Identification and requirements of enhancers that direct gene expression during zebrafish fin regeneration

John D. Thompson1,2, Jianhong Ou1, Nutishia Lee1,2, Kwangdeok Shin3, Valentina Cigliola1,2, Lingyun Song4, Gregory E. Crawford4, Junsu Kang3 and Kenneth D. Poss1,2,*

ABSTRACT
To identify candidate tissue regeneration enhancer elements (TREEs) important for zebrafish fin regeneration, we performed ATAC-seq from bulk tissue or purified fibroblasts of uninjured and regenerating caudal fins. We identified tens of thousands of DNA regions from each sample type with dynamic accessibility during regeneration, and assigned these regions to proximal genes with corresponding expression changes by RNA-seq. To determine whether these profiles reveal bona fide TREEs, we tested the sufficiency and requirements of several sequences in stable transgenic lines and mutant lines with homozygous deletions. These experiments validated new non-coding regulatory sequences near induced and/or essential genes during fin regeneration, including fgf20a, mdka and cx43, identifying distinct domains of directed expression for each confirmed TREE. Whereas deletion of the previously identified LEN enhancer abolished detectable induction of the nearby leptin b gene during regeneration, deletions of enhancers linked to fgf20a, mdka and cx43 had no effect or partially reduced gene expression. Our study generates a new resource for dissecting the regulatory mechanisms of appendage generation and reveals a range of requirements for individual TREEs in control of regeneration programs.

KEY WORDS: Chromatin, Enhancers, Fin regeneration, Genome-wide profiling, Regeneration, Zebrafish

INTRODUCTION
Adult teleosts like zebrafish and urodele salamanders can regenerate entire amputated fins and limbs. By contrast, regenerative healing of adult mammalian limbs is limited to the very tips of digits (Borgens, 1982; Morgan, 1901; Spallanzani, 1768). Illuminating concepts and targets of appendage regeneration will decipher how and why tissue regeneration occurs and inform strategies to improve human tissue repair.

Cell transplantation, genetic fate mapping and live-imaging strategies have been employed across many species and tissues over the past 15–20 years to answer cell-of-origin questions in models of regeneration from planarians to mice (Dor et al., 2004; Kikuchi et al., 2010; Kragl et al., 2009; Wagner et al., 2011). Relevant to appendage regeneration, the salamander limb, zebrafish fin or murine digit tip blastema is no longer considered a homogeneous mass of pluripotent cells, but rather as a composite of lineage-restricted cell compartments (Knopf et al., 2011; Kragl et al., 2009; Rinkevich et al., 2011; Tornini et al., 2016; Tu and Johnson, 2011). A second area of expansion has been the study of traditional developmental regulators such as Fgf, Wnt, Bmp, Notch, Igf and Hh proteins during regeneration (Payzin-Dogru and Whited, 2018; Sehring and Weidinger, 2020; Seifert and Muneoka, 2018). Because of their many roles during embryonic development, and as many labs possess expertise and tools that probe these pathways, these have been logical choices in which to establish early regulatory models of appendage regeneration.

Mammalian genomes likely encode all gene products required to regenerate an amputated limb, yet they lack the correct instructions for strategically modulating those gene products to accomplish limb regeneration. Certain non-mammalian vertebrates such as salamanders and zebrafish possess these instructions, which exist as cis-regulatory elements. Although the catalogue of defined cell dynamics and molecular factors in tissue regeneration is expanding, we know comparatively little about how genes involved in regenerative events are engaged upon injury. That is, what are the non-coding DNA sequences that recruit activators (or repressors) of gene programs during tissue regeneration, how are these sequences distributed throughout genomes of regeneration-competent animals and which transcription factors engage with these sequences? Regulatory sequences are key suspected targets for evolutionary change, and they are likely to provide insight into how regenerative capacity is lost or maintained during species evolution (Goldman and Poss, 2020; King and Wilson, 1975; Wittkopp and Kalay, 2011; Yang and Kang, 2019).

Distal-acting enhancer regulatory sequences can direct expression of their target genes and have been predominantly studied as a means for stage- and tissue-specific regulation during embryonic development. Recently, we identified an enhancer linked to the leptin b (lepb) gene, named LEN, that activates gene expression during fin and heart regeneration in zebrafish, revealing a new class of gene regulatory sequences we named tissue regeneration enhancer elements (TREEs) (Kang et al., 2016). TREEs are identifiable by chromatin profiling, which has revealed or suggested similar classes of regulatory elements in Drosophila imaginal wing discs (Harris et al., 2016; Harris et al., 2020; Vizcaya-Molina et al., 2018), acelo worms (Gehrke et al., 2019), mouse bone (Guenther et al., 2015) and zebrafish hearts (Goldman et al., 2017). In addition to informing regulatory mechanisms of regeneration, TREEs can be engineered in simple constructs to spatiotemporally target developmental factors to injured tissue and boost or inhibit regenerative events (Kang et al., 2016).

Here, we have profiled genome-wide changes in chromatin accessibility (ATAC-seq) and expression (RNA-seq) during...
regeneration of amputated zebrafish fins, assessing both whole tissue samples and purified fibroblasts from regenerating fins. These experiments revealed thousands of putative TREES, several of which we functionally characterized in new transgenic and mutant zebrafish lines. This study provides a resource and framework for the discovery and dissection of dynamic chromatin structure and enhancer elements during appendage regeneration.

RESULTS AND DISCUSSION

ATAC-seq and RNA-seq profiling of whole-fin regenerates

Immediately after amputation of a zebrafish caudal fin, nearby epidermal cells begin to migrate to cover the lesion site and form a wound epidermis, alternatively called a regeneration epidermis. This specialized epidermis matures and releases signals to the underlying stump cells to form a blastema, a proliferative mass of mainly fibroblasts and osteoblasts that act as progenitors for new structures (Akimenko et al., 2003; Ando et al., 2017; Johnson and Weston, 1995; McMillan et al., 2018; Poss et al., 2000; Sehring and Weidinger, 2020). By 2 days post amputation (dpa), each fin ray has formed a blastema. By 4 dpa, regenerates possess a major compartment of fibroblasts, in addition to a minor compartment of newly aligned osteoblasts in the process of depositing new ray matrix. Complete regeneration of new structures occurs within 2-3 weeks of amputation (Fig. 1A,B).

