked to the orthologous (orange) and non-orthologous (grey) genes in mice (left), and IREs induced (blue) or not induced (grey) in zebrafish (right). **C-E** Aggregation plots of H3K27ac ChIP-seq signals (zebrafish: normalized read counts, mouse: log₂ transformed read counts over corresponding input samples) at the genomic regions around peak center (from -2 kb to +2 kb) of shared (**C**), zebrafish-specific (**D**), or mouse-specific (**E**) IREs in zebrafish and mouse uninjured (orange) and regenerating (blue) hearts. **F** Distribution of AP-1 (left) or ETS (right) motif frequency of shared IREs between zebrafish and mouse hearts. *P*-values: paired Student's *t*-test. **G** An example of dynamic changes of shared IREs and their associated genes (*zeb2a/Zeb2*) upon heart injury in zebrafish and mice. The genomic regions of AP-1 and ETS motif are also shown, and values of motif frequency are listed at the bottom of the graph. Two IREs (with brown shadow) linked by dotted line are orthologous. Arrows under the genes stand for transcription direction. **H** Dot plots showing AP-1 (left) or ETS (right) motif frequency of zebrafish-specific IREs in zebrafish and orthologous sequences in mice. The orange lines represent mean values for zebrafish, and the blue lines stand for mice. *P*-values: paired Student's *t*-test. **I** Genome browser view showing dynamics of zebrafish-specific IREs and their associated genes (*lepb/Lep*) upon cardiac injury in zebrafish and mice. The genomic regions of AP-1 and ETS motif are shown, and values of motif frequency are also listed. Two IREs linked by dotted line are the IRE in zebrafish and orthologous sequence in mice. Arrows under the genes represent the direction of transcription. **J** Dot plots displaying AP-1 (left) or ETS (right) motif frequency of mouse-specific IREs in mice and orthologs in zebrafish. The orange lines represent mean values for zebrafish, and the blue lines stand for mice. *P*-values: paired Student's *t*-test. **K** An example displaying changes of mouse-specific IREs and their associated genes (*xbp1/Xbp1*) upon injury in zebrafish and mice. The genomic regions of AP-1 and ETS motif are shown, and values of motif frequency are also listed at the bottom of the graph. Two IREs linked by dotted line are the IRE in mice and its orthologous sequence in zebrafish. Arrows under the genes indicate transcription direction

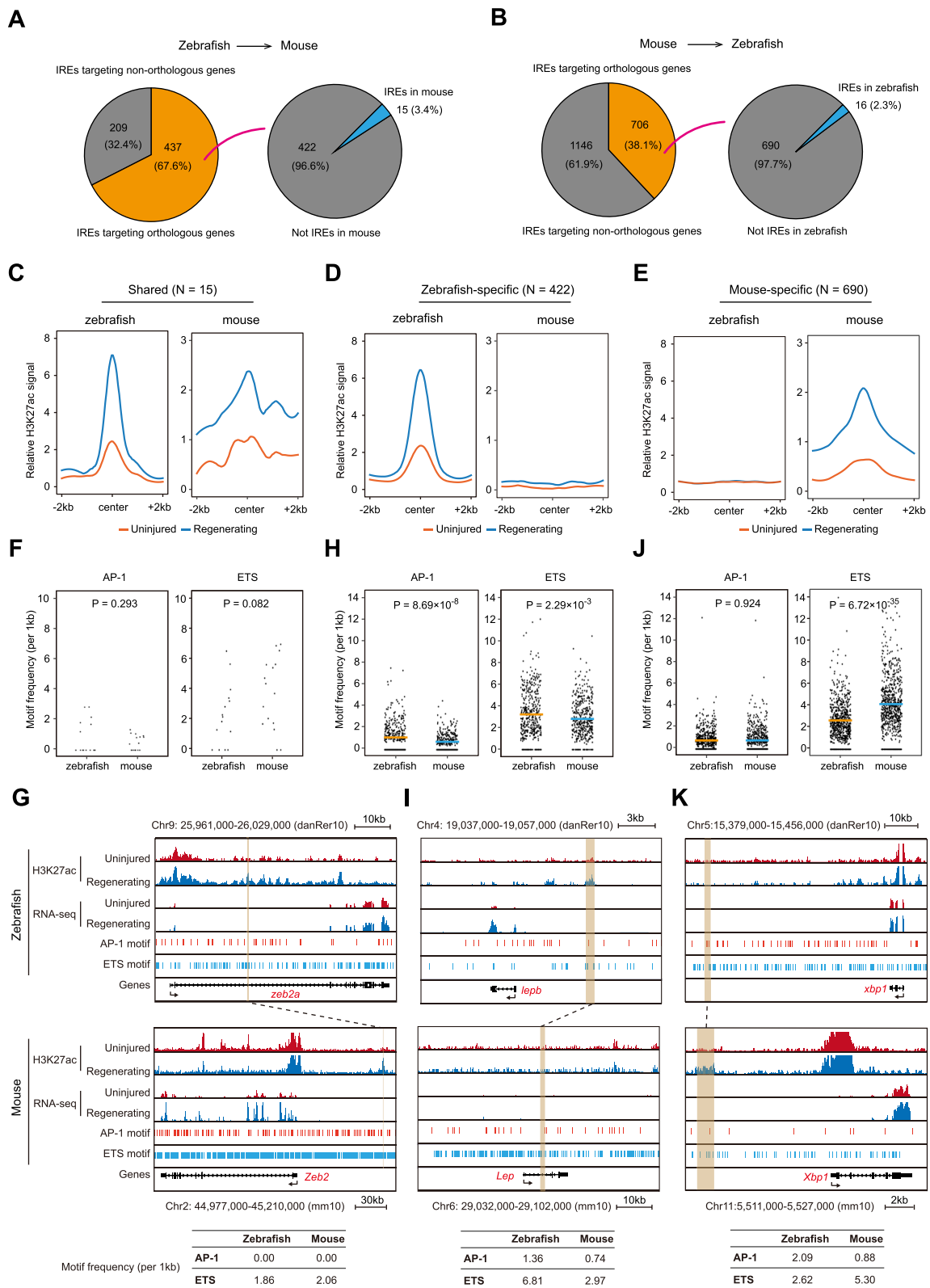


Fig. 5 (See legend on previous page.)

adjacent to the promoter of *Xbp1*, a critical transcription factor for stress responses [41], exhibits an elevated H3K27ac level upon heart injury. However, the zebrafish orthologous sequence of *Xbp1* IRE, which contains a lower ETS motif frequency and a higher AP-1 motif frequency, is not induced (Fig. 5K). Notably, *xbp1* exhibits transcriptional upregulation upon injury in zebrafish, again supporting the notion that many injury-responsive genes can be induced in zebrafish in an IRE-independent manner.

