Regeneration Rosetta: An interactive web application to explore regeneration-associated gene expression and chromatin accessibility

Andrea Rau[1,2], Sumona Dhara[3], Ava Udvadia[3], Paul Auer[2]

[1] GABI, INRA, AgroParisTech, Universite Paris-Saclay, 78350, Jouy-en-Josas, France
[2] Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA
[3] Department of Biological Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA

Abstract Time-course high-throughput assays of gene expression and enhancer usage in zebrafish provide a valuable characterization of the dynamic mechanisms governing gene regulatory programs during CNS axon regeneration. To facilitate the exploration and functional interpretation of a set of fully-processed data on regeneration-associated temporal transcription networks, we have created an interactive web application called *Regeneration Rosetta*. Using either built-in or user-provided lists of genes in one of dozens of supported organisms, our web application facilitates the (1) visualization of clustered temporal expression trends; (2) identification of proximal and distal regions of accessible chromatin to expedite downstream motif analysis; and (3) description of enriched functional gene ontology categories. By enabling a straightforward interrogation of these rich data without extensive bioinformatic expertise, *Regeneration Rosetta* is broadly useful for both a deep investigation of time-dependent regulation during regeneration in zebrafish and hypothesis generation in other organisms.

Keywords CNS axon regeneration; gene expression; chromatin accessibility; functional enrichment; zebrafish; R/Shiny

Introduction

Axon degeneration accompanying central nervous system (CNS) injury or disease leads to a permanent loss of function in human patients. This is largely due to an inability of mammals to reinitiate axon growth in adult CNS neurons[1]. In contrast to mammals, adult teleost fish can fully regenerate CNS axons that reinnervate appropriate targets, enabling functional recovery from CNS injury[2]. Interestingly, fish and mammals share common mechanisms for wiring the nervous system during development, and both are known to downregulate developmental growth and guidance signaling pathways during nervous system maturation[3-4]. Thus, what appears to set fish apart is the ability to re-initiate a sustained program of axon growth in response to CNS injury.

Transcriptional changes have long been correlated with the intrinsic capacity for regenerative axon growth[5,6]. In order to understand the precise mechanisms governing gene regulatory programs during CNS axon regeneration, Dhara et al. (2019)[7] recently identified the dynamic changes in gene expression and enhancer usage in zebrafish over the full time-course of axon regeneration in CNS neurons that are capable of successful regeneration. Adult zebrafish were subjected to optic nerve crush injury, and regenerating retinas were dissected at various time-points post injury in order to identify the interactions among expressed genes, open chromatin, and transcription factor expression during CNS axon regeneration.

These data on regeneration-associated temporal transcription networks in zebrafish represent a rich source of information with wide potential use and insight for the broader regeneration

community. To this end, we provide fully processed data from Dhara et al. (2019) in an interactive web application, *Regeneration Rosetta*, as a means to explore, visualize, and functionally interpret regeneration-associated gene expression and chromatin accessibility. Using either built-in lists of differentially expressed (DE) genes from Dhara et al. (2019) or user-provided gene lists in one of 69 supported organisms from Ensembl, our web application facilitates (i) customized visualization of clustered temporal expression trends during optic nerve regeneration; (ii) identification of proximal and distal regions of open chromatin relative to the gene list to expedite downstream motif analysis via the MEME suite[8]; and (iii) gene ontology (GO) functional enrichment analysis. Similarly, using either built-in lists of differentially accessible chromatin from Dhara et al. (2019) or user-provided genomic coordinates of accessible chromatin, the application identifies proximal and distal genes relative to their position and their corresponding enriched GO categories.

The *Regeneration Rosetta* app represents a new paradigm to facilitate data sharing and re-use in the field of regeneration. This type of data sharing directly promotes the National Institutes of Health guidelines for ensuring rigor and reproducibility in pre-clinical research (https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-researchas). As large-scale genomic data become more common, tools that allow rapid querying and exploration of fully processed data (without the need for additional coding or pre-processing steps) will be critical in accelerating advances in the regeneration field.

Methods

Experimental design, data generation, and bioinformatic analyses

Comprehensive experimental details may be found in Dhara et al. (2019). Briefly, whole retinas were dissected from 7-9 month old adult zebrafish at 0, 2, 4, 7, or 12 days post injury (dpi), following an optic nerve crush lesion (n=3 at each time point). After extracting total RNA from these samples, RNA-seq libraries were prepared and sequenced as previously described. Retinal ganglion cells (RGCs) were collected from retinas at each time point using fluorescent activated cell sorting (FACS) and chromatin was subsequently isolated. ATAC-seq libraries were prepared and sequenced as previously described[7]. One ATAC-seq library was dropped due to poor sequence quality.

For the RNA-seq data, after merging technical replicates and validating sequence quality we quantified transcript abundance[9] and tested differential expression with respect to the initial time point (0dpi)[10], controlling the false discovery rate at 5%. For the ATAC-seq data, sequences were aligned [11] to the zebrafish reference and open chromatin regions were called[12]. For each region with a p-value < 10e-10, a 500bp “peaklet” was defined by anchoring on the mode of the peak signal[13]. Chromatin accessibility was quantified by counting the number of overlapping reads for each retained peaklet, and differential accessibility was calculated with respect to the initial time point (0dpi)[14], controlling the false discovery rate at 5%. Software versions and parameters are provided in Dhara et al. (2019). All genomic coordinates and annotations are reported with respect to the GRCz10 Danio rerio genome assembly and Ensembl 90 gene annotation for the zebrafish.

