

RESEARCH ARTICLE SUMMARY

REGENERATION

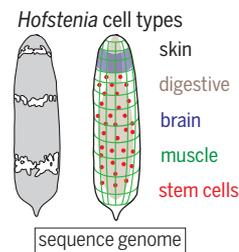
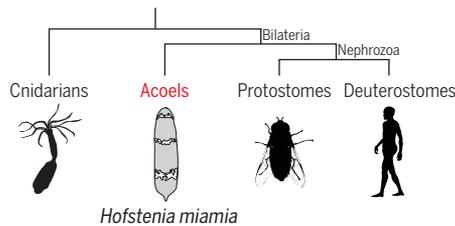
Acoel genome reveals the regulatory landscape of whole-body regeneration

Andrew R. Gehrke, Emily Neverett, Yi-Jyun Luo, Alexander Brandt, Lorenzo Ricci, Ryan E. Hulet, Annika Gompers, J. Graham Ruby, Daniel S. Rokhsar, Peter W. Reddien, Mansi Srivastava*

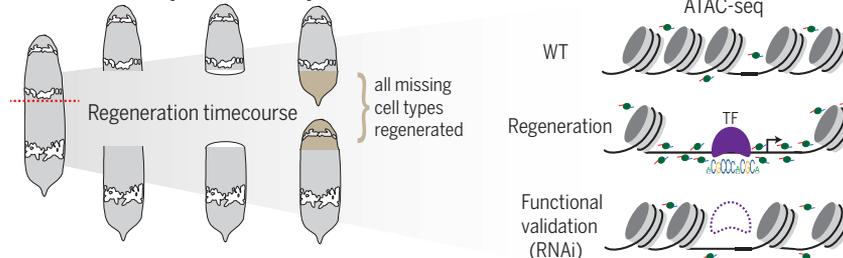
INTRODUCTION: Although all animals can heal wounds, some are capable of reconstructing their entire bodies from small fragments of the original organism. Whole-body regeneration requires the interplay of wound signaling, stem cell dynamics, and positional identity, all of which have been investigated at the protein-coding level of the genome. Little is known about how the noncoding portion of the genome responds to wounding to control gene

expression and to launch the process of whole-body regeneration. Understanding how these control points (regulatory regions) are activated and then operate during regeneration would uncover how genes connect into networks, ultimately restructuring entire body axes. Networks of transcriptional regulatory genes can reveal important mechanisms for how animals can grow new skin, muscles, or even entire brains.

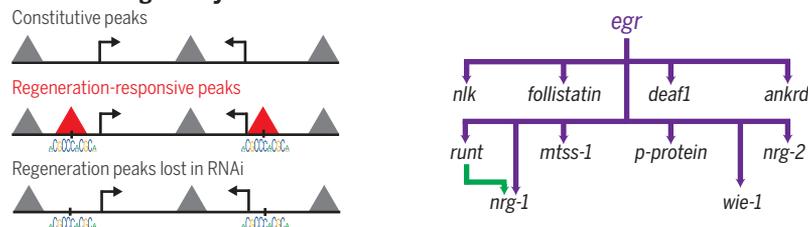
Hofstena genome sequencing and assembly



Regeneration ATAC-seq and RNA-seq



Regeneration Gene Regulatory Network



The regulatory landscape of whole-body regeneration. *Hofstena* represents the sister lineage of other bilaterians and regenerates extensively. We sequenced the genome of *H. miamia* and used ATAC-seq to identify thousands of regeneration-responsive regions of chromatin. Combining motif analysis, ATAC-seq, and RNAi, we identified Egr as a master regulator of regeneration in *Hofstena* and inferred an Egr-controlled GRN for regeneration.

RATIONALE: To identify regulatory regions involved in whole-body regeneration, we sequenced the genome of the highly regenerative acoel *Hofstena miamia*, commonly known as the three-banded panther worm. Equipped with this genome, we reasoned that applying the assay for transposase-accessible chromatin using sequencing (ATAC-seq) would

ON OUR WEBSITE

Read the full article at <http://dx.doi.org/10.1126/science.aau6173>

identify regulatory regions that change in response to amputation and during whole-body regeneration. Further, by analyzing the sequence motifs contained within these regulatory regions, we sought to predict which transcription factors (TFs) control regeneration gene networks.

RESULTS: The *Hofstena* genome assembly totals 950 megabases of sequence, with sufficient contiguity for functional genomics. ATAC-seq data revealed thousands of chromatin regions that respond dynamically during regeneration. A genome-wide scan for TF binding motifs in these regions identified the EGR (early growth response) motif as the most dynamic. By combining RNA interference (RNAi) and RNA-seq, we predicted a set of Egr target genes in *Hofstena*. We found that most of these target genes contained EGR binding motifs in neighboring regions of regeneration-responsive chromatin, which failed to respond under *egr*-RNAi. This functional validation allowed us to build a gene regulatory network (GRN) with Egr as a direct master regulator of downstream regeneration genes. Lastly, by quantifying the binding probabilities of TFs at individual motifs, we identified targets of TFs further downstream of Egr, extending the regeneration GRN.

CONCLUSION: Using our regulatory data, we inferred a GRN for launching whole-body regeneration in the acoel *H. miamia*, where the master regulator Egr acts as a putative pioneer factor to directly activate wound-induced genes. This network includes homologs of genes that are involved in regeneration in other species, suggesting that it can serve as a template for direct comparisons of regeneration pathways across distantly related animals. Our approach of combining genome-wide assays for chromatin accessibility with functional studies can be applied to extend the network further in time in *Hofstena* regeneration and to construct GRNs for regeneration in other systems. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mansi@oeb.harvard.edu
Cite this article as A. R. Gehrke et al., *Science* 363, eaau6173 (2019). DOI: 10.1126/science.aau6173

RESEARCH ARTICLE

REGENERATION

Acoel genome reveals the regulatory landscape of whole-body regeneration

Andrew R. Gehrke¹, Emily Neverett¹, Yi-Jyun Luo¹, Alexander Brandt², Lorenzo Ricci¹, Ryan E. Hulet¹, Annika Gompers¹, J. Graham Ruby³, Daniel S. Rokhsar⁴, Peter W. Reddien⁵, Mansi Srivastava^{1,6*}

Whole-body regeneration is accompanied by complex transcriptomic changes, yet the chromatin regulatory landscapes that mediate this dynamic response remain unexplored. To decipher the regulatory logic that orchestrates regeneration, we sequenced the genome of the acoel worm *Hofstenia miamia*, a highly regenerative member of the sister lineage of other bilaterians. Epigenomic profiling revealed thousands of regeneration-responsive chromatin regions and identified dynamically bound transcription factor motifs, with the early growth response (EGR) binding site as the most variably accessible during *Hofstenia* regeneration. Combining *egr* inhibition with chromatin profiling suggests that *Egr* functions as a pioneer factor to directly regulate early wound-induced genes. The genetic connections inferred by this approach allowed the construction of a gene regulatory network for whole-body regeneration, enabling genomics-based comparisons of regeneration across species.

The capacity to replace all missing tissues, i.e., whole-body regeneration, is present in nearly all animal phyla (1). Whereas candidate gene studies and high-throughput transcriptomics have yielded insight into genes involved in several aspects of regeneration, including wound response (2, 3), stem cell dynamics (4), and repatterning of tissue identities (5–7), how these genes are connected is largely unknown. A detailed understanding of the DNA binding site logic of regulatory transcription factors is needed to elucidate connections between genes, in particular to distinguish between direct (i.e., control via binding to cis-regulatory regions of a target gene) and indirect (i.e., control of a downstream gene via an intermediate gene) regulatory relationships. Active “tissue regeneration enhancer elements” have been documented in zebrafish heart and fin regeneration (8, 9), as well as in *Drosophila* imaginal discs (10), but we lack understanding of how the epigenome responds during whole-body regeneration.

We focused on the model system *Hofstenia miamia*, an acoel worm that possesses the ability to regenerate its entire body (11) (Fig. 1A). Besides its regenerative capacity, we chose *Hofstenia* because of its plentiful and accessible embryos

(Fig. 1A) and its phylogenetic position as the likely sister group (Xenacoelomorpha) to all other bilaterians (Nephrozoa) (11–13) (Fig. 1B).