Taken together, our comparative epigenetic analyses of zebrafish and mouse cardiac IREs reveal that the interspecies variations in injury inducibility of enhancers are linked with the changes in AP-1 and ETS motif frequencies.

The abrogation of AP-1 or ETS motifs attenuates the activation of IREs in mouse cardiomyocytes

To validate the functional importance of AP-1 and ETS binding motifs for the induction of IREs upon injury, we chose mouse cardiomyocyte cell line HL-1 as a model system for its capacity of maintaining cardiac-specific phenotype [42], and simulated myocardial infarction that results from LAD in vivo by treating HL-1 cells with hypoxia [43]. Upon a hypoxia treatment of 24 h, HL-1 cells exhibited a dramatic increase of *Hif1 α* , which is a reliable biomarker of hypoxia [44], at both the mRNA level (Fig. 6A) and the protein level (Fig. 6B, Fig. S6A), suggesting the hypoxia responses were effectively induced.

Next, we chose two enhancers that are predicted to be mouse-specific IREs (E1 and E2) for functional validation. Both enhancers exhibited significant upregulation of H3K27ac level upon cardiac injury in mice, whereas their orthologous sequences in zebrafish showed no changes (Fig. 6C). Consistently, associated genes of E1

(*Pde1a*) and E2 (*Gda*) were significantly up-regulated upon injury in mouse hearts (Fig. S6B). In HL-1 cells, the expression levels of *Pde1a* and *Gda* were also up-regulated significantly upon hypoxia (Fig. 6D), indicating the reliability of our in vitro model. The 800 bp E1 enhancer contains 6 ETS motifs and 1 AP-1 motif, and the E2 enhancer, which is 459 bp in length, contains 3 ETS motifs and 1 AP-1 motif (Fig. 6E, Fig. S6C). Using dual luciferase reporter assay in HL-1 cells, we found that wildtype E1 or E2 exhibited no detectable enhancer activity in the normal culture condition, but were remarkably activated with a similar extent upon hypoxia treatment (Fig. 6F-G), thus demonstrating that these two enhancers are bona fide IREs.

To explicitly test the functional contribution of ETS and AP-1 motifs in the induction of IREs, we mutated all ETS motifs or AP-1 motifs in both enhancers to AAA AAAAA or AAAAAAA, respectively. For both enhancers, mutating either the ETS or the AP-1 motifs led to a reduction in the enhancer activation following hypoxia treatment (Fig. 6F-G), suggesting both types of motifs indeed contribute to the injury responsiveness of the IREs. Of note, deleting the single AP-1 motif in E1 led to a significant loss of enhancer activation, while the reduction in hypoxia response of E1 caused by ETS motif deletion were relatively modest (Fig. 6F-G). In contrast, mutation of the single AP-1 motif in E2 only resulted in a mild decrease in enhancer activation, whereas the deletion of the three ETS motifs in E2 nearly completely abolished its responses to hypoxia treatment (Fig. 6F-G). Therefore, the AP-1 and ETS family transcription factors may function in a sequence context-dependent manner to regulate the injury responsiveness of IREs. Taken together, our findings reveal that the reduction in ETS or AP-1 motif frequencies attenuated the injury responses of IREs in mouse cardiomyocytes in vitro.

(See figure on next page.)

Fig. 6 Response capability of cardiac IREs requires sufficient AP-1 or ETS motif frequency. **A** Bar plots showing *Hif1 α* mRNA expression level under normal and hypoxic conditions as measured by qPCR, using *Gapdh* for normalization. Error bars denote SEM ($N=3$). P -value: unpaired Student's t -test (**, $P<0.01$). **B** Western blot for HIF1 α protein levels collected from HL-1 cells under normal and hypoxic conditions, using GAPDH for negative control. **C** Genome browser view showing H3K27ac ChIP-seq signals at the genomic regions around two mouse-specific IREs (E1 and E2) in mouse uninjured (red) and regenerating (blue) samples, as well as corresponding orthologous sequences in zebrafish. **D** Bar plots showing mRNA relative expression level of *Pde1a* (left) and *Gda* (right) under normal and hypoxic conditions, using *Gapdh* for normalization. Error bars denote SEM ($N=3$). P -value: unpaired Student's t -test (*, $P<0.05$). **E** Schematic of IRE sequences without mutation, and with ETS or AP-1 motif mutations (orange: ETS motif, yellow: AP-1 motif, grey: mutated sequence). Values of E1 and E2 length are listed below the graph. **F** Luciferase reporter assay for wildtype, ETS or AP-1 motif mutated mouse IREs under normal (orange) and hypoxic (blue) conditions. Empty vectors are used for negative control. Error bars denote SEM ($N=4$). P -values: unpaired Student's t -test (*, $P<0.05$; **, $P<0.01$; n.s., not significant). **G** Lollipop plot showing activation levels quantified by fold change of relative luciferase activity from **(D)** of two mouse IREs without mutation and with ETS or AP-1 motif mutations upon hypoxia, using the empty vector for normalization. **H** A model summarizing the inter-species variations between zebrafish and mouse IREs and their associated genes. Left, AP-1 and ETS motif frequencies of shared IREs in zebrafish and mice are similar, and nearby genes are up-regulated upon injury. Middle, orthologous sequences of zebrafish-specific IREs in mice are not activated by injury with reduction in motif frequency, and nearby genes show no changes in expression level. Right, for mouse-specific IREs, orthologous sequences in zebrafish are with decreased ETS motif frequency, but nearby genes are still with upregulation, which is mediated by IRE-independent mechanisms. Red pentagrams stand for ETS motifs, and the blue for AP-1 motifs. Rectangles with solid line indicate IREs identified through H3K27ac ChIP-seq data analysis, and rectangles with dotted line represent orthologous sequences of IREs. Black rectangles are IRE-associated genes or orthologs in the other species