To chart dynamic chromatin changes occurring during zebrafish fin regeneration during a stage of rapid cell division with concomitant skeletal patterning, we performed an assay for transposase-accessible chromatin followed by high-throughput sequencing (ATAC-seq) and RNA-seq on freshly amputated (0 dpa) and regenerating (4 dpa) fin tissue. We achieved an average of 55 million total reads per ATAC-seq library in this study, varying between 31 and 147 million reads for each of the libraries reported here, with an average of 24 million unique reads per library (Tables S1, S2, S3). First, to survey heterogeneous cell populations, we collected all cells from the regenerating regions of the fin plus one ray segment proximal to the amputation plane (Fig. 1C). We found 4265 transcripts differentially expressed and 50,873 chromatin regions differentially accessible at 4 dpa compared with 0 dpa (Fig. S1A; Tables S4, S5). We then plotted 13,892 of these differential ATAC-seq peaks to the closest differentially expressed transcript to assess genes potentially modulated by putative regulatory sequences during regeneration (Fig. 1D,E; Table S6). Most sequences displaying increased accessibility during regeneration were in intergenic regions (46.0%), 3 kb upstream of transcriptional start sites (25.7%) or within introns (18.0%). The remaining regions were located in UTRs, exons or immediate downstream regions of genes (10.3%) (Fig. S1B). ChromVar analysis indicated that the most variable transcription factor binding sites at 4 dpa compared to 0 dpa are predicted binding sites of components of the AP-1 complex, known to promote cell proliferation and survival (Fig. S1C) (Karin et al., 1997; Schep et al., 2017). To examine differential transcription factor binding accessibility, footprinting analysis was performed on a subset of transcription factor consensus sequences from our differential 4 dpa ATAC dataset, revealing minor changes in accessibility near the selected sequences FOS:JUN, EPAS1 and ZNF263 (Fig. S1D-F). To identify upstream regulators and canonical pathways involved in fin regeneration, we analyzed these RNA-seq data using ingenuity pathway analysis (IPA) (Kramer et al., 2014). The most highly activated upstream regulators were ERBB2, an EGFR co-receptor implicated in zebrafish fin regeneration (Rojas-Munoz et al., 2009), and MYC, a transcription factor central to cell growth and proliferation (Meyer and Penn, 2008) implicated in regeneration of liver, Müller glia and lateral line neuromasts in zebrafish (Jagtap et al., 2020; Lee et al., 2016; Mitra et al., 2019). The most downregulated upstream factor was TP53, a major cell cycle and apoptotic regulator that was described to be suppressed to facilitate salamander limb regeneration (Fig. S1G; Table S7) (Yun et al., 2013). Thyroid hormone metabolism, which has been reported to impact fin regeneration in medaka (Sekimizu et al., 2007), was predicted by IPA to be the most increased canonical pathway in our dataset, whereas melatonin degradation was predicted to be one of the most downregulated pathways (Fig. S1H; Table S8).

To assess gene regulation during an earlier stage – active formation of the initial regeneration blastema – we profiled chromatin accessibility and RNA levels at 1 dpa compared with 0 dpa, and found 5400 transcripts differentially expressed and 25,398 chromatin regions differentially accessible (Fig. S2A; Tables S9, S10). We were able to merge the ATAC-seq and RNA-seq data by pairing 8720 dynamic chromatin regions to nearby differentially expressed genes (Fig. 1F,G; Table S11). Regions more accessible at 1 dpa were again mainly intergenic (45.4%), within 3 kb upstream of the start site (27.2%), or intronic (17.5%), with the remaining 9.9% in UTRs, exons or immediate downstream regions of genes (Fig. S2B). Of the 4265 transcripts differentially expressed at 4 dpa compared with 0 dpa, 61.7% were also differentially expressed at 1 dpa (Fig. 1H). Similar to in the 4 dpa dataset, ChromVar analysis revealed that the most variable transcription factor binding sites at 1 dpa compared with 0 dpa implicate the AP-1 complex (Fig. S2C) (Karin et al., 1997; Schep et al., 2017). Footprinting analysis near the selected transcription factor consensus sequences FOS:JUN, EPAS1 and ZNF263 again revealed accessibility changes (Fig. S2D-F). Upstream factor analysis of the differential 1 dpa RNA-seq dataset by IPA showed similar results as the differential 4 dpa dataset, with upregulated upstream factors MYC and ERBB2, and the downregulated upstream factor TP53 (Fig. S2G; Table S12). IPA of canonical pathways upregulated at 1 dpa showed an increase in stress response, DNA damage and DNA replication pathways (Fig. S2H; Table S13). We interpret the overlap in differential transcript expression to include stable regenerative programs that are initiated at or around 1 dpa. Differences may reflect transient programs like injury response and blastema formation likely to be enriched at 1 dpa, as well as programs underlying patterning that would be initiated after this stage.

Comparative analysis of merged differential ATAC peaks paired with nearby differential transcripts revealed 1178 transcripts with nearby differential peaks shared between the 1 dpa and 4 dpa dataset, with 2981 unique to the 1 dpa dataset and 4637 unique to the 4 dpa dataset (Fig. S2I). GO Term analysis of the shared transcripts revealed involvement in a variety of regenerative and developmental processes, such as cartilage, connective tissue, skeletal system and pectoral fin development, as well as appendage, epithelial and embryonic organ morphogenesis (Fig. S2J). The 1 dpa dataset showed greater enrichment with dynamic accessibility near differential transcripts involved in neuronal, mesenchymal and cell differentiation, whereas the 4 dpa dataset had greater enrichment with dynamic accessibility near differential transcripts involved in cell cycle processes such as DNA replication, translation and mitotic cell cycle processes (Fig. S2K,L).

TREES near fgf20a and mdka control distinct epidermal expression domains

To examine putative TREES indicated by the dataset using functional tests, we employed two distinct strategies: one method to test sufficiency and one method to test requirement. To assess the
sufficiency of noncoding sequences from our ATAC-seq datasets to
direct gene expression, we used a transgenic enhancer reporter system
with the candidate sequence inserted immediately upstream of a 95-
bp minimal mouse c-fos promoter and EGFP gene cassette (Goldman
et al., 2017) (Fig. 2A). To examine the importance of these same
regions in their natural genomic context, we used CRISPR to excise
DNA containing the putative cis-regulatory element and generate
homozygous mutant deletion lines. We prioritized putative TREE
regulatory regions that were near genes previously implicated in fin
growth and regeneration, or involved in other regeneration contexts.