Assembly and annotation of the *Hofstenia miamia* genome

We sequenced the genome of *Hofstenia* at an average coverage of 89.6× via Illumina sequencing of paired-end [300- and 500-base pair (bp) inserts] and mate pair (3- and 7-kb inserts) libraries derived from a single individual and a pool of five animals, respectively (table S1A) (14). The sequenced animals were descendants of a wild population of individuals, and we estimated the polymorphism level as 0.426% (or 1 single nucleotide polymorphism every 235 bases) (fig. S1A). The draft assembly, achieved by the sequential application of SOAPdenovo (15), PRICE (16), SSPACE (17), and the Chicago method (18), totals 950 Mb of sequence (11.8% gaps) (14). Half of the sequence (N50) is contained in 294 scaffolds longer than 1 Mb in this assembly (fig. S1A), and this contiguity was validated by mapping long-read Oxford Nanopore sequencing and by flagging potential misjoins using REAPR (14). To predict protein-coding genes, we generated a new transcriptome assembly to train Augustus (19) and recovered 22,632 gene models, 97% of which were supported by transcriptome data (14). The *Hofstenia* genome is 53% repetitive, of which the majority are long terminal repeats derived from retrotransposons (fig. S1B and table S1B). The *Hofstenia* genome annotation contained 90% of the expected metazoan BUSCO (20) genes (84% complete BUSCOs), yielding a representative protein-coding gene set (fig. S1A) (14). A principal components analysis on gene content in 36 metazoan genomes nested the *Hofstenia* genome among nonbilaterian

and protostome genomes, indicating that it contains a standard complement of animal genes (Fig. 1C, fig. S2A, and table S1, C and D).

Chromatin is highly dynamic during regeneration

To understand chromatin dynamics during the process of whole-body regeneration, we applied the assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) (21) to identify changes in the epigenome following injury. To capture how *Hofstenia* chromatin responds during regeneration, animals were amputated transversely, and wound sites from both resulting fragments (anterior-facing wound sites from tail fragments and posterior-facing wound sites from head fragments) were harvested for ATAC-seq at “zero” (control), 3, 6, 12, 24, and 48 hours post-amputation (hpa) (Fig. 1D, fig. S2, B to E, and table S2). This time course encompasses many aspects of regeneration, including the wound response (i.e., transcriptional changes initiated immediately upon amputation) at early time points and the formation and patterning of new tissue at later time points (11, 22).

Regeneration induced dynamic patterns of change in chromatin over time, including regions of chromatin (or “peaks”) that “opened” or “closed.” We identified nearly 18,000 changes to chromatin regions (adjusted $p < 0.05$, Wald test) (14) during initiation of head regeneration in tail fragments (see fig. S2, E and F, for head fragment data analysis) (Fig. 1E). This regeneration-responsive chromatin was most frequently located in introns and intergenic space (Fig. 1F and table S3). Notably, the majority of peaks appeared within the first 6 hpa (Fig. 1G and fig. S2F). As regeneration progressed, the number of newly emerging dynamic peaks decreased, until 24 to 48 hpa, when chromatin regions tended to close (Fig. 1G, fig. S2F, and table S3).

We sought to identify if there were specific transcription factor (TF) binding motifs associated with regions of change during the first 48 hours of regeneration. We used chromVAR (23) to identify matches to known TF binding sites in the *Hofstenia* genome and to calculate a genome-wide “chromatin variability” score, a measure of how dynamic chromatin regions containing a particular motif were during regeneration. We reasoned that TFs with binding sites that become more or less accessible during regeneration, i.e., whose binding sites lie in variable chromatin, are candidate regulators of regeneration. Of the 386 motifs that we examined with chromVAR, some were more variably accessible in head fragments (e.g., TCF) or in tail fragments (e.g., NFY), likely corresponding to TF families that play important roles in tail and head formation, respectively (table S4, A and B). Many motifs showed highly variable accessibility in both head and tail fragment data (fig. S3A and table S4C). Notably, the binding site for an early growth response (*Egr*) protein emerged as the most significantly variable (adjusted $p < 10^{-10}$, Brown-Forsythe test), i.e., associated with the most

¹Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

²Department of Chemistry, University of California, Berkeley, CA 94703, USA. ³Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA. ⁴Department of Molecular and Cell Biology, University of California, Berkeley, CA 94703, USA. ⁵Whitehead Institute for Biomedical Research, Howard Hughes Medical Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02138, USA. ⁶Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA.

*Corresponding author. Email: mansi@oeb.harvard.edu

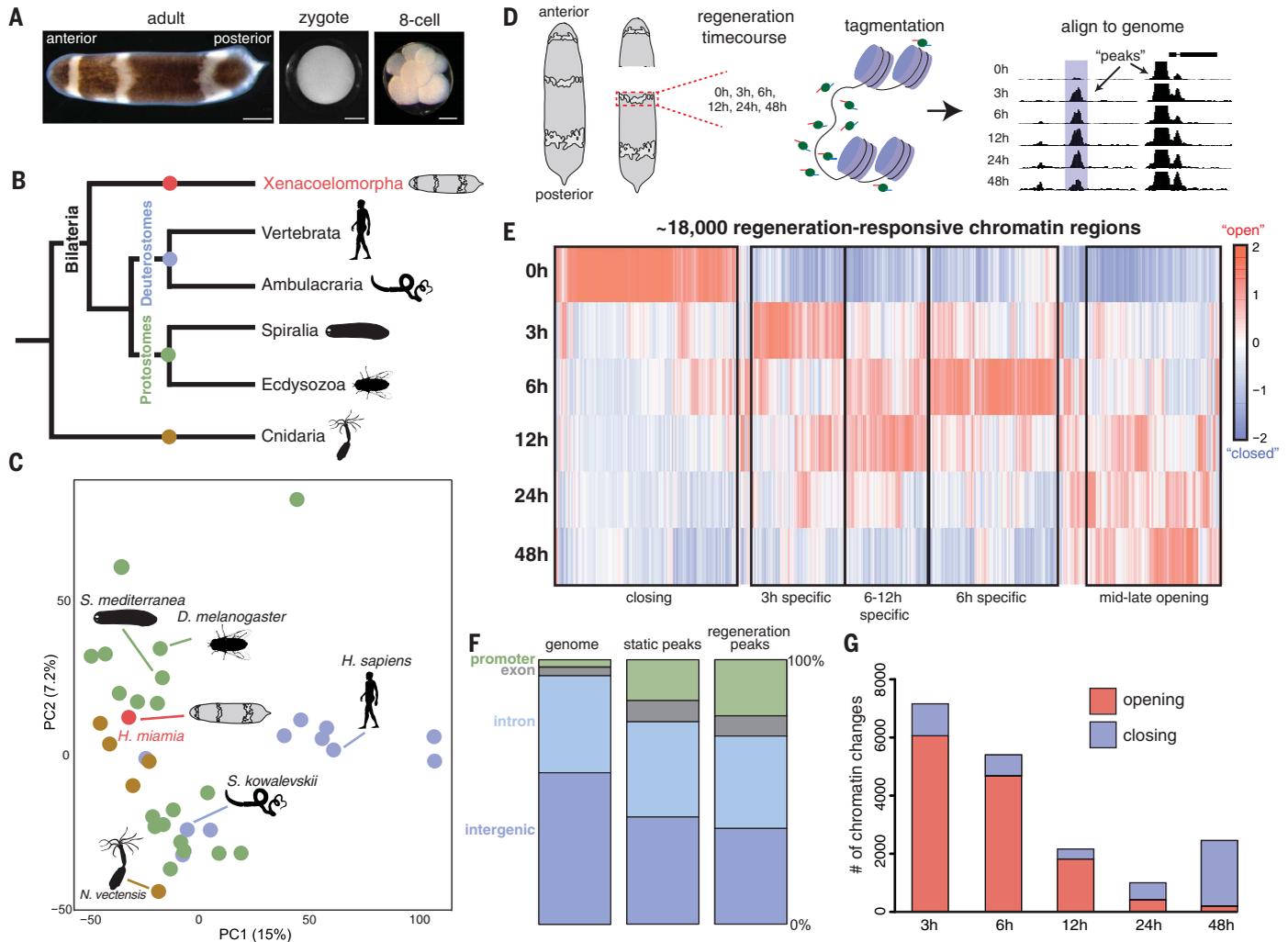


Fig. 1. The genome of the acoeel *Hofstenia miamia* and global chromatin dynamics of whole-body regeneration. (A) *Hofstenia miamia* adult, zygote, and eight-cell embryo. Scale bars represent 500 μ m for adult, 100 μ m for embryos. (B) Schematic phylogenetic tree showing the placement of acoeels (Phylum: Xenacoelomorpha) as the likely sister group to all other bilaterians, according to (12). (C) Principal components analysis of metazoan genomes showing that *Hofstenia* contains a standard complement of animal genes. Each dot represents a species with a sequenced genome and is colored green (protostome), blue (deuterostome), or brown (nonbilaterian species). For a full list of species in this graph, see fig. S2A and table S1, C and D. (D) Schematic of chromatin profiling workflow for heatmap in (E). Gray box