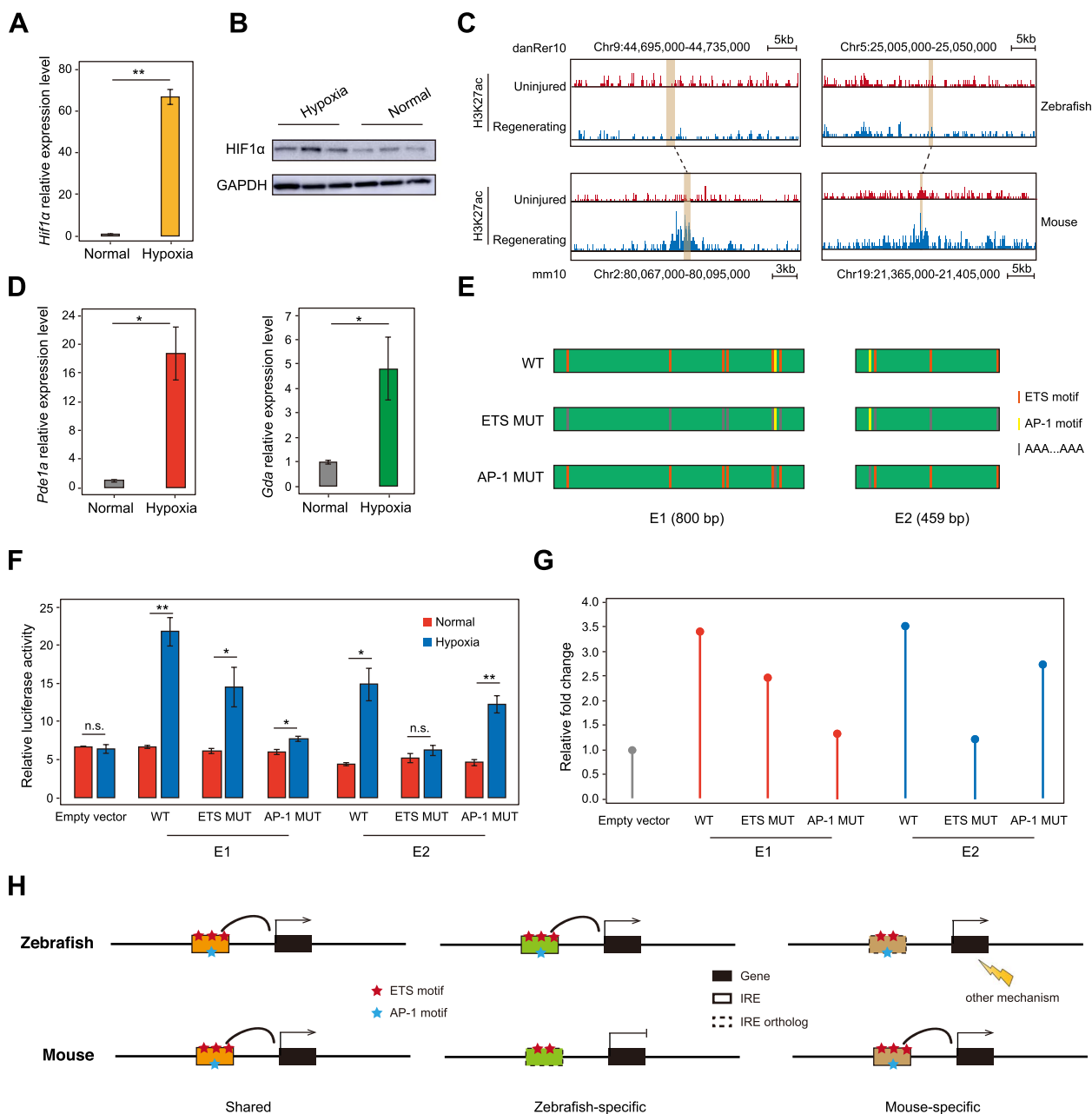


Fig. 6 (See legend on previous page.)

Discussion

Enhancer elements play critical roles in orchestrating the transcriptional landscapes in development and tissue homeostasis [45, 46]. Previous studies have demonstrated that a set of enhancers are activated in response to injury signals and promote tissue regeneration in organisms such as zebrafish [8, 24] and *Drosophila* [9]. However, the functional significances of IREs in mammals remain elusive. In the present study, we performed a comparative epigenomic analysis on cardiac IREs identified in

zebrafish versus mice, in which the cardiac regeneration capacity rapidly diminishes by postnatal day 7. We showed that enhancer elements activated upon injury in the regenerative mouse heart are also activated in the non-regenerative heart. Of note, the extent of IRE activation is significantly higher in P1 than P8 mouse heart, as well as up-regulation of IRE-associated genes, suggesting that IREs may partially contribute to cardiac regeneration in mice. These results are consistent with the findings that the epigenomic H3K27ac chromatin landscape

responds differently to injury in mouse heart with and without regenerative capacity [14]. We further demonstrated that the genes associated with, and potentially regulated by the IREs in zebrafish and mice vary considerably. We reason that while the IREs can be activated across a wide range of species including mammals, the regulatory functions of mammalian IREs likely shift more toward wound clearance and healing, rather than regeneration promotion.

The type, number, and combination of transcription factor binding sites (TFBSs) are critical features that govern enhancer functions [47]. Our analyses showed that the IREs are enriched for the AP-1 and ETS family TFBSs in both zebrafish and mice, suggesting an evolutionarily conserved mechanism of IRE activation across species. AP-1 transcription factor family is involved in modulating many cellular processes, including cell proliferation, differentiation and apoptosis [48]. Previous studies have shown that AP-1 binding sites are the most highly enriched motifs in zebrafish IREs activated during zebrafish heart or caudal fin regeneration, pinpointing their critical roles in the regeneration process [10, 49]. Apart from AP-1, we revealed that the ETS family transcription factor binding motifs are also enriched in both zebrafish and mouse cardiac IREs. ETS transcription factors are linked with various processes during embryonic development [50]. Of note, a fundamental property across regeneration is the reactivation of genes that are essential for embryogenesis [51]. Furthermore, the deletion of ETS-1, a member of the ETS family of transcription factors, results in abnormal ventricular morphology, indicating its important role in mammalian heart development [52]. Moreover, the interaction between AP-1 and ETS transcription factors in regulatory elements contributes to gene expression regulation [53, 54]. Therefore, it is tempting to speculate that AP-1 and ETS may act in a combinatorial or synergistic fashion to promote the activities of IREs upon cardiac injury. Apart from AP-1 and ETS, other factors, such as KLF1 [55], could also play a regulatory role in IRE-associated cardiac damage responses and regeneration. We anticipate that the future application of single-cell epigenomic profiling techniques, such as scATAC-seq (single-cell assay for transposase-accessible chromatin using sequencing), on the cardiac regeneration process would reveal more transcription factors responsible for the activation of IREs, and elucidate the functions of IREs in different cardiac cell types.