Fig. 1. Chromatin accessibility profiling of regenerating zebrafish fins. (A) By 4 dpa, the fin regeneration blastema is formed, and proliferating fibroblasts and
osteoblasts actively restore lost mesenchymal tissue, extracellular matrix and bone. Zebrafish are able to fully regenerate their fins by 2-3 weeks after injury.
(B) Immediately following fin ray amputation, epidermal cells migrate to cover the site of injury and signal to the underlying mesenchyme, which forms a
regeneration blastema. The blastema is formed by 2 dpa, and osteoblasts align at the basal epidermal layer and begin to deposit matrix by 3-4 dpa. (C) Overview
of sample collection and enzymatic dissociation for ATAC-seq and RNA-seq profiling of 0 dpa, 1 dpa and 4 dpa fin tissue. Uninjured samples were collected from
~1.5 mm portions of tissue from the middle region of fins, while regenerating samples were collected at 1 or 4 dpa from the regenerating tissue and one joint
proximal to the amputation plane. Each sample contained 15 pooled fin clips. (D) Dot plot of differential ATAC-seq chromatin regions linked to nearby differential
transcripts in 4 dpa versus 0 dpa whole-fin samples. Each point indicates a separate ATAC-seq peak. (E) Heat map of increased transcripts in 4 dpa versus 0 dpa
regenerating whole-fin tissue linked to nearby differentially accessible chromatin regions. Uninjured 0 dpa samples for both 4 dpa and 1 dpa datasets were used to
consider batch effects when performing differential comparison. (F) Dot plot of differential ATAC-seq chromatin regions linked to nearby differential transcripts in 1
dpa versus 0 dpa whole-fin samples. Each point indicates a separate ATAC-seq peak. (G) Heat map of upregulated transcripts in 1 dpa versus 0 dpa regenerating
whole-fin tissue linked to nearby differentially accessible chromatin regions. Uninjured 0 dpa samples for both 4 dpa and 1 dpa datasets were used to
consider batch effects when performing differential comparison. (H) Venn diagram comparison of differentially expressed genes in 0 versus 1 dpa compared with
0 versus 4 dpa RNA-seq data. Many of the transcripts that are differential at 4 dpa from 0 dpa are also differential at 1 dpa. (I,J) Browser tracks of TREEs near
fgf20a, fgf20aE48 (I) and mdka, mdkaE20 (J). Both genes increase transcript levels during regeneration. Tracks represent normalized combined replicates.
fgf20a encodes a ligand expressed by epithelial cells in the distal tip of the regenerating fins, and fgf20a null mutations block the regeneration of amputated fins (Shibata et al., 2016; Whitehead et al., 2005). We discovered a putative TREE 48 kb downstream of the transcriptional start site of fgf20a that displays increased chromatin accessibility at both 1 and 4 dpa. We subcloned this 953 bp fragment, referred to as fgf20aE48 (Figs 1I and 2A), and used it to generate eight independent lines of fgf20aE48:EGFP zebrafish. We compared the EGFP expression domains in these enhancer reporter lines with those of the published HGn21A:EGFP line, in which a construct containing a heat shock minimal promoter and EGFP expression cassette randomly inserted 1.6 kb upstream of the fgf20a transcriptional start site (Shibata et al., 2016). Both fgf20aE48 and HGn21A direct EGFP expression in larval brain (4 of 8 fgf20aE48 lines) and in larval skeletal muscle (3 of 8 fgf20aE48 lines, data not shown) (Fig. 2B–C’), with HGn21A:EGFP fluorescence being considerably stronger in each domain. During fin regeneration, fgf20aE48-directed fluorescence was induced in a similar distal region of fin epidermis at 4 dpa as HGn21A:EGFP (Fig. 2G–G’). In general, the EGFP fluorescence domains were similar but broader than endogenous fgf20a mRNA detected by in situ hybridization. To assess whether this was a result of persistent EGFP fluorescence, we examined EGFP RNA localization in HGn21A:EGFP and fgf20aE48:EGFP regenerates. Wound epidermal expression of EGFP mRNA at 4 dpa in both reporter lines was more restricted and similar to endogenous fgf20a,
although much stronger in HGn21A:EGFP regenerates versus fgf20aE48:EGFP regenerates (Fig. S4AE,AF). Despite the similar domains at 4 dpa, the dynamics of fgf20aE48 and HGn21A: EGFP expression were distinct. In uninjured adults, HGn21A:EGFP expression was present at the very distal edge of the caudal fin, whereas fgf20aE48:EGFP expression was undetectable (Fig. 2H-K). fgf20aE48-directed EGFP was not detectable at 1 dpa, was present in the 4 dpa regeneration epidermis and was undetectable again at 7 dpa, whereas HGn21A:EGFP was strong throughout these stages (Fig. 2L-Q). CRISPR deletion of the fgf20aE48 region did not notably affect fgf20a transcript levels in uninjured fins (Fig. 2R) or 2 days post fertilization (dpf) embryos (Fig. S4AJ). We compared expression levels normalized to a standard gene (actb2), as well as relative increases in expression from the 0 dpa levels of the corresponding genotype. Following amputation, fgf20aE48ΔΔ fins trended toward increased fgf20a expression at 1 dpa (Fig. 2R,S), although with variability in expression. At 4 dpa, fgf20aE48ΔΔ fins had similar fgf20a levels as wild-type regenerates (Fig. 2R), with the extent of increase versus 0 dpa samples being lower than wild types (Fig. 2S). The expression domain of fgf20a was not altered in 4 dpa fgf20aE48ΔΔ fins (Fig. 2D,E). The lengths of fgf20aE48ΔΔ fin regenerates were slightly decreased, by ~5-7%, at 4, 7 and 14 dpa (Fig. 2T), while fin patterns were grossly normal (Fig. S4F-J). Together, our results indicate that the fgf20aE48 regulatory sequence is likely to direct a component of fgf20a expression, particularly at 4 dpa, and there are other essential regulatory sequences that maintain sufficient levels of Fgf20a ligand to permit regeneration.