indicates a region of chromatin ("peak") that is regeneration-responsive (14). (E) Heatmap of all chromatin peaks that show significant changes during tail fragment ("new head") regeneration. Groups of chromatin regions that show similar temporal patterns of opening or closing are highlighted by black boxes. See fig. S2, B to F, for head fragment ("new tail") data. (F) Bar plots showing the proportions of total genome sequence, of static peaks, and of peaks dynamic during regeneration contained within promoters, exons, introns, and intergenic regions. Most regeneration-responsive peaks lie in introns and intergenic space. (G) Number of "new" (i.e., not present in any previous time point) chromatin changes at each time point. Animal silhouettes were obtained from PhyloPic (<http://phylopic.org/>).

dynamic chromatin, in our combined analyses of head and tail regeneration in *Hofstenia* (Fig. 2A, fig. S3A, and table S4C). Regions containing EGR binding motifs were closed at 0 hpa, then open at 3 hpa, with these sites remaining accessible until 48 hpa, when the chromatin regions encompassing these motifs closed (Fig. 2B and fig. S3B).

Egr is a master regulator of regeneration

As the EGR motif emerged as the top candidate from the chromVAR analysis and Egr homologs

have known roles upon wounding in other animals (24–30), we sought to understand the expression dynamics and function of this family of TFs during *Hofstenia* regeneration. The *Hofstenia* genome and transcriptome only contains a single gene encoding an Egr ortholog, *egr* (fig. S3C). In situ hybridization showed that this gene was up-regulated at both anterior- and posterior-facing wound sites of amputated animals as early as 1 hpa (Fig. 2C). Notably, wound-induced *egr* expression occurred in multiple cell types, including epidermis, muscle, neurons, and neoblasts, the stem cell population that underlies the re-

generative capacity of *Hofstenia* (fig. S3D). We inhibited the expression of *egr* using RNA interference (RNAi) (*egr*-RNAi) and found that *egr*-RNAi animals did not regenerate blastemas, resulting in a failure to make new heads or new tails (100%, $n = 150$) (Fig. 2C and fig. S3E).

These data suggest that *egr* might lie at the top of a gene regulatory network (GRN) associated with the initiation of regeneration. To test this hypothesis, we performed RNA sequencing (RNA-seq) at 0, 1, 3, 6, and 12 hpa using the same amputation strategy as for the ATAC-seq experiment. We identified 61 genes that were significantly

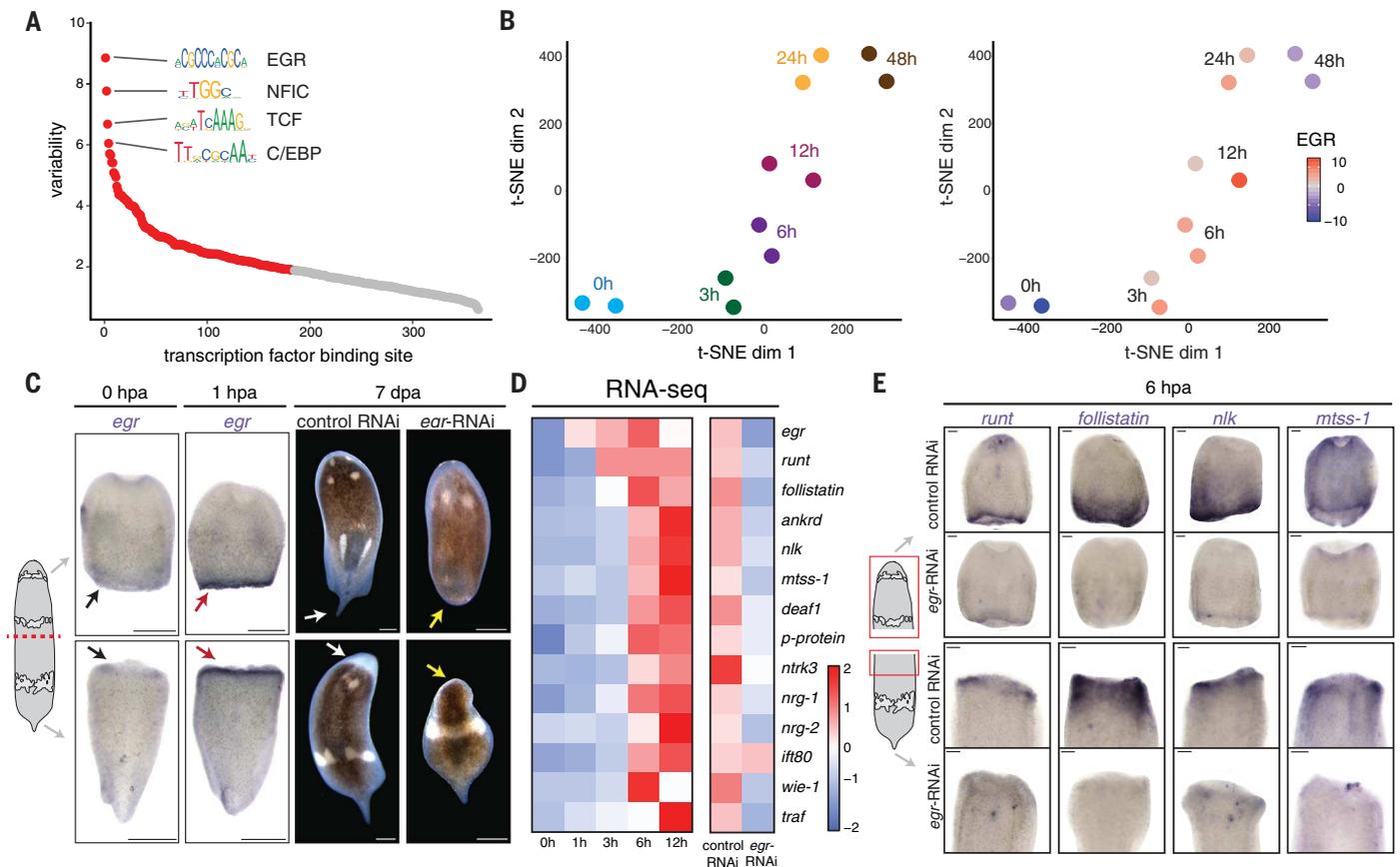


Fig. 2. Chromatin dynamics, expression, and functional assays implicate *egr* as a master regulator of regeneration in *Hofstenia*.

(A) Plot of chromVAR chromatin variability scores for 386 transcription factor binding motifs across all time points of regeneration assayed in both head and tail fragments, indicating EGR as the most variable motif. (B) t-Distributed stochastic neighbor embedding (t-SNE) plot showing groupings of samples, including biological replicates, based on chromatin accessibility of the consensus peak set. Samples are colored on the basis of their time point. Right panel shows the accessibility of EGR binding motifs overlaid on the same t-SNE plot, indicating opening of chromatin (red color) at sites by 3 hpa, followed by closing (blue color) at 48 hpa (see fig. S3B for combined tail and head fragment analysis). (C) Left: *egr* was expressed at the wound sites of head and tail fragments (red arrows) at 1 hpa relative to 0 hpa (black arrows) (26/28). Scale bars, 500 μ m. Right:

Compared to control RNAi animals that regenerated normally (white arrows), 100% ($n = 150$) *egr*-RNAi head and tail fragments failed to regenerate (yellow arrows). Phenotypes shown here were observed 7 days post-amputation (dpa). Dashed line in schematic shows plane of amputation. Scale bars, 200 μ m. (D) Heatmap of RNA-seq data showing the expression of 13 validated wound-induced genes in tail fragment samples during regeneration, including 6 hpa *egr*-RNAi animals. Twelve out of 13 wound-induced genes failed to be up-regulated ($p < 0.05$, likelihood ratio test) at 6 hpa in the *egr*-RNAi condition. Additionally, *egr*, which was targeted via RNAi, was also down-regulated relative to controls. (E) In situ hybridization validation of *egr*-mediated control of a subset of genes shown in (D). *egr*-RNAi animals failed to show wound-induced expression of *runt* (20 out of 22), *follistatin* (19 out of 19), *nlk* (22 out of 23), and *mtss-1* (19 out of 21) relative to controls. Scale bars, 200 μ m.

up-regulated (adjusted $p < 0.05$, likelihood ratio test) (14) in both head and tail fragments in at least one of the four time points. *egr* was the most significantly wound-induced gene at 1 hpa in both head and tail fragments (Fig. 2D and table S5). Twenty genes from this list yielded RNA probes with detectable expression, enabling us to validate by in situ hybridization 13 genes, in addition to *egr*, as wound-induced genes (up-regulated at wound sites by 6 hpa) (fig. S3F and table S5).