We further revealed that the inter-species variations in IREs between zebrafish and mice are intimately linked to the alterations in frequencies of AP-1 and ETS transcription factor binding motifs. Enhancers undergo rapid changes in sequence and turnover in function

during evolution [35]. Consistent with this notion, only a small number of IREs are retained from zebrafish to mice (Fig. 6H, left). Cross-species comparison between IREs and their orthologous sequences reveals that the zebrafish-specific IREs contain significantly higher frequencies of both AP-1 and ETS motifs than their non-injury-responsive orthologous sequences in mice (Fig. 6H, middle). In contrast, the mouse-specific IREs only exhibit higher frequencies of ETS motifs, but not AP-1 motifs, than their zebrafish orthologous sequences (Fig. 6H, right). In another word, gaining more ETS motifs within the ancestral IRE sequences is sufficient to render these IREs injury-responsive in mice. Using hypoxia treatment on mouse cardiomyocytes as an *in vitro* model to mimic myocardial ischemia *in vivo*, we showed that both AP-1 and ETS motifs contribute to the effective activation of IREs, suggesting that the gain or loss of either type of motifs could potentially lead to the functional turnover of IREs during evolution. However, more systematical loss-of-function and gain-of-function studies, ideally performed *in vivo*, are needed to unambiguously demonstrate the functional contribution of the AP-1 and ETS motifs in IRE activation and cardiac regeneration.

The variations in IREs between different species are linked with different transcriptome responses upon cardiac injury. Of particular note, the transcriptional activation for genes associated with zebrafish-specific IREs was dampened in mice, which may partially contribute to the reduced regeneration capacity in mice (Fig. 6H, middle). Interestingly, the genes associated with mouse-specific IREs were similarly upregulated in zebrafish, suggesting these genes may adopt a promoter-proximal mechanism rather than utilizing distal regulatory elements to achieve injury-induced transcription activation (Fig. 6H, right).

We note that several limitations associated with the currently available datasets may hinder the precise interpretation of the functions of IREs. First, the lack of stage-matched scRNA-seq and scATAC-seq datasets for mouse and zebrafish heart regeneration led us to use bulk genomic and epigenomic data in this study, which represent the averaged gene expression and enhancer activities of all cell types within the cardiac tissue compartments. This may lead to the underestimation of IRE-driven gene activation, and prevent us from assessing the cell-type-specific functions of IREs. Second, the transcriptomic and epigenetic profiling for both mouse and zebrafish hearts was only performed at limited time points post-injury. The lack of temporal resolution of the data may lead to the omission of transiently activated IREs, and also cause underestimation in the induction of IRE activity and gene expression.

Importantly, while most enhancers preferentially regulate target genes within their vicinity, enhancers are known to exert long-range regulatory effects by forming 3D interactions with their target genes. However, how 3D genome architecture changes during the process of regeneration and wound repair, and how the IREs interact with their targets have not been experimentally characterized. We anticipate that the application of 3C-based techniques, such as Hi-C [56], capture Hi-C [57] and HiChIP [58], would enable the generation of high-resolution chromatin interaction maps in the context of regeneration and the precise annotation of the regulatory functions of IREs, thereby enabling more accurate association of IREs with their target genes, and further promoting the understanding of the functional significances of IREs. In summary, our results suggest that the evolution of enhancer regulatory landscapes imposes a profound impact on transcriptional programs during tissue repair and regeneration across different species.

Conclusions

In this study, we performed comparative genomic analyses on IREs during heart regeneration in zebrafish and mice, and revealed that the inter-species variations in the distribution of AP-1 and ETS transcription factor binding motifs may play an important role in defining the functions of *cis*-regulatory elements during injury responses. Our findings are informative for understanding the molecular mechanisms of transcriptional remodeling in response to injury across different species, which will be potentially beneficial for developing new therapeutic strategies in regenerative medicine.

Materials and methods

ChIP-seq analysis

Neonatal mouse heart regeneration associated H3K27ac and H3K27me3 ChIP-seq data from GSE123867 [14] were aligned to mm10 genome assembly using bowtie2 v2.4.2 [59] with default parameters. Reads not uniquely mapped were discarded, and PCR duplicates were removed using Picard MarkDuplicates v2.25.2 (<http://broadinstitute.github.io/picard/>). Genomic tracks with bigwig format were generated using bamCoverage in deepTools v3.5.1 [60].

To identify IREs for P1 mouse hearts, we first merged bam files from two biological replicates (1.5 dpi ChIP or 1.5 dpi input), and used MACS2 v2.2.7.1 [61] to call peaks against the input control with the command: `macs2 callpeak -g mm -broad -broad-cutoff 1e-5`. Peaks were annotated using `annotatePeaks` in HOMER v4.11.1 [62], and those defined as promoters (-2 kb to +2 kb from TSS) were depleted. We then applied R package DiffBind v2.14.1 [63] to find peaks with significantly higher

H3K27ac ChIP-seq signals in 1.5 dpi than 1.5 dps. The cutoffs were fold change greater than 1.5 and false discovery rate (FDR) less than 0.05. Non-IREs were defined as peaks identified in P1 + 1.5dpi after excluding IREs.

H3K27ac ChIP-seq data of zebrafish heart regeneration from GSE75894 [8] were processed as described above with slight modifications. We set reference genome to the zebrafish genome (danRer10), and adjusted cutoffs for differential peak identification to fold change > 1.5 and p -value < 0.01.

To generate H3K27ac density plots of genomic regions of interest, we used `computeMatrix` and `plotProfile` or `plotHeatmap` from deepTools v3.5.1 [60]. Smoothed density plots were obtained with `geom_smooth()` function (`span = 0.3`) in R package ggplot2.