mdka encodes a heparin-binding growth factor that activates MAPK or AKT pathways and can promote repair, proliferation, the head, eye, notochord and caudal fin fold (Fig. 3B,B′). We subcloned a 1.3 kb ∼-coding sequence (Gramage et al., 2015; Lien et al., 2006; Nagashima et al., 2019). In zebrafish, Mdka is implicated in heart and retina regeneration (Gramage et al., 2015; Lien et al., 2006; Nagashima et al., 2019). We found that mdka transcript levels increase sharply during fin regeneration. A putative TREE 20 kb upstream of the mdka start site displayed increased chromatin accessibility at 1 and 4 dpa (Fig. 1J). We subcloned a 1.3 kb mdkaE20 fragment corresponding to this region into our reporter construct (Fig. 3A). This sequence directed EGFP expression throughout transgenic zebrafish larvae, mainly in the head, eye, notochord and caudal fin fold (Fig. 3B,B′), but with minimal expression in uninjured adult fins (Fig. 3C). Eight out of 10 independent stable transgenic reporter lines generated with this fragment directed mdkaE20:EGFP expression in epidermal and/or mesenchymal cell expression during fin regeneration, similar to the ISH pattern of mdka (Fig. 3D,F,F′), with EGFP mRNA detection by ISH in mdkaE20:EGFP regenerates showing similar results (Fig. S4AG). mdkaE20-directed EGFP was present at 1 dpa both proximal and distal to the injury site, and this expression persisted until ~7 dpa, fading by 21 dpa (Fig. 3G-J). Deletion of the entire mdka-coding sequence (mdkaΔΔ) did not impact animal viability, and mutant fin regenerates measured ~4-8% shorter than wild-type clutchmates at 4, 7 and 14 dpa, respectively (Fig. 3K), with overall patterning being normal (Fig. S4L-P). Homozygous deletions of the mdkaE20 TREE via CRISPR resulted in similar mdka expression levels as in wild-type fins at 0, 1 and 4 dpa (Fig. 3L), and in similar increases at 1 dpa and 4 dpa over 0 dpa as wild types (Fig. 3M), although embryonic expression of mdka was reduced by 38% in 2 dpf mdkaE20ΔΔ embryos (Fig. S4AJ). The expression pattern of mdka in mdkaE20ΔΔ regenerating adult fins was similar to wild types (Fig. 3D,E), and mdkaE20ΔΔ fins regenerated normally without any detectable patterning defects (Fig. 3N; Fig. S4R-V). Thus, our data indicate that mdkaE20 is used to direct embryonic mdka expression, directs expression mainly in epidermis and/or distal blastema during regeneration, and that it is not required to activate mdka expression after fin amputation.

**Genome-wide profiling of uninjured fin and blastemal fibroblasts**

Profiles of purified cell populations can provide lineage-specific information that is masked by bulk tissue analysis. Fibroblasts are the predominant cell type in zebrafish fins and in the regeneration blastema, and recent findings indicate that blastemal fibroblasts possess positional information for regeneration (Tornini et al., 2016). To genetically label uninjured and regenerating fibroblasts for purification, we took advantage of the fact that immediate upstream regulatory sequences of tph1b, a gene encoding the rate-limiting enzyme in serotonin synthesis, direct expression in blastemal fibroblast progenitors upon injury. Thus, if a high tamoxifen dose is administered to tph1b:CreER; ubi:switch zebrasfish 2 days after a proximal fin amputation, most blastemal fibroblasts and their progeny permanently express mCherry (Fig. 4A). If this fin is allowed to fully regenerate (i.e. for 30 days or more), a high percentage of fibroblasts stably expressing mCherry occupy the fin (Tornini et al., 2016).

We performed ATAC-seq and RNA-seq on stably labeled fin fibroblasts sorted from 0 dpa and regenerating 4 dpa tph1b:CreER; ubi:switch tissue (Fig. 4A). This analysis revealed 4629 differentially regulated transcripts and 33,732 differentially accessible chromatin regions (Fig. S3A; Tables S14, S15). We were able to merge the ATAC-seq and RNA-seq data by pairing 10,197 differential chromatin regions with nearby differential RNA transcripts (Fig. 4B,C; Table S16). Comparative analysis of the fibroblast differential transcriptome to the corresponding whole-fin 4 dpa differential transcriptome showed overlap, with 51.9% of the differential fibroblast transcripts present in differential whole-fin 4 dpa data (Fig. 4D). Regions with increased accessibility in fibroblasts were mainly in intergenic regions (65.8%), introns (17.9%) or exons (10.0%). The remaining 6.4% of differential chromatin regions in 4 dpa fibroblasts were within 3 kb upstream or downstream of a gene, or in UTR regions (Fig. S3B). Similar to in the 4 dpa dataset, ChromVar analysis revealed that the most variable transcription factor binding sites in the fibroblast data implicate components of the AP-1 complex (Fig. S3C) (Karlin et al., 1997; Schep et al., 2017), and footprinting analysis again identified changes in accessibility near the three selected transcription factor consensus sequences FO5:JUN, EPAS1 and ZNF263 (Fig. S3D-F). As in whole-fin tissue, IPA indicated the most highly activated upstream regulator as ESRB2, and the most downregulated upstream regulator as TP53 (Fig. S3G; Table S17). IPA of canonical pathways upregulated in fibroblasts revealed a notable increase in oxidative phosphorylation, with one of the most downregulated pathways again being melatonin degradation (Fig. S3H; Table S18).