To determine whether the wound-induced activity of these 13 genes is under the control of *egr*, we performed RNA-seq after inhibition of *egr* via RNAi. Twelve out of 13 of these genes showed significantly lower expression at 6 hpa in *egr*-RNAi animals compared to controls (Fig. 2D and table S5). To validate these results, we

performed in situ hybridization for four potential target genes after *egr*-RNAi and amputation and found a decrease in expression when *egr* was inhibited (Fig. 2E and fig. S3G). Given that EGR binding sites show high chromatin variability, *egr* is wound-induced within 1 hpa in multiple cell types, is necessary for regeneration, and lies upstream of the majority of other validated early (1 to 12 hpa) wound-induced genes, we infer that *egr* is likely a master regulator of early regeneration.

Egr directly regulates wound-induced genes

We next sought to understand the mechanism(s) by which Egr protein regulates other wound-induced genes. We filtered our list of genome-wide regeneration-responsive chromatin peaks

for those that contained a predicted EGR binding site and assigned each peak to its closest gene (table S7, A to C) (14). To distinguish between spurious sites and true binding events, we performed ATAC-seq on *egr*-RNAi animals at 6 hpa. By focusing on peaks that experienced a significant decrease ($p < 0.05$, Wald test) in accessibility in *egr*-RNAi relative to control RNAi (which represent likely bona fide Egr targets), we could identify the Egr regeneration cistrome, i.e., the genome-wide cis-acting targets of Egr protein during regeneration (table S7, C and D) (14).

Among the list of putative *egr* targets, we found that 10 out of the 12 genes affected by *egr*-RNAi, and *egr* itself, had a predicted EGR binding site present in a neighboring dynamic peak (Fig. 3A and table S7). Analysis of ATAC-seq data in *egr*-RNAi at 6 hpa revealed that

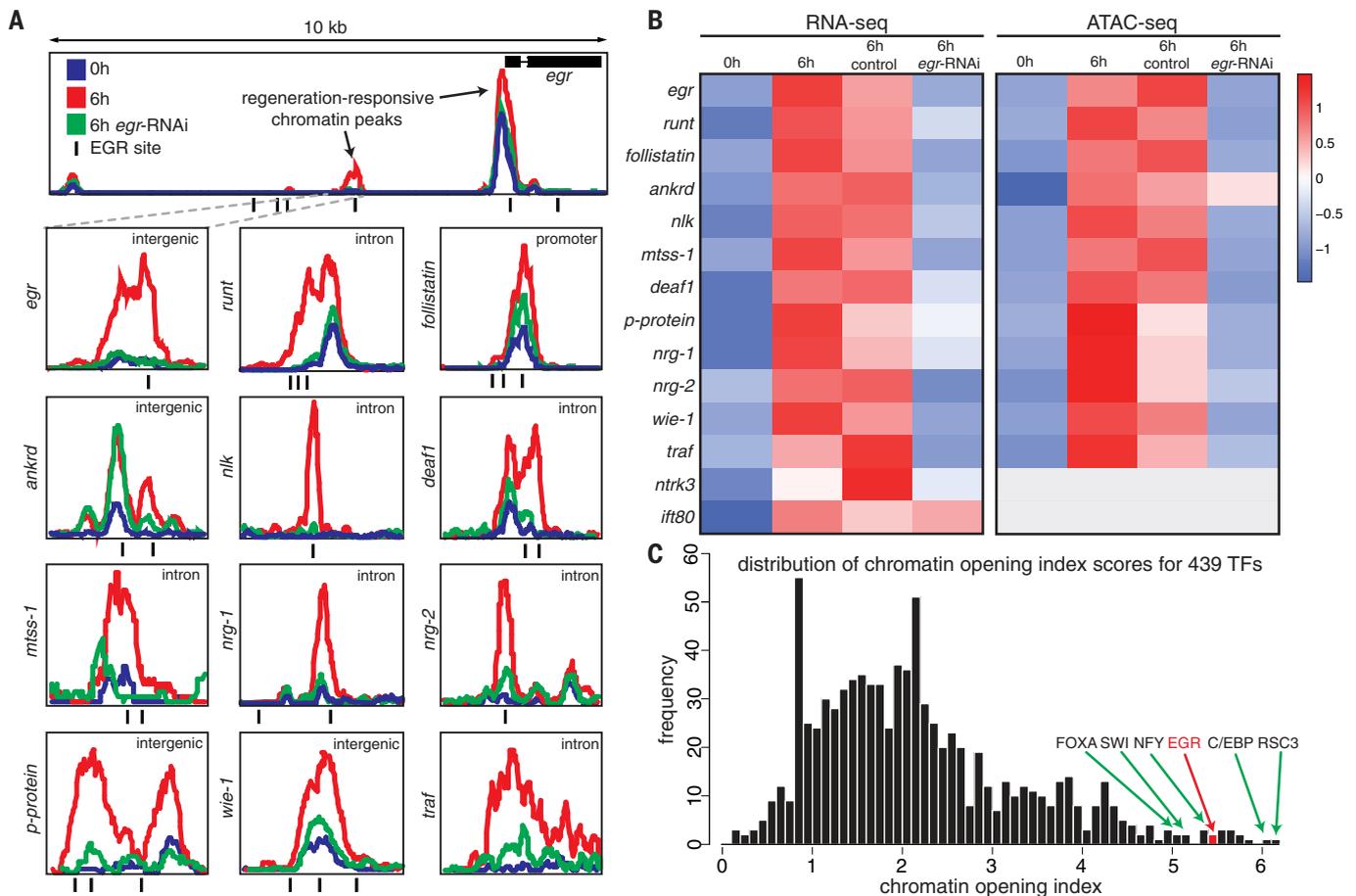


Fig. 3. Egr is a direct transcriptional regulator of wound-induced genes and likely has pioneer factor activity. (A) ATAC-seq tail fragment profiles at 0 hpa (blue), 6 hpa (red), and *egr*-RNAi 6 hpa (green) at the loci of validated wound-induced genes. Predicted EGR sites are denoted as black tick marks below. A 10-kb window containing the *egr* locus is shown, with magnified views (~800 bps) of regeneration-responsive peaks in the loci of *egr* and other wound-induced genes shown below. See fig. S4 for images of all peaks in the context of larger surrounding genomic regions. (B) Heatmaps

these dynamic peaks depend upon the action of Egr, as the regeneration-responsive chromatin regions that contain EGR-binding sites failed to open at 6 hpa in *egr*-RNAi animals and instead resembled the chromatin profile of the control (time 0) fragments (Figs. 3A and fig. S4). These experiments provide functional validation of predicted binding sites in the loci of wound-induced genes and suggest that Egr directly regulates their transcription.

We propose that the induction of *egr* expression upon amputation launches a cascade of transcriptional events, with the 10 wound-induced genes identified here as a subset of direct Egr targets. The Egr targets encode diverse molecular functions, including homologs of two transcription factors (*runt*, *deaf1*), components of several signaling pathways (*follistatin*, *nrg-1*, *nrg-2*, *nlk*), and cytosolic proteins with varied functions (*p-protein*, *ankrd*, *mtss-1*), as well as a protein with unclear homology (*wound induced expression-1*; *wie-1*). These targets are likely up-regulated in different cell types to mediate varied down-

stream effects: for example, *follistatin* was specifically expressed and up-regulated in longitudinal muscle after injury in *Hofstenia* (figs. S3H and S6B), much like its ortholog in planarians (31).