RNA-seq analysis

Single-end sequencing reads from RNA-seq during neonatal mouse heart regeneration from GSE123863 [14] were mapped to the mouse mm10 reference using HISAT2 v2.2.1 [64] with default settings. Uniquely aligned reads were extracted by `grep` command with `NH:i:1` tag. Genome browser tracks (bigwig format) were generated using `bamCoverage` from deepTools v3.5.1 [60]. Then we applied `featureCounts` v2.0.1 [65] to count reads with the parameter “-t exon -g gene_id”, and calculated transcripts per million (TPM) using StringTie v2.1.5 [66] to quantify gene expression with the parameter “-A”. TPM for gene or transcript i is defined as following formula:

$$TPM_i = \frac{q_i/l_i}{\sum_j(q_j/l_j)} \times 10^6$$

where q_i is read counts mapped to transcript, l_i denotes the transcript length, and $\sum_j(q_j/l_j)$ represents the sum of mapped reads to transcript normalized by transcript length. Fold change of gene expression level between injured samples and the control was calculated using R package DESeq2 v1.26.0 [67].

Zebrafish heart regeneration-associated RNA-seq data from GSE75894 [8] were analyzed as described above with changing reference genome to danRer10 assembly.

To define IRE-associated genes, we used the closest function from BEDTools v2.30.0 [68] to find the gene located nearest to each IRE in both zebrafish and mice.

For scRNA-seq data analysis, we first downloaded the count matrices of gene expression from GSE159032 for zebrafish [27] and GSE153480 for mice [28]. The matrices were scaled with factor 10,000 and normalized using `log1p` per cell (`log1p` returns the natural logarithm of a number plus a pseudo-count of 1, i.e. `log(1 + num)`). We

selected IRE-associated genes with significant up-regulation (fold change > 1.5, FDR < 0.1) upon cardiac injury identified in bulk RNA-seq data, and calculated the average fold change (injured vs. uninjured) of these genes at the single cell level under the guidance of cell type annotations. All cell types of interest were summarized into 10 types as follows: cardiomyocyte, endothelial cell, fibroblast, epicardial cell, perivascular cell, smooth muscle cell, macrophage, monocyte, T cell, and B cell. One gene was assigned to one cell type with the largest change in gene expression level. Heatmaps showing gene expression changes for different cell types upon injury were plotted using the pheatmap package.

Enhancer clustering analysis

We downloaded the H3K4me1 peak file from GSE82736 (https://ftp.ncbi.nlm.nih.gov/geo/series/GSE82736/GSE82736/suppl/GSE82736_ENCFF099IOG_peaks_mm10.bed.gz), and annotated these peaks by annotatePeaks in HOMER v4.11.1 [62]. Peaks that were not located at promoters (-2 kb to +2 kb from TSS) were defined as enhancers. Then we used computeMatrix and plotHeatmap from deepTools v3.5.1 [60] to perform clustering analysis for enhancers with the parameter “-kmeans 3”. All enhancers were classified into three types: poised enhancers (low H3K27ac, high H3K27me3), primed enhancers (low H3K27ac, low H3K27me3), and active enhancers (high H3K27ac, low H3K27me3).

To annotate the inactive IREs, we downloaded cCRE data from ENCODE Encyclopedia (https://api.wenglab.org/screen_v13/downloads/mm10-ccREs.bed) and extracted potential enhancer elements using the command “grep ELS” for further analysis.

Motif enrichment analysis

We used findMotifsGenome.pl from HOMER v4.11.1 [62] to perform motif enrichment analysis for both mouse and zebrafish IREs with the parameter “-size 500”. Genomic regions that match the GC-content distribution of the IRE sequences were selected as the background control. The transcription factor binding motifs with *p*-value less than 1×10^{-10} were considered as significantly enriched.

Motif frequency calculation

Genomic sequences of IREs were scanned to find and count AP-1 and ETS transcription factor binding motifs using a custom python script. The International Union of Pure and Applied Chemistry (IUPAC) codes for AP-1 and ETS motifs summarized by results of motif enrichment analysis are as follows: AP-1

(TKASTMA), and ETS (VVGGAWVY). For 14 transcription factors from ETS family, IUPAC codes are listed below: EHF (DNVMGGAAR), ELF1 (SCGGAAGY), ELF3 (WNVMGGAARY), ELF4 (VMGGAAG), ELF5 (WNVMGGAAGT), ERG (CMGGAARY), ETS1 (RVMGGAARY), ETV1 (CMGGAWG), ETV2 (VMGGAWR), ETV4 (SMGGAWRB), FLI1 (VMGGAWR), GABP α (VMGGAAG), PU.1 (DVRGGAAGTG), and SPIB (DWDVRGAAVYS). IUPAC codes for 7 transcription factors from AP-1 family are as follows: ATF3 (VTKANTCAB), BATF (TKASTMA), FOS (TKANTYA), FRA1 (VTKANTMAB), FRA2 (TKWSYMA), JUN (TGASTCA), and JUNB (TKASTCA). Motif frequency was calculated as

$$f = \frac{1000 \times c}{s}$$

where *f* is motif frequency (per kilobase), *c* represents motif counts, and *s* is sequence size of the IRE.

To reveal relationship between motif frequency and binding sites of transcription factor, we randomly generated 10,000 genomic sequences with the length of 1 kb using random and getfasta function from BEDTools v2.30.0 [68]. Peaks of AP-1 or ETS transcription factors in different types of cell lines were obtained from ReMap database [69]. To generate random peaks as control, we used shuffle function from BEDTools v2.30.0 [68]. The relative overlapping ratio was calculated using the following formula:

$$r = \frac{n_i}{n_{i=0}}$$

where *r* is relative overlapping ratio, *n_i* denotes the number of sequence with certain motif frequency intersecting with true (or random) peaks, and *n_{i=0}* represents the number of sequence without AP-1 or ETS motif overlapping with true (or random) peaks. The intersection was performed using intersect function from BEDTools v2.30.0 [68].

Functional enrichment analysis

To compare biological functions of IRE-associated genes, including GO biological process (BP) and KEGG pathway, between mice and zebrafish as comprehensively as possible, we used Metascape [70] with following parameters: min overlap = 1, *p*-value cutoff = 1 and min enrichment = 1. All genes were set as the enrichment background. Terms with *q*-value less than 0.05 were considered significantly enriched.