Comparative analysis of merged differential ATAC peaks paired with nearby differential transcripts revealed 651 transcripts with nearby differential peaks shared between the fibroblast and 4 dpa dataset, with 4460 unique to the fibroblast dataset and 2419 unique to the 4 dpa dataset (Fig. S3I). GO Term analysis of the shared transcripts revealed overlapping involvement in wound healing and regeneration processes, as well as axonal development (Fig. S3J). The fibroblast dataset showed greater enrichment with dynamic accessibility near differential transcripts involved in sensing external stimuli, cell growth and neuronal guidance, whereas the 4 dpa whole-fin dataset showed greater enrichment with dynamic accessibility near differential transcripts involved with vasculature and epithelium (Fig. S3K,L).
We selected a putative TREE showing increased accessibility in 4 dpa fibroblast samples, more prominent than in 1 or 4 dpa whole-fin samples, and located 24 kb downstream of the cx43 locus exhibit shortened ray segments and an overall shortened fin (Fig. S4AH). We saw strong cassette EGFP fluorescence in regenerating fin rays at 4 dpa, which lingered until 7 dpa and faded once regeneration was completed (Fig. 5H-J). Deletion of cx43E24 by CRISPR revealed similar cx43 transcript levels at 0 and 1 dpa in cx43E24ΔΔ and wild-type regenerates, and similar increases in expression from 0 to 1 dpa in both genotypes (Fig. 5K,L). By contrast, cx43 levels were 26.1% lower in cx43ΔΔ or wild type (M). Each biological replicate pool of three fins each was performed in technical triplicate, and paired two-tailed t-tests were performed to indicate significance (n=6). (N) Fin regeneration in mdkaE20ΔΔ fish is similar to wild type (n=14). Paired two-tailed t-tests were performed to indicate significance. Data are means±s.d.
was also reduced by 33% in cx43E24Δ/Δ embryos (Fig. S4AJ). No difference in the expression pattern of cx43 mRNA was observed in cx43E24Δ/Δ 4 dpa fins compared with wild type (Fig. 5E,F). Overall cx43E24Δ/Δ fin regeneration lengths at 4 dpa and 7 dpa were normal compared with wild types (Fig. 5M; Fig. S4X-AB). However, we found that ray segments distal to the amputation plane in regenerating cx43E24Δ/Δ 7 dpa fin tissue were 6.2% shorter than wild types, and those proximal to the amputation plane were 5.2% shorter (Fig. 5N). These changes are modest, but given the established phenotype of sof mutants (Iovine et al., 2005), and as we found fgf20aE48Δ/+ and mdkaE20A/+ ray segment lengths were no different from wild-type clutches mates (Fig. S4AI), they are possibly of biological significance. Thus, our results indicate that cx43E24 is sufficient to direct blastemal fibroblast expression following fin amputation and contributes to induction and/or maintenance of cx43 expression at 4 dpa, although additional elements regulate cx43 expression.

Deletion of the LEN element abolishes injury-induced expression of leptin b
LEN was originally identified as a sequence region ∼6 kb upstream of the lepb transcriptional start site that becomes enriched with the H3K27Ac mark during zebrafish heart regeneration. During fin regeneration, lepb is sharply induced from essentially undetectable levels, and LEN was shown to be inactive in uninjured fins but able to direct EGFP expression in regenerating fibroblasts when linked to a lepb 2 kb minimal promoter (Kang et al., 2016) (Fig. 6A, B). To date, the chromatin signature of LEN during fin regeneration has not been assessed. In our datasets, lepb levels were elevated in 0 dpa fibroblasts, similar to the cx43 expression data and possibly again due to the double amputation protocol. RNA levels increased from 0 to 4 dpa in fibroblasts, and sharply from 0 to 1 or 4 dpa in whole-fin samples. LEN was accessible in both whole-fin samples and in purified fibroblast samples, although the peak was not dynamic during regeneration and would not have been considered differential during regeneration by our bioinformatic analyses (Fig. 6C). To explain this, it is possible that LEN is not regulated by context-specific chromatin remodeling and is held accessible in a primed or poised state prior to injury-induced activation (Heinz et al., 2015). Alternatively, ATAC-seq is not informative for this locus versus other marks of active regulatory sequences.

To test the requirement for LEN in control of endogenous lepb expression during regeneration, we designed gRNAs to generate a ∼3.4 kb deletion encompassing LEN. Homozygous LEN deletion mutants survived and regenerated fins normally; this was expected, as lepb-coding sequence mutants we previously generated also survive and regenerate normally (Kang et al., 2016) (Fig. 6D). Importantly, qPCR and transcriptome sequencing from regenerating wild-type and LENΔ/Δ fin regenerates revealed no detectable elevation of lepb at 2 dpa in the absence of LEN (Fig. 6E,F,H). RNA-seq of LENΔ/Δ fins also indicated normal upregulation of the lepb-linked gene impdhlb during regeneration, suggesting specificity of LEN to the lepb target gene (Fig. 6G,H). Therefore, LEN is genetically required for lepb regulation during regeneration, and there is no apparent compensation by alternative enhancer elements at this locus.
Distinct chromatin accessibility patterns in regenerating osteoblasts and fibroblasts

While this manuscript was being prepared, Johnson and colleagues published a report assessing DNA methylation and chromatin accessibility during zebrafish fin regeneration in purified osteoblast (sp7+) and non-osteoblast (sp7−) cells at 0 and 4 dpa (Lee et al., 2020). We compared our purified cell datasets and observed that while the number of differentially expressed genes during regeneration was similar between fibroblasts and osteoblasts, only 41.9% of transcripts differentially expressed during regeneration in fibroblasts were also indicated as differentially expressed in osteoblasts (Fig. 7A). We used the methylation datasets from Lee et al. to determine the methylation states of the enhancers identified in our study, finding that methylation states of fgf20aE48, mdkaE20, cx43E24 and LEN were not dynamic between uninjured and regenerating fins, as a key conclusion from their work predicts. Interestingly, each of these enhancers contained a cluster of unmethylated regions in both sp7+ and sp7− datasets (Fig. S5G-J). We examined chromatin accessibility across datasets for sequences validated in transgenic lines from Lee et al. and from the current study (Fig. 7B-H; Fig. S5A-F). Interestingly, the osteoblast TREEs cdc42ep3+2k and chst3b+5k exhibited significantly decreased accessibility in fibroblasts during regeneration, opposing the methylation datasets from Lee et al. and demonstrating distinct chromatin accessibility patterns between fibroblasts and osteoblasts during zebrafish fin regeneration.