This preliminary regeneration GRN appears to have an autoregulatory loop, as the *egr* locus itself has an EGR binding site contained within a regeneration-responsive peak that was lost upon *egr*-RNAi. Two wound-induced genes, the homologs of a growth factor receptor (*ntrk3*) and a tumor necrosis factor (TNF) receptor-associated factor (*traf*), showed significantly diminished expression in *egr*-RNAi, but the lack of EGR binding sites in their loci suggests either the presence of cryptic EGR binding sites or that they are indirectly regulated by Egr. The one validated wound-induced gene, a homolog of an intracellular transport protein (*ift80*), that did not lose its wound-induced expression activity in the *egr*-RNAi condition also did not have a regeneration-responsive chromatin region containing an EGR site, suggesting that alternative, *egr*-independent pathways also operate

of RNA-seq (transcript reads) and ATAC-seq (genomic DNA reads) profiles of validated wound-induced genes, showing correspondence of statistically significant changes of both gene expression profile and corresponding chromatin accessibility (tables S3E, S5A, and S7) (14). (C) Frequency distribution of PIQ-derived chromatin opening index scores for 439 JASPAR motifs at the 6 hpa time point of tail fragment regeneration. Green arrows indicate factors among the top 11 TFs that have previously reported pioneer factor activity, with the score for the EGR motif highlighted in red.

during early stages of *Hofstenia* regeneration (Fig. 3B). We noted a correspondence of chromatin profiles at EGR binding sites with RNA expression levels of the associated genes in our analysis of validated wound-induced genes (Fig. 3B), suggesting a causal relationship between specific chromatin loci and their target genes.

Egr shows pioneer factor-like activity

Peaks containing EGR sites in the loci of wound-induced genes appeared to have a nearly binary response, where chromatin at most EGR sites seemed to switch from “closed” to “open” (e.g., in the loci of *egr*, *nlk*, *nrg-1*) (Fig. 3A, fig. S4, and table S7). These chromatin differences were not observed in the *egr*-RNAi condition, where chromatin failed to open at these loci at 6 hpa relative to control RNAi (Fig. 3, A and B). The pronounced opening of chromatin at EGR sites during regeneration suggests that Egr could be functioning as a pioneer factor.

Pioneer factors are TFs that bind to closed chromatin, resulting in an opening of the region

to provide access to other TFs (32). To assess pioneer-like activity of TFs in *Hofstenia* in a systematic and unbiased manner, we used the protein interaction quantitation (PIQ) algorithm (33), which calculates a “chromatin opening index” that has been used to both validate and predict new pioneer TFs in human data, on the basis of the frequency of transposase cutting sites surrounding TF footprints (14, 33). Pioneer factor footprints show increased levels of local transposase integration relative to those of other TFs. When applied to 6 hpa ATAC-seq data in *Hofstenia*, we found that the EGR motif exhibited a high chromatin opening index, ranking among the top five of the 439 binding sites analyzed. Further, EGR was nested among other TFs that have known pioneer activity (Fig. 3C and table S8). Thus, we suggest that *egr* is a master control gene that regulates other wound-induced genes by acting as a pioneer factor in *Hofstenia*, representing a potentially previously unknown role for this protein family. We note that *Hofstenia* Egr can also likely act as a more traditional TF regulator by binding to open promoter regions of certain targets (e.g., *follicistatin*).

Extending the Egr-controlled regeneration GRN

Connecting Egr and its direct targets allowed us to assemble the first tier of a GRN for early stages of regeneration. Our data indicate that Egr directly regulates at least two other genes encoding TFs, *runt* and *deaf1*. We next sought to predict and validate targets of these two Egr-controlled TFs. *deaf1*-RNAi animals regenerated normally, but *runt*-RNAi animals failed to regenerate tails (28 out of 33) and heads (19 out of 29). The phenotype of *runt*-RNAi was binary: either complete or no regeneration (Fig. 4A). Furthermore, *runt* homologs are known to be wound-induced and are required for regeneration in planarians (2).

runt was the only detected Runt family homolog in *Hofstenia* (fig. S6A). The predicted RUNT binding motif exhibited a relatively low chromatin opening index score (table S8), suggesting that changes in ATAC-seq peaks associated with this motif during regeneration might be less substantial than for those associated with Egr. We found that regeneration-responsive chromatin peaks were significantly lower in peaks with

RUNT sites relative to those for EGR sites (Fig. 4B) ($p < 10^{-10}$, two-sample t test). These results suggest that Runt is not acting as a pioneer factor and thus would not be expected to elicit large changes in chromatin accessibility.

We sought to assess changes in the binding of Runt protein to its motif genome-wide without relying on large changes in chromatin accessibility. We used PIQ to assess, for every RUNT site in open chromatin in the *Hofstenia* genome, whether the Runt protein was predicted to be “bound” (binding score ≥ 0.7) or “unbound” (binding score < 0.7) (33) on the basis of transposase accessibility at binding sites (footprinting) (table S9, A and B). By plotting the distribution of binding indices for RUNT sites in 0 hpa, 6 hpa, and 6 hpa *egr*-RNAi data, we recovered a distinct population of TF binding sites with high binding indices at 6 hpa but not at 0 hpa, which remain “unbound” in *egr*-RNAi animals (Fig. 4C and table S9A). Because *egr* is required for activation of *runt*, this population of sites represents RUNT sites that become “newly” bound during regeneration, likely upon Runt up-regulation by Egr.