Orthologous enhancers prediction

To characterize orthologous sequences of zebrafish IREs in mice, we first identified orthologs of zebrafish IRE-associated genes in mice using Ensembl BioMart [71]. If a given gene in zebrafish has more than one ortholog in mice, we only select the ortholog with the highest similarity quantified by percentage of query gene identical to target gene. Genomic sequences around TSSs of orthologs (from -100 kb to +100 kb) were extracted using getfasta function from BEDTools v2.30.0 [68]. Pairwise alignment between each zebrafish IRE and the genomic sequence around corresponding orthologous gene was performed by EMBOSS Needle v6.6.0.0 [72] with the parameters “-gapopen 10.0 -gapextend 0.5 -aformat3 markx3”. Zebrafish and mouse reference genomes, danRer10 (chr1-25) and mm10 (chr1-19 and chrX) assemblies, were used.

Prediction of orthologous genes or enhancers in zebrafish of mouse IREs was performed with the same procedure as described above.

Evolutionary conservation analysis

The evolutionary conservation scores quantified by phastCons60way [73] or phyloP60way [74] were retrieved from UCSC genome browser data portal (<http://genome.ucsc.edu>) as bigwig files. The average conservation scores in 1 kb centering IRE or non-IRE midpoint were computed using subcommand computeMatrix and plotted by plotProfile from deepTools v3.5.1 [60].

Cell culture

HL-1 mouse cardiomyocytes (a kind gift from Prof. Alex Chang, Shanghai Institute of Precision Medicine, China) were cultured in RPMI 1640 medium (Catalog No. 11875093, Thermo Fisher Scientific) containing 10% fetal bovine serum (FBS) (Catalog No. 12103C, Sigma-Aldrich), 100 U/mL penicillin, and 0.1 mg/mL streptomycin in 5% CO₂ at 37 °C.

Induction of hypoxia

HL-1 cells (Catalog No. BNCC288890, BeNa Culture Collection) were grown under standard conditions to approximately 80% confluency, and maintained in a quiescent state for 12 h. To induce hypoxia, HL-1 cells were transferred to a chamber (MIC-101, Billups-Rothenberg) and flushed with a gas mixture of 95% N₂ and 5% CO₂ for 24 h.

Quantitative reverse transcription PCR (qRT-PCR)

Total RNA was isolated with TRIzol Reagent (Catalog No. 15596018, Thermo Fisher Scientific) according to manufacturer's instructions, and cDNA was synthesized with PrimeScript RT Master Mix Kit (Catalog No. RR036A,

TaKaRa). Real time qPCR was performed using LightCycler 480 II (Roche). We used $\Delta\Delta CT$ method to quantify relative gene expression level, and *Gapdh* was as endogenous control. Primer sequences are as follows (from 5' to 3'): *Hif1 α* forward primer: GTTACGATTGTGAAGTTA, reverse primer: AAGGAATGAGATTAGGAA; *Pde1a* forward primer: ATCTTATCAACCGCTTCAA, reverse primer: TGCTGTAACCAACTTCTAA; *Gda* forward primer: AGGAATAGCAGTGGTAAT, reverse primer: AACAAAGCATAGGTAACATT; *Gapdh* forward primer: CCTGGTCACCAGGGCTGC, reverse primer: CGC TCCTGGAAGATGGTGATG.

Western blotting

The protein extracts were prepared in RIPA lysis buffer containing 100 U protease inhibitors (Catalog No. 20101ES60, Yeasen Biotech). Samples were run on 10% SDS-PAGE gels, and then transferred to PVDF membranes (IPVH00010, Millipore). The membranes were blocked with 5% no-fat milk in TBS with 0.1% Tween 20, and incubated with primary antibodies (anti-HIF1 α : Catalog No. 36169, Cell Signaling Technology; anti-GAPDH: Catalog No. 5174, Cell Signaling Technology) at 4 °C overnight. HRP-conjugated goat anti-rabbit IgG (Catalog No. 33101ES60, Yeasen Biotech) secondary antibodies were used. The signals were recorded with ECL reagents (Millipore ECL plus kit), and visualized using an ImageQuant LAS4000 (GE Healthcare).

Luciferase reporter assay

Luciferase reporter plasmid constructs cloned with synthetic enhancer sequences were confirmed by Sanger sequencing. HL-1 cells were transfected with related plasmid using Lipo8000 (Catalog No. C0533, Beyotime). Firefly and renilla luciferase activity were measured by Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions. Renilla luciferase was used as an internal control to facilitate normalization for transfection efficiency. For each enhancer candidate and construct, we used the average of the replicates as the final activity together with the standard error of mean (SEM). Enhancer activation level quantified by relative fold change of average relative luciferase activity was calculated as follows:

$$FC = \frac{s_i(H) / s_i(N)}{s_c(H) / s_c(N)}$$

where *FC* denotes relative activation level of the sequence (enhancer, enhancer with mutations, or empty vector), *s_i(H)* and *s_i(N)* are average relative luciferase signal of samples under hypoxia and normal condition, respectively. Similarly, *s_c(H)* and *s_c(N)* are mean relative luciferase activity of samples with control (empty vector) under hypoxic and normal condition.

Abbreviations

IREs	Injury-responsive enhancers
ChIP-seq	Chromatin immunoprecipitation sequencing
RNA-seq	RNA sequencing
MI	Myocardial infarction
CMS	Cardiomyocytes
LAD	Left anterior descending
dpi	Days post-injury
dps	Days post-sham
TSSs	Transcriptional start sites
cCRE	Candidate <i>cis</i> -regulatory element
MEA	Motif enrichment analysis
scRNA-seq	Single-cell RNA sequencing
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
TFBSs	Transcription factor binding sites
scATAC-seq	Single-cell assay for transposase-accessible chromatin using sequencing
FDR	False discovery rate
TPM	Transcripts per million
IUPAC	The International Union of Pure and Applied Chemistry
BP	Biological Process
SEM	Standard error of mean

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09283-8>.

Additional file 1: Fig. S1. Comparison of enhancer landscapes between mouse P1 and P8 hearts.

Additional file 2: Fig. S2. Motif associated analyses of zebrafish and mouse cardiac IREs.

Additional file 3: Fig. S3. Comparison of IRE-gene pairs between zebrafish and mice.

Additional file 4: Fig. S4. Analyses of IRE-associated genes involved in different biological functions between zebrafish and mice.

Additional file 5: Fig. S5. Sequence conservation and motif frequency analyses of zebrafish and mouse cardiac IREs.

Additional file 6: Fig. S6. Two mouse IREs tested in HL-1 cells with hypoxia treatment.