Fig. 5. A TREE linked to cx43 directs blastema gene expression. (A) RNA-seq and ATAC-seq browser tracks indicating the cx43E24 TREE near the upregulated cx43 transcript in regenerating fibroblast and whole-fin tissue. Tracks represent normalized combined replicates. (B) The cx43E24-fos:EGFP enhancer reporter construct. The 945 bp candidate TREE (blue arrow) was placed upstream of a 96 bp c-fos minimal promoter (yellow arrow) to direct fluorescent EGFP expression (green arrow). (C) Larval cx43E24:EGFP animals at 5 days post fertilization (dpf) displaying EGFP expression in trunk muscle (mus) and early pectoral fins. High magnification view of a larval pectoral fin (arrowhead) (C'). Scale bars: 500 μm. (D) The three dorsal-most caudal fin rays in some strongly expressing cx43E24:EGFP animals express EGFP in uninjured tissue, although this is not present in all lines. Scale bar: 300 μm. (E-G') Longitudinal sections indicate cx43 expression (violet) in fibroblasts (white asterisks) and joint osteoblasts in wild-type fish (E) and cx43E24Δ/Δ fish (F) by in situ hybridization. Expression in forming joints is marked by black arrowheads. cx43E24:EGFP reporter fish exhibit a similar EGFP immunofluorescence pattern to endogenous cx43, although EGFP expression is detected in most blastemal osteoblasts (white arrowheads) in addition to blastemal fibroblasts (white asterisk) (G,G'). Scale bars: 100 μm. (H) cx43E24:EGFP fish display mesenchymal EGFP expression by 4 dpa, fading by 21 dpa. Arrowheads indicate amputation plane. Scale bars: 300 μm. (K,L) qPCR analysis of cx43 transcript levels in cx43E24Δ/Δ and wild-type fins. Expression is induced at 4 dpa but to a lesser extent than in wild-type fins. Each biological replicate pool of three fins each was performed in technical triplicate, and paired two-tailed t-tests were performed to indicate significance (n=8). (M) Lengths of regenerating fins are similar in cx43E24Δ/Δ and wild-type fish (n=14). Paired two-tailed t-tests were performed to indicate significance. (N) Fin ray segment length at 7 dpa was decreased by 6.2% distal to the amputation plane and 5.2% shorter proximal to the amputation plane in regenerating cx43E24Δ/Δ fins compared with wild-type fins (n=14). Paired two-tailed t-tests were performed to indicate significance. Data are mean±s.d.
Fig. 6. LEN is essential for lepb expression during fin regeneration. (A,B) LENP2:EGFP fish display EGFP expression in the blastema by 2 dpa, as reported by Kang et al. (2016). Arrow indicates amputation plane. Scale bars: 300 μm. (C) Browser tracks indicating the LEN TREE (arrow) near the upregulated lepb transcript in regenerating fibroblast and whole fin tissue. While LEN is accessible in whole fin and fibroblast tissue, its accessibility is not considered dynamic during regeneration. Tracks represent normalized combined replicates. (D) Lengths of regenerating fins are similar in transcript in regenerating fibroblast and whole fin tissue. While regeneration. (H) RNA-seq of by Kang et al. (2016). Arrow indicates amputation plane. Scale bars: 300 μm. (Fig. 5F). Interestingly, LEN is essential for lepb expression during fin regeneration. (A,B) LENP2:EGFP fish display EGFP expression in the blastema by 2 dpa, as reported by Kang et al. (2016). Arrow indicates amputation plane. Scale bars: 300 μm. (C) Browser tracks indicating the LEN TREE (arrow) near the upregulated lepb transcript in regenerating fibroblast and whole fin tissue. While LEN is accessible in whole fin and fibroblast tissue, its accessibility is not considered dynamic during regeneration. Tracks represent normalized combined replicates. (D) Lengths of regenerating fins are similar in LENΔ/+ and wild-type fish (n=12). Paired two-tailed t-tests were performed to indicate significance. Data are mean±s.d. (E,F) qPCR analysis of lepb transcript levels in LENΔ/+ and WT fins. lepb is undetectable in 2 dpa regenerating LENΔ/+ fish. Each biological replicate pool of three fins each was performed in technical duplicate, and paired two-tailed t-tests were performed to indicate significance. (n=5). Data are mean±s.d. (G) Genomic region immediately surrounding LEN. lepb and impdh1b are induced during fin regeneration. (H) RNA-seq of LENΔ/+ and wild-type 0 and 2 dpa regenerates. Tracks indicate an absence of lepb expression at 2 dpa in LENΔ/+ fish, whereas impdh1b transcript levels are increased similarly in both wild type and LENΔ/+ upon fin amputation. Tracks represent individual replicates.

Conclusions

We provide a resource of the chromatin accessibility landscape of regenerating zebrafish caudal fins during blastema formation and active regenerative outgrowth, revealing thousands of putative TREEs from whole-fin tissue. Profiling purified blastemal fibroblasts further allowed us to filter out noise from the heterogeneity of the whole-fin samples and detect fibroblast-relevant signals. The resources presented in this study, combined with other complementary datasets (Lee et al., 2020), identify many enhancers and putative enhancers that direct gene expression during appendage regeneration. This resource also describes enriched motifs and network analysis for predicted transcriptional regulators, as well as regeneration-related regulatory sequences to examine for conservation between species or for ectopically expressing gene cassettes during fin regeneration.

Differentially accessible sequences we attempted to validate were consistently able to direct reporter gene expression that approximates expression of a nearby endogenous gene. Only in the case of LEN did we see observe that enhancer deletion abrogated all detectable endogenous gene expression, whereas other TREE deletions had partial or no detectable effects on expression during regeneration. Studies of enhancers in other contexts have indicated that individual elements can be redundant, or their loss can lead to compensation or chromosome structural changes that impact expression (Andersson and Sandelin, 2020; Cannavò et al., 2016; El-Sherif and Levine, 2016; Heinz et al., 2015; Scholes et al., 2019). Thus, we suspect that additional regulatory elements, either detectable in ATAC-seq datasets or cloaked, contribute to expression levels of fgf20a, mdka, and cx43. Although we have shown that transposase accessibility assays, and, most likely, assays for other chromatin features like H3K27Ac, might not directly identify TREEs that are essential for regeneration-responsive expression, assessment of interactions between genomic regions near induced genes, using chromosome conformation capture (Grob and Cavalli, 2018) or tiled deletion approaches (Diao et al., 2017),
may help identify additional regulatory sequences related to the genes studied here or other genes involved in regeneration.