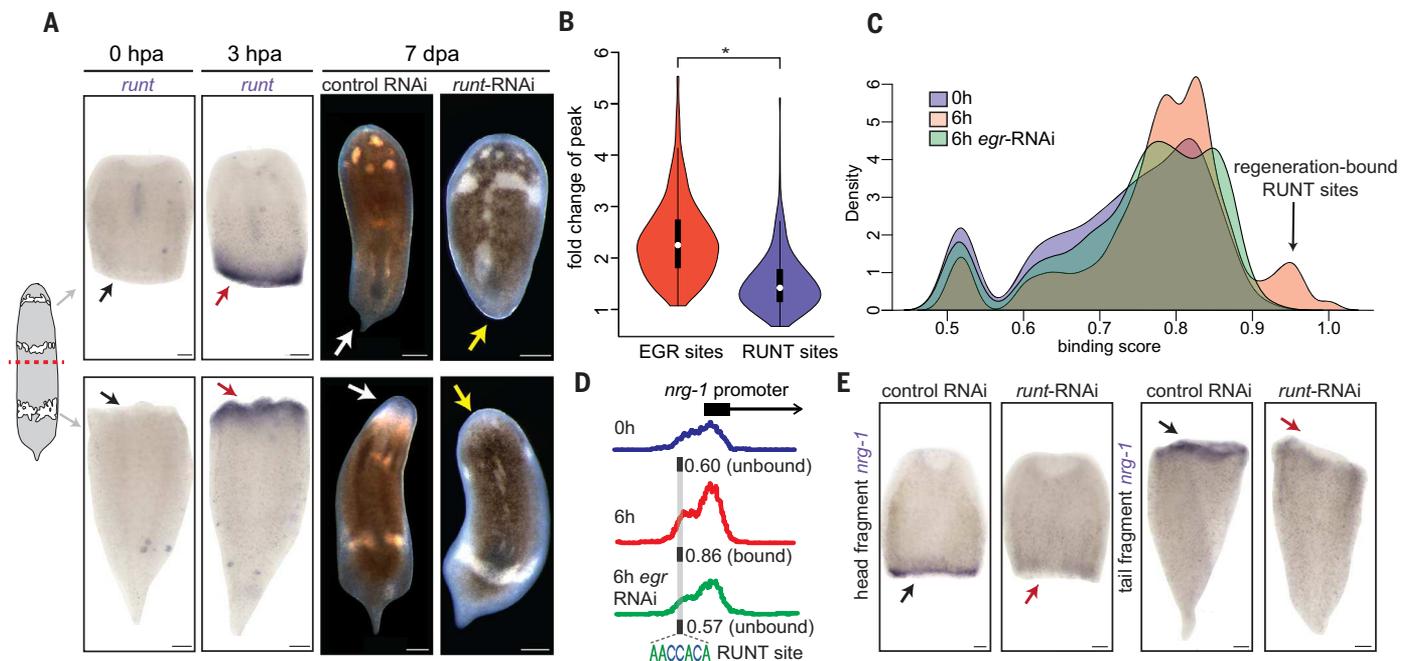


Fig. 4. *runt* is a downstream target of Egr, is wound-induced and necessary for regeneration, and directly controls *nrg-1*. (A) Left: In situ hybridization showing that *runt* is wound-induced at 3 hpa (red arrows) relative to the 0 hpa control (black arrows). Scale bars, 300 μ m. Right: In contrast to normal regeneration in controls (white arrows), *runt*-RNAi resulted in lack of tail (28 out of 33) and head regeneration (19 out of 29) (yellow arrows). Phenotypes observed 7 days post-amputation (dpa) are shown. Dashed line in schematic shows plane of amputation. Scale bars, 200 μ m. (B) Violin plot comparing the distribution of fold change of all regeneration-responsive peaks at 6 hpa (relative to 0 hpa) at EGR and RUNT sites. Regeneration-responsive peaks that contain RUNT sites show significantly lower magnitude of change compared to those that contain EGR sites. Asterisk indicates statistical significance in a two-sample t test.

(C) Distribution (viewed as density plots) of RUNT binding scores derived from PIQ in the 0 hpa, 6 hpa, and 6 hpa *egr*-RNAi data samples. RUNT sites that become highly “bound” at 6 hpa, and are dependent on *egr*, are indicated by a black arrow. (D) Schematic of the *nrg-1* promoter with ATAC-seq peaks from 0 hpa, 6 hpa, and 6 hpa *egr*-RNAi data samples shown below. The location of a predicted RUNT binding site in this locus is shown as a black line, with the corresponding PIQ binding score for each condition shown next to the site. PIQ scores above 0.7 indicate a “bound” state for the TF at the given site. (E) In situ hybridization of the *nrg-1* gene in head and tail fragments at 6 hpa in control and *runt*-RNAi conditions. Normal wound-induced expression in controls is highlighted by black arrows, loss of expression in *runt*-RNAi is highlighted by red arrows (18 out of 38). Animals were amputated as per the schematic in (A). Scale bars, 200 μ m.

Finally, to predict potential targets of bound RUNT sites during regeneration we sorted RUNT sites in descending order of the magnitude of change between the 6- and 0-hpa binding indices and examined the top 10% as a conservative estimate of “newly bound” sites. To identify the RUNT sites most dependent on the Egr-initiated wound response, we ranked these sites by the magnitude of change in binding indices between 6- and 6-hpa *egr*-RNAi data (table S9). Notably, the second-highest ranked RUNT site (“newly” bound and downstream of *egr*) in this list was closest to the gene *neuregulin-1* (*nrg-1*), which was identified as a wound-induced gene in *Hofstenia* in our RNA-seq analysis. *nrg-1* is an epidermal growth factor (EGF) family ligand whose homologs are known to operate in regeneration in planarians (34) and vertebrates (35, 36) (Fig. 4D, fig. S6C, and table S9). Our ATAC-seq analysis had identified *nrg-1* as being under direct control by Egr in *Hofstenia* (Fig. 3A and fig. S5A). Owing to the substantial change in RUNT binding score at this locus, we hypothesized that *nrg-1* is also under direct control of Runt in *Hofstenia* (Fig. 4D). To test this, we inhibited *runt* via RNAi, amputated animals, and assessed the expression of *nrg-1* at 6 hpa using *in situ* hybridization. We found that *nrg-1* expression was lost or decreased in the absence of *runt* activity (18 out of 38), validating that *egr-runt-nrg-1* functions as a regulatory cascade in *Hofstenia* (Fig. 4E and fig. S5B).

Regulatory networks and the evolution of whole-body regeneration

Egr family members are well known for their roles as immediate early genes (IEGs) in many

injury and regeneration contexts (24, 25), including in neuronal (26), cardiac (27), liver (24), and immune cells (28) of vertebrates, and have been reported as significantly induced during injury and regeneration in cnidarians (29), planarians (2, 3), and sea stars (30). Here, we identified whole-body regeneration regulatory elements in the acoel *Hofstenia* to construct a regeneration GRN launched upon wounding and controlled by the master regulator Egr (Fig. 5A).

To assess the relevance of our findings to other species, we generated ATAC-seq data for the regenerating planarian (*Schmidtea mediterranea*) and found that, as observed in *Hofstenia*, the EGR motif is among the most variably accessible during regeneration (Fig. 5B and table S10). This finding indicates that TFs of the Egr family might also globally regulate chromatin and act as master regulators of regeneration in planarians, which is consistent with the early wound-induced expression of several Egr homologs, one of which is necessary for neural regeneration, in this system (2, 37) (Fig. 5C). Furthermore, homologs of the Egr-controlled GRN are either required for regeneration or are up-regulated during regeneration in distantly related species, including vertebrates (35, 36), planarians (2, 34, 38, 39), and sea stars (30) (Fig. 5C and fig. S6). Constructing regeneration GRNs in additional animals will be crucial to understanding if these networks are conserved or evolved convergently.

Understanding the gene networks that underlie complex processes requires a knowledge of regulatory interactions. In this study, we used the acoel genome to map functional binding sites and to construct a GRN initiating whole-

body regeneration. By viewing whole-body regeneration through an epigenomic lens, we uncovered a regulatory mechanism underlying dynamic gene transcription during regeneration. This approach can be applied to established and emerging model systems to gain a deeper mechanistic understanding of the transcriptional cascades that underlie the phenomenon of regeneration.

Material and methods summary

Short-read sequence data were assembled to generate a draft genome for *Hofstenia miamia*, which was validated using long-read sequencing. ATAC-seq data were obtained from wound sites at “zero” (control), 3, 6, 12, 24, and 48 hpa. These data were used to define regions of chromatin that changed significantly as regeneration progressed. Motif analysis in these regions was used to rank TF binding sites on the basis of dynamic overlying chromatin. The requirement of the top-ranked TF (Egr) for regeneration in *Hofstenia* was studied via RNAi. RNA-seq data were obtained from wild-type and control and *egr*-RNAi regenerating worms to identify wound-induced genes and the subset of those under Egr control. The results of these analyses were validated by RNAi and *in situ* hybridization. To assess direct transcriptional control, regeneration-responsive chromatin regions were assigned to their nearest gene, and ATAC-seq data from control and *egr*-RNAi worms were analyzed to test for significant differences. To extend the GRN, we used binding site footprinting in our ATAC-seq data to identify dynamically bound TFs, which were linked to their nearest gene as putative targets and validated using RNAi.