Additional file 7: Table S1. Coordinates of IREs and non-IREs in neonatal P1 mouse hearts.

Additional file 8: Table S2. Locations of IREs identified during zebrafish heart regeneration.

Additional file 9: Table S3. Results of motif enrichment analysis for zebrafish and mouse IREs.

Additional file 10: Table S4. Significantly enriched GO terms for IRE-associated genes that have orthologous genes in both zebrafish and mice.

Additional file 11: Table S5. Significantly enriched KEGG pathways for IRE-associated genes that have orthologous genes in both zebrafish and mice.

Additional file 12: Table S6. Three types of genes associated with IREs in zebrafish and mice.

Additional file 13: Table S7. Significantly enriched GO terms for three types of IRE-associated genes in zebrafish and mice.

Additional file 14: Table S8. Significantly enriched KEGG pathways for three types of IRE-associated genes in zebrafish and mice.

Additional file 15: Table S9. Coordinates of three classes of zebrafish and mouse cardiac IREs.

Additional file 16: Table S10. Public datasets analyzed in this paper.

Acknowledgements

We are grateful to Prof. Alex Chang for providing HL-1 cell line and helping with induction of hypoxia, and thank members of Bian lab for constructive criticism and generative discussions.

Authors' contributions

LJ designed the experiments, conducted the computational analysis, and wrote the paper. YS conducted the experiments, analyzed the data and wrote the paper. QB conceived the study, designed the experiments, and wrote the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (31970585, 32170544, and 31801056) and Natural Science Foundation of Shanghai (23ZR1435700). QB is also supported by The Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (TP2018044), and the Innovative Research Team of High-Level Local Universities in Shanghai (SHSMU-ZLCX20211700).

Availability of data and materials

This paper analyses publicly available data. These accession numbers for the datasets are listed and summarized in Table S10.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 2 December 2022 Accepted: 29 March 2023

Published online: 05 April 2023

References

- Townsend N, Kazakiewicz D, Lucy Wright F, Timmis A, Huculeci R, Torbica A, et al. Epidemiology of cardiovascular disease in Europe. *Nat Rev Cardiol.* 2022;19:133–43.
- Witman N, Murtuza B, Davis B, Arner A, Morrison JL. Recapitulation of developmental cardiogenesis governs the morphological and functional regeneration of adult newt hearts following injury. *Dev Biol.* 2011;354:67–76.
- Gemberling M, Bailey TJ, Hyde DR, Poss KD. The zebrafish as a model for complex tissue regeneration. *Trends Genet.* 2013;29:611–20.
- Uygur A, Lee RT. Mechanisms of Cardiac Regeneration. *Dev Cell.* 2016;36:362–74.
- Sun X, Chuang JC, Kanchwala M, Wu L, Celen C, Li L, et al. Suppression of the SWI/SNF Component Arid1a Promotes Mammalian Regeneration. *Cell Stem Cell.* 2016;18:456–66.
- Cui M, Wang Z, Bassel-Duby R, Olson EN. Genetic and epigenetic regulation of cardiomyocytes in development, regeneration and disease. *Development.* 2018;145:dev171983.
- Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet.* 2011;12:283–93.
- Kang J, Hu J, Karra R, Dickson AL, Tornini VA, Nachtrab G, et al. Modulation of tissue repair by regeneration enhancer elements. *Nature.* 2016;532:201–6.
- Vizcaya-Molina E, Klein CC, Serras F, Mishra RK, Guigo R, Corominas M. Damage-responsive elements in *Drosophila* regeneration. *Genome Res.* 2018;28:1852–66.
- Wang W, Hu CK, Zeng A, Alegre D, Hu D, Gotting K, et al. Changes in regeneration-responsive enhancers shape regenerative capacities in vertebrates. *Science.* 2020;369:eaaz3090.