MATERIALS AND METHODS

**Zebrafish maintenance and procedures**

Wild-type or transgenic zebrafish of the outbred Ekkwill (EK) or EK/AB strain ranging from 4 to 12 months were used for all experiments. Transgenic and mutant zebrafish lines are presented in Table S19. Water temperature was maintained at 27.5°C, and fish were kept on a 14/10 light/dark cycle. Fish were anesthetized in 0.75% v/v 2-phenoxyethanol (Sigma-Aldrich) in fish water. Fins were amputated at 50% of their original length using razor blades (VWR). All work with zebrafish was performed in accordance with Duke University guidelines.

Enhancer reporter zebrafish lines (\(\text{fgf20aE48:EGFP}\), \(\text{mdkaE20:EGFP}\), and \(\text{cx43E24:EGFP}\)) were generated using I-SceI transgenesis. Enhancer fragments (954 bp \(\text{fgf20aE48}\), 1310 bp \(\text{mdkaE20}\), and 946 bp \(\text{cx43E24}\)) were subcloned via Gateway Cloning using Gateway LR Clonase II (Thermo Fisher Scientific) upstream of a mouse \(\text{c-fos}\) minimal promoter, EGFP and poly-A tail, flanked by I-SceI sites (Goldman et al., 2017). Primers are provided in Table S20. Reporter construct (500 µl of 25 ng/µl), 333 U/ml I-SceI (NEB), 1x I-SceI buffer and 0.05% Phenol Red (Sigma-Aldrich) solution were injected into one-cell zebrafish embryos. F0 founders were discovered by screening for fluorescence in outcrossed F1 embryos.
Deletion lines (LENΔmdkaE20, fgf20aE48Δ/Δmdka, mdkaE20Δ/ΔmdkaE20 and cx43E24Δ/ΔmdkaE20) were created by CRISPR/Cas9 genome editing (3440 bp LEN deletion, 2517 bp fgf20aE48 deletion, 1263 bp mdkaE20 deletion and 1658 bp cx43E24 deletion). Double-stranded DNA (dsDNA) templates were generated from a Universal gRNA 3’ oloigor combined with the CRISPR Primers (designed by CRISPRscan), listed in Table S20, using Phusion DNA Polymerase (New England Biolabs), then purified using a DNA Clean and Concentrator-5 kit (Zymo). The purified dsDNA templates were used for in vitro transcription to generate gRNA using a T7 or SP6 RNA Polymerase kit (New England Biolabs), then purified using an RNA Clean and Concentrator-5 kit (Zymo). A gRNA pair (500 pL of 30 ng/μl) of Phenol Red (Sigma-Aldrich) with 0.05% Phenol Red (Sigma-Aldrich) solution was injected into one-cell zebrafish embryos. Two to 4 dpf embryos were genotyped by PCR and agarose gel electrophoresis using primers in Table S20 to determine gRNA efficiency. To screen F0 founders, genomic DNA was extracted from outcrossed F1 embryos and genotyped by PCR and agarose gel electrophoresis with KAPA2G Genotyping Mix (Kapa Biosystems) using the primers in Table S20. Deletions were confirmed via Sanger sequencing (Eton Bioscience).

RNA isolation and qPCR
Caudal fin tissue from three fish or from 15 2 dpf embryos, was pooled for each biological replicate and collected in Tri-Reagent (Sigma). RNA was isolated by ethanol precipitation followed by purification with an RNA Clean & Concentrator Kit-5 (Zymo). cDNA was synthesized from 1 μg of total RNA with a First Strand cDNA Synthesis Kit (Roche). Quantitative PCR was performed on a Roche LightCycler 480 using LightCycler 480 SYBR Green 1 Master Mix (Roche) with primers listed in Table S20. fgf20aE48Δ/Δ, mdkaE20Δ/Δ, and cx43E24Δ/Δ qPCRs were completed in technical triplicates for each time point, and LENΔ/ΔA qPCRs in technical duplicates. All experiments normalized transcript expression levels to actb2 as a control. PRISM software was used to perform statistical analysis with paired t-tests and to generate graphs (GraphPad PRISM 8.4.0).

Library preparation and sequencing
For whole-fin libraries, biological triplicate pools of caudal fin clips were collected at 0 dpf (freshly amputated), 1 dpf or 4 dpf time points via razor blade (VWR) from 15 fish per pool. Cells were dissociated using 1.18 U/ml Liberase DH (Roche) in Hank’s Balanced Salt Solution (Gibco) stirring at 37°C for 2 h, collected in 15 min increments, then strained through a 70 μm nylon cell strainer (Corning). Samples were quantitated for 100 K live cells using the Countess Automated Cell Counter (Thermo Fisher Scientific). ATAC library preparation was performed as described previously (Buenrostro et al., 2013).

For whole-fin 0/1 dpf, RNA-seq, triplicate Ekkwll strain caudal fin biological replicate pools were collected at 0 dpf, 1 dpf or 4 dpf time points via razor blade (VWR) from 15 fish per pool aged between 3 and 12 months. LENΔ/ΔA and corresponding wild-type tissue were collected similarly but at 0 dpf and 2 dpf. Fin tissue was collected in Tri-Reagent (Sigma) and isolated by ethanol precipitation. RNA was further purified with the RNA Clean & Concentrator Kit-5 (Zymo). Libraries were prepared using a Stranded mRNA-seq Kit (KAPA).

For fibroblast libraries, triplicate pth1b::creER<sup>T2</sup> W6301, ubi:switch<sup>C1707</sup> caudal fins were clamped and treated at 2 dpf for 12 h in 4 μM tamoxifen (Sigma-Aldrich) (Tornini et al., 2016), and allowed to regenerate fully for 2 months. Seventy-five fish fin clips were combined for each biological replicate pool, collected at 0 dpf or 4 dpf time points. Cells were dissociated using 1.18 U/ml Liberase DH (Roche) in Hank’s Balanced Salt Solution (Gibco) stirring at 37°C for 2 h, collected in 15 min increments, strained through a 70 μm nylon cell strainer (Corning), then sorted for 100 K mCherry<sup>+</sup> fluorescent cells on a B-C Astrios cell sorter by the Duke Cancer Institute Flow Cytometry Shared Resource. ATAC-seq library preparation was performed as detailed by Buenrostro et al. (2013). Libraries were prepared using a Stranded mRNA-seq Kit (KAPA).