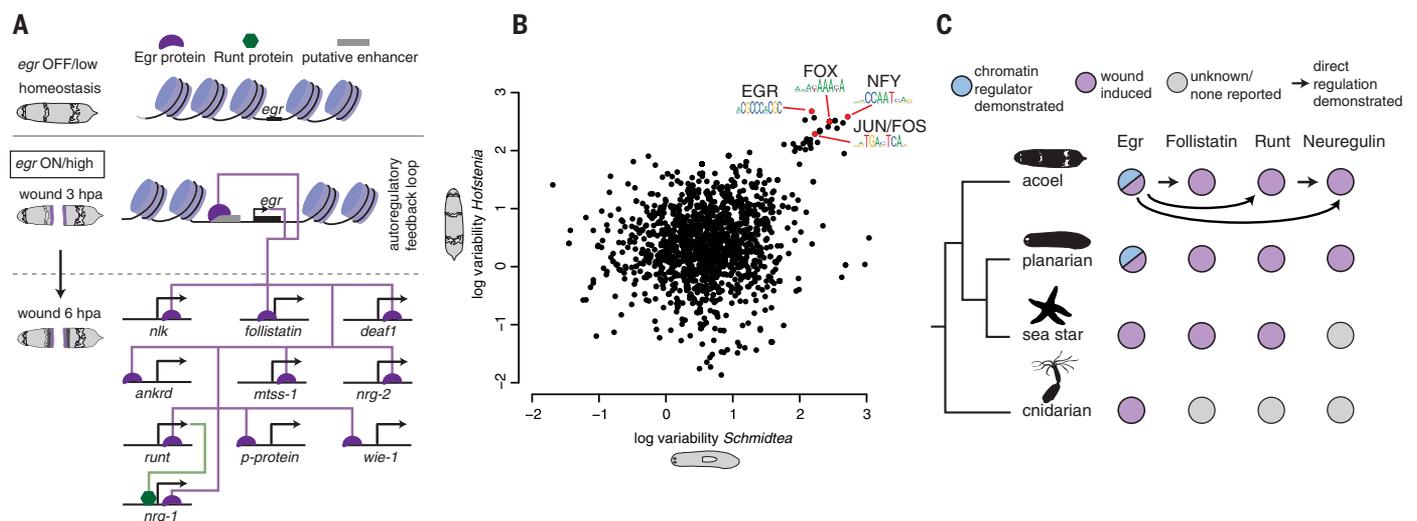


Fig. 5. Model and evolution of a gene regulatory network for early regeneration. (A) Model of a gene regulatory network. *egr* (purple color at wound sites) is wound-induced early, forms an autoregulatory loop, and acts as a likely pioneer factor to directly regulate a suite of downstream genes. This includes the TF *runt* (green color at wound sites), which binds to the promoter of *nrg-1* to up-regulate its expression. Purple shapes and lines indicate Egr protein and its activity, respectively; green shape and line indicate Runt protein and its activity. Blue disks indicate nucleosomes. (B) chromVAR chromatin variability scores for 6 hpa ATAC-seq data from the planarian *S. mediterranea*

and the acoel *H. miamia*. Select motifs and their associated TF families that are highly variable in both taxa are shown in red. (C) Schematic comparing a portion of the early regeneration network described in this work to what is known in other species capable of whole-body regeneration. Blue color indicates a role in regulating genome-wide chromatin accessibility as determined in this study, purple indicates that the gene is known to be wound-induced. Black arrows indicate direct regulation determined by this study, and gray circles indicate unknown or unstudied expression or function. Sea star and cnidarian silhouettes were obtained from PhyloPic (<http://phylopic.org/>).

REFERENCES AND NOTES

- A. E. Bely, K. G. Nyberg, Evolution of animal regeneration: Re-emergence of a field. *Trends Ecol. Evol.* **25**, 161–170 (2010). doi: [10.1016/j.tree.2009.08.005](https://doi.org/10.1016/j.tree.2009.08.005); pmid: 19800144
- D. Wenemoser, S. W. Lapan, A. W. Wilkinson, G. W. Bell, P. W. Reddien, A molecular wound response program associated with regeneration initiation in planarians. *Genes Dev.* **26**, 988–1002 (2012). doi: [10.1101/gad.187377.112](https://doi.org/10.1101/gad.187377.112); pmid: 22549959
- O. Wurtzel et al., A generic and cell-type-specific wound response precedes regeneration in planarians. *Dev. Cell* **35**, 632–645 (2015). doi: [10.1016/j.devcel.2015.11.004](https://doi.org/10.1016/j.devcel.2015.11.004); pmid: 26651295
- S. J. Zhu, B. J. Pearson, (Neo)blast from the past: New insights into planarian stem cell lineages. *Curr. Opin. Genet. Dev.* **40**, 74–80 (2016). doi: [10.1016/j.gde.2016.06.007](https://doi.org/10.1016/j.gde.2016.06.007); pmid: 27379899
- C. P. Petersen, P. W. Reddien, Smed- β catenin-1 is required for anteroposterior blastema polarity in planarian regeneration. *Science* **319**, 327–330 (2008). doi: [10.1126/science.1149943](https://doi.org/10.1126/science.1149943); pmid: 18063755
- K. A. Gurley, J. C. Rink, A. Sánchez Alvarado, β -Catenin defines head versus tail identity during planarian regeneration and homeostasis. *Science* **319**, 323–327 (2008). doi: [10.1126/science.1150029](https://doi.org/10.1126/science.1150029); pmid: 18063757
- C. P. Petersen, P. W. Reddien, Polarized *notum* activation at wounds inhibits Wnt function to promote planarian head regeneration. *Science* **332**, 852–855 (2011). doi: [10.1126/science.1202143](https://doi.org/10.1126/science.1202143); pmid: 21566195
- J. Kang et al., Modulation of tissue repair by regeneration enhancer elements. *Nature* **532**, 201–206 (2016). doi: [10.1038/nature17644](https://doi.org/10.1038/nature17644); pmid: 27049946
- J. A. Goldman et al., Resolving heart regeneration by replacement histone profiling. *Dev. Cell* **40**, 392–404.e5 (2017). doi: [10.1016/j.devcel.2017.01.013](https://doi.org/10.1016/j.devcel.2017.01.013); pmid: 28245924
- R. E. Harris, L. Setiawan, J. Saul, I. K. Hariharan, Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature *Drosophila* imaginal discs. *eLife* **5**, e11588 (2016). doi: [10.7554/eLife.11588](https://doi.org/10.7554/eLife.11588); pmid: 26840050
- M. Srivastava, K. L. Mazza-Curll, J. C. van Wolfswinkel, P. W. Reddien, Whole-body acoeel regeneration is controlled by Wnt and Bmp-Admp signaling. *Curr. Biol.* **24**, 1107–1113 (2014). doi: [10.1016/j.cub.2014.03.042](https://doi.org/10.1016/j.cub.2014.03.042); pmid: 24768051
- J. T. Cannon et al., Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89–93 (2016). doi: [10.1038/nature16520](https://doi.org/10.1038/nature16520); pmid: 26842059
- G. W. Rouse, N. G. Wilson, J. I. Carvajal, R. C. Vrijenhoek, New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. *Nature* **530**, 94–97 (2016). doi: [10.1038/nature16545](https://doi.org/10.1038/nature16545); pmid: 26842060
- Material and methods are available as supplementary materials.
- R. Li et al., De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010). doi: [10.1101/gr.097261.109](https://doi.org/10.1101/gr.097261.109); pmid: 20019144
- J. G. Ruby, P. Bellare, J. L. Derisi, PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. *G3* **3**, 865–880 (2013). doi: [10.1534/g3.113.005967](https://doi.org/10.1534/g3.113.005967); pmid: 23550143
- M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011). doi: [10.1093/bioinformatics/btq683](https://doi.org/10.1093/bioinformatics/btq683); pmid: 21149342
- N. H. Putnam et al., Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016). doi: [10.1101/gr.193474.115](https://doi.org/10.1101/gr.193474.115); pmid: 26848124
- M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntentically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008). doi: [10.1093/bioinformatics/btn013](https://doi.org/10.1093/bioinformatics/btn013); pmid: 18218656
- F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). doi: [10.1093/bioinformatics/btv351](https://doi.org/10.1093/bioinformatics/btv351); pmid: 26059717
- J. D. Buenostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013). doi: [10.1038/nmeth.2688](https://doi.org/10.1038/nmeth.2688); pmid: 24097267
- A. A. Raz, M. Srivastava, R. Salvamoser, P. W. Reddien, Acoel regeneration mechanisms indicate an ancient role for muscle in regenerative patterning. *Nat. Commun.* **8**, 1260 (2017). doi: [10.1038/s41467-017-01148-5](https://doi.org/10.1038/s41467-017-01148-5); pmid: 29084955
- A. N. Schep, B. Wu, J. D. Buenostro, W. J. Greenleaf, chromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017). doi: [10.1038/nmeth.4401](https://doi.org/10.1038/nmeth.4401); pmid: 28825706
- M. T. Pritchard, L. E. Nagy, Ethanol-induced liver injury: Potential roles for *egr-1*. *Alcohol. Clin. Exp. Res.* **29** (Suppl), 146S–150S (2005). doi: [10.1097/al.0c0000189286.81943.51](https://doi.org/10.1097/al.0c0000189286.81943.51); pmid: 16344600
- S. Bahrami, F. Drabløs, Gene regulation in the immediate-early response process. *Adv. Biol. Regul.* **62**, 37–49 (2016). doi: [10.1016/j.jbior.2016.05.001](https://doi.org/10.1016/j.jbior.2016.05.001); pmid: 27220739
- B. Pérez-Cadahía, B. Drobic, J. R. Davie, Activation and function of immediate-early genes in the nervous system. *Biochem. Cell Biol.* **89**, 61–73 (2011). pmid: 21326363
- L. M. Khachigian, Early growth response-1 in cardiovascular pathobiology. *Circ. Res.* **98**, 186–191 (2006). doi: [10.1161/01.RES.0000200177.53882.c3](https://doi.org/10.1161/01.RES.0000200177.53882.c3); pmid: 16456111
- D. Gómez-Martín, M. Díaz-Zamudio, M. Galindo-Campos, J. Alcocer-Varela, Early growth response transcription factors and the modulation of immune response: Implications towards autoimmunity. *Autoimmun. Rev.* **9**, 454–458 (2010). doi: [10.1016/j.autrev.2009.12.006](https://doi.org/10.1016/j.autrev.2009.12.006); pmid: 20035903
- R. Eiran et al., Early and late response of *Nematostella vectensis* transcriptome to heavy metals. *Mol. Ecol.* **23**, 4722–4736 (2014). doi: [10.1111/mec.12891](https://doi.org/10.1111/mec.12891); pmid: 25145541
- G. A. Cary, A. Wolff, O. Zueva, J. Pattinato, V. F. Hinman, Analysis of sea star larval regeneration reveals conserved processes of whole-body regeneration across the metazoa. *BMC Biol.* **17**, 16 (2019). doi: [10.1186/s12915-019-0633-9](https://doi.org/10.1186/s12915-019-0633-9); pmid: 30795750
- M. L. Scimone, L. E. Cote, P. W. Reddien, Orthogonal muscle fibres have different instructive roles in planarian regeneration. *Nature* **551**, 623–628 (2017). pmid: 29168507
- K. S. Zaret, S. E. Mango, Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016). doi: [10.1016/j.gde.2015.12.003](https://doi.org/10.1016/j.gde.2015.12.003); pmid: 26826681
- R. I. Sherwood et al., Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014). doi: [10.1038/nbt.2798](https://doi.org/10.1038/nbt.2798); pmid: 24441470
- K. Lei et al., Egf signaling directs neoblast repopulation by regulating asymmetric cell division in planarians. *Dev. Cell* **38**, 413–429 (2016). doi: [10.1016/j.devcel.2016.07.012](https://doi.org/10.1016/j.devcel.2016.07.012); pmid: 27523733
- J. E. Farkas, P. D. Freitas, D. M. Bryant, J. L. Whited, J. R. Monaghan, Neuregulin-1 signaling is essential for nerve-dependent axolotl limb regeneration. *Development* **143**, 2724–2731 (2016). doi: [10.1242/dev.133363](https://doi.org/10.1242/dev.133363); pmid: 27317805
- M. Gemberling, R. Karra, A. L. Dickson, K. D. Poss, Nrg1 is an injury-induced cardiomyocyte mitogen for the endogenous heart regeneration program in zebrafish. *eLife* **4**, e05871 (2015). doi: [10.7554/eLife.05871](https://doi.org/10.7554/eLife.05871); pmid: 25830562
- S. Fraguas, S. Barberán, M. Iglesias, G. Rodríguez-Esteban, F. Cebriá, *egr-4*, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians. *Development* **141**, 1835–1847 (2014). doi: [10.1242/dev.101345](https://doi.org/10.1242/dev.101345); pmid: 24700819
- R. H. Roberts-Galbraith, P. A. Newmark, Follistatin antagonizes activin signaling and acts with Notum to direct planarian head regeneration. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1363–1368 (2013). doi: [10.1073/pnas.1214053110](https://doi.org/10.1073/pnas.1214053110); pmid: 23297191
- M. A. Gaviño, D. Wenemoser, I. E. Wang, P. W. Reddien, Tissue absence initiates regeneration through Follistatin-mediated inhibition of Activin signaling. *eLife* **2**, e00247 (2013). doi: [10.7554/eLife.00247](https://doi.org/10.7554/eLife.00247); pmid: 24040508
- A. R. Gehrke, E. Neveerett, Y.-J. Luo, A. Brandt, L. Ricci, R. E. Hulett, A. Gompers, J. G. Ruby, D. S. Rokhsar, P. W. Reddien, M. Srivastava, Acoel genome reveals the regulatory landscape of whole-body regeneration. *Zenodo* (2019); doi: [10.5281/zenodo.2547750](https://doi.org/10.5281/zenodo.2547750)