11. Porrello ER, Mahmoud AI, Simpson E, Hill JA, Richardson JA, Olson EN, et al. Transient regenerative potential of the neonatal mouse heart. *Science*. 2011;331:1078–80.
12. Laflamme MA, Murray CE. Heart regeneration. *Nature*. 2011;473:326–35.
13. Porrello ER, Mahmoud AI, Simpson E, Johnson BA, Grinsfelder D, Canseco D, et al. Regulation of neonatal and adult mammalian heart regeneration by the miR-15 family. *Proc Natl Acad Sci U S A*. 2013;110:187–92.
14. Wang Z, Cui M, Shah AM, Ye W, Tan W, Min YL, et al. Mechanistic basis of neonatal heart regeneration revealed by transcriptome and histone modification profiling. *Proc Natl Acad Sci U S A*. 2019;116:18455–65.
15. Cui M, Atmanli A, Morales MG, Tan W, Chen K, Xiao X, et al. Nr1 promotes heart regeneration and repair by regulating proteostasis and redox balance. *Nat Commun*. 2021;12:5270.
16. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
17. Consortium EP, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopedia of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699–710.
18. Dheilly E, Battistello E, Katanayeva N, Sungalee S, Michaux J, Duns G, et al. Cathepsin S Regulates Antigen Processing and T Cell Activity in Non-Hodgkin Lymphoma. *Cancer Cell*. 2020;37(674–89): e12.
19. Dhawan P, Richmond A. Role of CXCL1 in tumorigenesis of melanoma. *J Leukoc Biol*. 2002;72:9–18.
20. MacGrogan D, Munch J, de la Pompa JL. Notch and interacting signalling pathways in cardiac development, disease, and regeneration. *Nat Rev Cardiol*. 2018;15:685–704.
21. Wu CL, Yin R, Wang SN, Ying R. A Review of CXCL1 in Cardiac Fibrosis. *Front Cardiovasc Med*. 2021;8:674498.
22. Tan Y, Duan X, Wang B, Liu X, Zhan Z. Murine neonatal cardiac B cells promote cardiomyocyte proliferation and heart regeneration. *NPJ Regen Med*. 2023;8:7.
23. Poss KD, Wilson LG, Keating MT. Heart regeneration in zebrafish. *Science*. 2002;298:2188–90.
24. Goldman JA, Kuzu G, Lee N, Karasik J, Gemberling M, Foglia MJ, et al. Resolving Heart Regeneration by Replacement Histone Profiling. *Dev Cell*. 2017;40(392–404): e5.
25. Garces de Los Fayos Alonso I, Liang HC, Turner SD, Lager S, Merkel O, Kenner L. The Role of Activator Protein-1 (AP-1) Family Members in CD30-Positive Lymphomas. *Cancers*. 2018;10:93.
26. Oikawa T, Yamada T. Molecular biology of the Ets family of transcription factors. *Gene*. 2003;303:11–34.
27. Hu B, Lelek S, Spanjaard B, El-Sammak H, Simoes MG, Mintcheva J, et al. Origin and function of activated fibroblast states during zebrafish heart regeneration. *Nat Genet*. 2022;54:1227–37.
28. Wang Z, Cui M, Shah AM, Tan W, Liu N, Bassel-Duby R, et al. Cell-Type-Specific Gene Regulatory Networks Underlying Murine Neonatal Heart Regeneration at Single-Cell Resolution. *Cell Rep*. 2020;33:108472.
29. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, et al. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res*. 2019;47:D729–35.
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–9.
31. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
32. Hortells L, Johansen AKZ, Yutzey KE. Cardiac Fibroblasts and the Extracellular Matrix in Regenerative and Nonregenerative Hearts. *J Cardiovasc Dev Dis*. 2019;6:29.
33. Simoes FC, Riley PR. Immune cells in cardiac repair and regeneration. *Development*. 2022;149:dev199906.
34. Zuppo DA, Tsang M. Zebrafish heart regeneration: Factors that stimulate cardiomyocyte proliferation. *Semin Cell Dev Biol*. 2020;100:3–10.
35. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. *Cell*. 2015;160:554–66.
36. Hegarty SV, Sullivan AM, O’Keeffe GW. Zeb2: A multifunctional regulator of nervous system development. *Prog Neurobiol*. 2015;132:81–95.
37. Gladka MM, Kohela A, Molenaar B, Versteeg D, Kooijman L, Monshouwer-Kloots J, et al. Cardiomyocytes stimulate angiogenesis after ischemic injury in a ZEB2-dependent manner. *Nat Commun*. 2021;12:84.
38. Begeman IJ, Shin K, Osorio-Mendez D, Kurth A, Lee N, Chamberlain TJ, et al. Decoding an organ regeneration switch by dissecting cardiac regeneration enhancers. *Development*. 2020;147:dev194019.
39. Ganguly R, Mohyeldin A, Thiel J, Kornblum HI, Beullens M, Nakano I. MELK-a conserved kinase: functions, signaling, cancer, and controversy. *Clin Transl Med*. 2015;4:11.
40. Saito Y, Li L, Coyaud E, Luna A, Sander C, Raught B, et al. LGL2 rescues nutrient stress by promoting leucine uptake in ER(+) breast cancer. *Nature*. 2019;569:275–9.
41. Romero-Ramirez L, Cao H, Nelson D, Hammond E, Lee AH, Yoshida H, et al. XBP1 is essential for survival under hypoxic conditions and is required for tumor growth. *Cancer Res*. 2004;64:5943–7.
42. Claycomb WC, Lanson NA Jr, Stallworth BS, Egeland DB, Delcarpio JB, Bahinski A, et al. HL-1 cells: a cardiac muscle cell line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci U S A*. 1998;95:2979–84.
43. Kuznetsov AV, Javadov S, Sickinger S, Frotschnig S, Grimm M. H9c2 and HL-1 cells demonstrate distinct features of energy metabolism, mitochondrial function and sensitivity to hypoxia-reoxygenation. *Biochim Biophys Acta*. 2015;1853:276–84.
44. Lee JW, Bae SH, Jeong JW, Kim SH, Kim KW. Hypoxia-inducible factor (HIF-1)alpha: its protein stability and biological functions. *Exp Mol Med*. 2004;36:1–12.
45. Huang J, Liu X, Li D, Shao Z, Cao H, Zhang Y, et al. Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev Cell*. 2016;36:9–23.
46. Cai W, Huang J, Zhu Q, Li BE, Seruggia D, Zhou P, et al. Enhancer dependence of cell-type-specific gene expression increases with developmental age. *Proc Natl Acad Sci U S A*. 2020;117:21450–8.
47. Jindal GA, Farley EK. Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev Cell*. 2021;56:575–87.
48. Shaulian E, Karin M. AP-1 as a regulator of cell life and death. *Nat Cell Biol*. 2002;4:E131–6.
49. Beisaw A, Kuenne C, Guenther S, Dallmann J, Wu CC, Bentsen M, et al. AP-1 Contributes to Chromatin Accessibility to Promote Sarcomere Disassembly and Cardiomyocyte Protrusion During Zebrafish Heart Regeneration. *Circ Res*. 2020;126:1760–78.
50. Sharrocks AD. The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol*. 2001;2:827–37.
51. Goldman JA, Poss KD. Gene regulatory programmes of tissue regeneration. *Nat Rev Genet*. 2020;21:511–25.
52. Ye M, Coldren C, Liang X, Mattina T, Goldmuntz E, Benson DW, et al. Deletion of ETS-1, a gene in the Jacobsen syndrome critical region, causes ventricular septal defects and abnormal ventricular morphology in mice. *Hum Mol Genet*. 2010;19:648–56.
53. Bergelson S, Daniel V. Cooperative interaction between Ets and AP-1 transcription factors regulates induction of glutathione S-transferase Ya gene expression. *Biochem Biophys Res Commun*. 1994;200:290–7.
54. Bassuk AG, Leiden JM. A direct physical association between ETS and AP-1 transcription factors in normal human T cells. *Immunity*. 1995;3:223–37.
55. Ogawa M, Geng FS, Humphreys DT, Kristianto E, Sheng DZ, Hui SP, et al. Kruppel-like factor 1 is a core cardiomyogenic trigger in zebrafish. *Science*. 2021;372:201–5.
56. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289–93.
57. Hughes JR, Roberts N, McGowan S, Hay D, Giannoulitou E, Lynch M, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet*. 2014;46:205–12.
58. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*. 2016;13:919–22.
59. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
60. Ramirez F, Dunder F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014;42:W187–91.
61. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.

62. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576–89.
63. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481:389–93.
64. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12:357–60.
65. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
66. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–5.
67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
68. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
69. Hammal F, de Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res*. 2022;50:D316–25.
70. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523.
71. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)*. 2011;2011:bar030.
72. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res*. 2015;43:W580–4.
73. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15:1034–50.
74. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of non-neutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20:110–21.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