Sequencing for the Whole Fin 0/4, Fibroblast and LENΔ/ΔA libraries was performed on a HiSeq 4000, 50 bp SE, by the Duke Center for Genomic and Computational Biology. Sequencing for the Whole Fin 0/1 libraries was performed on a DNBSEQ-MGI2000, 100 bp PE, by BGI Americas.

Histology and imaging
Histology was performed on cryosections of fin regenerates fixed in 4% paraformaldehyde as described previously by Tornini et al. (2016). RNA probes were amplified from 4 dpf fin cDNA using primers listed in Table S20, transcribed with T7 RNA polymerase (New England Biolabs), and labelled using a DIG RNA Labeling Mix (Roche). In situ hybridization was performed using an InSituPro VSI robot (Intavis). Immunohistochemistry to enhance visualization of EGFP in enhancer reporter fish was performed with anti-GFP antibodies (rabbit, Invitrogen, A-11122; 1:200) followed by anti-rabbit antibodies (Alexa Fluor 488 conjugated; goat, Invitrogen, A-11034; 1:500). Imaging was performed on a Leica DM6000B compound microscope.

Anesthetized zebrafish were imaged using a Zeiss AxioZoom microscope at the time points indicated. All raw images were processed using either ZEN (Zeiss) or FIJI (Schindelin et al., 2012) software. Lengths of fin regenerates were measured using ZEN Blue software (Zeiss), averaging the lengths of the second and third dorsal- and ventral-most rays for each animal. Fin segment lengths were measured by taking the distance between the third and fourth joint on the second and third dorsal-most rays in each fin. PRISM software was used to perform paired two-tailed t-tests on fin amputations and generate graphs (GraphPad PRISM 8.4.0). Adobe Illustrator was used to format graphs (Adobe Illustrator 23.0.6).

Bioinformatic analysis
RNA-seq reads were trimmed using Trim Galore (Krueger, 2015) (0.4.1, with -q 15—paired) and then mapped with TopHat (Trapnell et al., 2009) (v 2.1.1, with parameters --b2 very-sensitive --no-coverage-search and supplying the UCSC danRer10 refSeq annotation). Gene-level read counts were obtained using the htsseq-count (Anders et al., 2015) (v 1.6.1) by the reads with MAPQ greater than 30. Bioconductor package DESeq2 (Love et al., 2014) (v 1.26.0) was employed for differential expression (DE) analysis. A fold change greater than two and a false discovery rate of less than 0.05 was used to filter the significant DE genes. The zebrafish gene IDs were mapped to mouse homologous gene IDs to perform pathway and upstream regulator analysis using QIAGEN Ingenuity Pathway Analysis software (Krämer et al., 2014).

ATAC-Seq reads were trimmed using Trim Galore (0.4.1, with -q 15) then mapped with bowtie2 (Langmead and Salzberg, 2012) (2.2.5, with parameters –very-sensitive) to zebrafish genome (UCSC danRer10 refSeq annotation). The mapped reads were filtered by a MAPQ greater than 30 using samtools (Li et al., 2009) (v 1.5) and duplicated reads were removed by picard (v 1.91). The peaks were called by MACS2 (Zhang et al., 2008) (v 2.1.0, with –nomodel –shift 37 –extsize 73 -q 0.05 -q 1.5e9). The bioconductor package DiffBind (Ross-Innes et al., 2012) (v 2.14.0) was employed for differential open region analysis. A fold change greater than two and a P-value of less than 0.05 was used to filter the significant differential peaks. The enriched motifs were analyzed using chromVAR (Schep et al., 2017) (v 1.8.0). The motifs were a collection of jasper2018, jaspa2013 and csibsp_1.02 from package motifDB (v 1.28.0) and were merged by a distance smaller than 1e-9 calculated by MotifV::motifDistances function (v 1.42.0). The merged motifs were exported by motifStack (Ou et al., 2018b) (v 1.30.0). The figure footprints were plotted by ATACseqQC package (version 1.12.3) (Ou et al., 2018a).

Bioconductor package ChIPseekAnno (Zhu et al., 2010) (v 3.21.4), trackViewer (Ou and Zhu, 2019) (v 1.23.2) and complexHeatmap (Gu et al., 2016) (v 2.2.0) were used to annotate the peaks, plot tracks and generate heatmaps. The differential peaks were annotated by nearest gene start site to the center of peaks. ATAC-Seq peaks were paired to RNA-Seq differential expression data by annotated nearest gene symbols. The methylation/ unmethylation data were downloaded from GSE126702 (Lee et al., 2020) and unlogged using the trackViewer package (version 1.24.1) (Ou and Zhu, 2019).

Acknowledgements
We thank Jim Burris, Karina Olivierri, Larry Frauen, Colin Dolan, Eileen Gu, Daniel Stutts, Tim Curtis, Nicolette Mariano and Shane Miller of the Duke Zebrafish Core for
zebrafish care; Aleixas Saifi for assistance with ATAC-seq; Atsushi Kawakami for the Hpn21A line; Valerie Tornini for guidance on transgenic lines; and Chen-Hui Chen for comments on the manuscript. This work is dedicated to our friend, mentor and colleague Stephen L. Johnson.

Competing interests
The authors declare no competing or financial interests.

Author contributions

Funding
This work was supported by a Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung Postdoctoral Fellowship to V.C., by University of Wisconsin-Madison start-up funds (AAC8355, AAC8979, AAC8429 and AAG9756) to J.K. and by grants from the National Institutes of Health (ROI R01 AS076342 and R01 GM074057) to K.D.P. Deposited in PMC for release after 12 months.

Data availability
The ATAC-seq and RNA-seq data have been deposited in GEO under accession number GSE149660.

Supplementary information
Supplementary information available online at https://dev.biologists.org/lookup/doi/10.1242/dev.191262.supplemental

Peer review history
The peer review history is available online at https://dev.biologists.org/lookup/doi/10.1242/dev.191262Reviewer-comments.pdf

References

Development (2020) 147, dev191262. doi:10.1242/dev.191262