ACKNOWLEDGMENTS

We thank The Bauer Core Facility at Harvard University for technical support with sequencing, T. Sackton (Harvard University) for bioinformatics/statistical assistance, J. Gaspar (Harvard University) for NGmerge and removeChrom scripts, A. Freedman (Harvard University) for assistance with RNA-seq analysis, N. Jeffery and T. Ryan Gregory (University of Guelph) for genome size estimates, and A. Hochwagen (New York University) for help with preparing high-quality genomic DNA. We thank T. Capellini (Harvard University) and all members of the Srivastava Lab for critical reading of the manuscript. **Funding:** M.S. is supported by the Milton Fund of Harvard University, Searle Scholars Program, Smith Family Foundation, and the National Science Foundation (award no. 1652104). A.R.G. is supported by the Helen Jay Whitney Foundation. Y.-J.L. is supported by the Human Frontier Science Program. D.S.R. and A.B. were supported by NIH grant 1R01HD080708-01, and A.B. was supported by the Biomedical Big Training Program at UC Berkeley and NIH grant 5T32LM012417-03. D.S.R. acknowledges support from the Marthella Foskett Brown Chair in Biological Sciences. P.W.R. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** A.R.G. and M.S. conceived the project and designed experiments. M.S. and P.W.R. initiated the genome and regeneration transcriptome sequencing. A.R.G. performed ATAC-seq and RNA-seq library construction and data analysis, chromVAR and PIQ bioinformatics, and Oxford Nanopore Technology sequencing analysis. A.R.G., E.N., L.R., and A.G. performed in situ hybridization and RNAi experiments. Y.-J.L., A.B. J.G.R., D.S.R. and A.R.G. performed genome assembly and statistics. Y.-J.L. constructed an improved transcriptome. R.E.H. performed phylogenetic analysis of genes. A.R.G. and M.S. wrote the paper with input from all of the authors. **Competing interests:** D.S.R. is a member of the scientific advisory board of Dovetail Genomics LLC and a minor shareholder. **Data and materials availability:** The *Hofstenia* genome assembly has been deposited at GenBank under the accession SCFE00000000. Other genomic and transcriptomic data for this project can be found under NCBI BioProject PRJNA512373. Data for *Schmidtea* ATAC experiments can be found under BioProject PRJNA515075. Additionally, *Hofstenia* data are available for download and genome browsing via <http://srivastavalab.rc.fas.harvard.edu>. Source code for ATAC-seq, RNA-seq, chromVAR, PIQ, Oxford Nanopore, and REAPR pipelines can be found at zenodo (40). *Hofstenia* are distributed under a nonrestrictive materials transfer agreement allowing work for academic and nonprofit purposes.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6432/eaau6173/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S10
References (41–64)

28 June 2018; resubmitted 8 November 2018
Accepted 8 February 2019
[10.1126/science.aau6173](https://doi.org/10.1126/science.aau6173)

Acoel genome reveals the regulatory landscape of whole-body regeneration

Andrew R. Gehrke, Emily Neverett, Yi-Jyun Luo, Alexander Brandt, Lorenzo Ricci, Ryan E. Hulett, Annika Gompers, J. Graham Ruby, Daniel S. Rokhsar, Peter W. Reddien and Mansi Srivastava

Science **363** (6432), eaau6173.
DOI: 10.1126/science.aau6173

Acoel-regeneration regulatory landscapes

Some animals, including some types of worms, can undergo whole-body regeneration and replace virtually any missing cell type. Gehrke *et al.* sequenced and assembled the genome of *Hofstenia miamia*, a regenerative acoel worm species (see the Perspective by Alonge and Schatz). They identified a variable motif corresponding to regulation of the early growth response (*egr*) gene that was involved in regeneration. RNA interference experiments and validation in a second species showed that the protein Egr is a pioneer factor that stimulates regeneration.

Science, this issue p. eaau6173; see also p. 1152

ARTICLE TOOLS	http://science.sciencemag.org/content/363/6432/eaau6173
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2019/03/13/363.6432.eaau6173.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/363/6432/1152.full
REFERENCES	This article cites 64 articles, 11 of which you can access for free http://science.sciencemag.org/content/363/6432/eaau6173#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)