Integration of RNA-seq and ATAC-seq data

To link regions of accessible chromatin with gene expression, we calculated peaklet-to-gene distance based on the coordinates of the peaklet mode and the gene’s transcription start site

(TSS). A proximal peaklet was then defined as one that overlaps the TSS and/or is within ±1kb of the TSS, while a distal peaklet was defined as one within ±100kb of the TSS but not proximal. Users can optionally remove exonic peaklets from these lists, defined as those within 50bp of exonic regions but not overlapping a TSS. To identify genes that are proximal or distal to a given set of accessible chromatin (whether user-provided or through the built-in lists of differentially accessible chromatin provided in the app), users may choose to include all genes or only a subset of those identified to be DE at a particular time point.

For peaklets identified as proximal or distal to the query set of genes, a FASTA file of sequences and BED file of genomic coordinates may be downloaded by the user for further analysis; in addition, a CSV file providing the potentially many-to-many correspondences of proximal and distal peaklets to genes may also be downloaded.

Gene expression visualization and queries for alternative species

Several built-in gene lists, based on the results described in Dhara et al. (2019), are directly available within the *Regeneration Rosetta* app. These include the lists of DE genes (based on an FDR adjusted p-value < 0.01) compared to 0dpi, as well as pre-identified clusters with expression patterns roughly corresponding to established events in the regeneration process (down-regulation during early-, mid-, or late-regeneration, growth toward the midline, midline crossing, target selection, and brain innervation). Users may also employ the *Regeneration Rosetta* app to explore gene sets for one of 68 species (Table S1) in addition to zebrafish by providing the relevant Ensembl gene IDs[15]. These converted gene lists can be further narrowed to include only those genes found to be DE post-injury.

For a given set of genes, expression heatmaps using log fold-changes, log transcripts per million (TPM), or Z-scores of these measures are produced using *ComplexHeatmap*[16], where transcript clusters are identified for a given number of clusters using the K-means algorithm, and rows are ordered within each cluster according to hierarchical clustering (Euclidean distance, complete linkage). Samples may also be hierarchically clustered. A high-resolution heatmap may be resized and downloaded by the user.

Functional enrichment analysis

The *Regeneration Rosetta* performs on-the-fly functional enrichment analyses of GO terms for Biological Processes (BP), Cellular Components (CC), and Molecular Function (MF) for a given gene set using *topGO* (weight01 algorithm, Fisher test statistic, and gene universe defined as the set of expressed transcripts from Dhara et al. (2019). P-values are not adjusted for multiple correction, and only GO terms with raw p-values < 0.05 are reported; tables of enriched GO terms are displayed in an HTML table in the app and may be optionally downloaded as a CSV file, Excel spreadsheet, or PDF file.

Technical details of the Regeneration Rosetta

The *Regeneration Rosetta* interactive web app was built in R using the *Shiny* and *flexdashboard* packages. In addition to the other software packages already cited above, it makes use of the *data.table* and *RSQLite* R packages for fast data manipulation, *DT* for rendering HTML tables using JavaScript, *readxl* for parsing data from Excel spreadsheets, *dplyr* for data manipulation, and *tokenizers* to convert user-provided gene IDs into tokens.

Results

Regeneration Rosetta yields insight into cholesterol and lipid biosynthesis regulation during regeneration

Cholesterol biosynthesis pathways were found to be enriched during regeneration in Dhara et al. (2019) and have been previously shown to be important in axon regeneration in mouse[17]. To more deeply investigate their behavior during optic nerve regeneration in zebrafish and demonstrate some of the capabilities of the *Regeneration Rosetta*, we input a list of 125 genes with cholesterol-related GO terms obtained from the Mouse Genome Informatics (MGI) database[18] (Table S2), corresponding to 330 expressed zebrafish transcripts, into the *Regeneration Rosetta* app (Fig. 1A).

Of these 330 transcripts, 67 were DE in Dhara et al. (2019); focusing on this subset of transcripts, the *Regeneration Rosetta* produces a clustered heatmap of expression Z-scores across distinct stages of optic nerve regeneration (Figure 1B). Using on the associated GO terms from MGI, we found that the twenty transcripts with peak expression early in regeneration during the phase of growth towards the midline (2-4 dpi) were predominantly enriched in cholesterol metabolic genes, while the majority of those peaking during the midline crossing and target selection phases (4-7 dpi) were enriched in cholesterol biosynthetic pathways (Figure S1). Interestingly, transcripts differentially down-regulated during regeneration were enriched in negative regulation of cholesterol biosynthetic processes.

Among the cholesterol biosynthetic genes, we observed upregulation during mid-regeneration (4-7 dpi) of SREBF2, a known cholesterol master-regulatory transcription factor[18,19,20]. After downloading the FASTA files of the sequences for peaklets proximal and distal to the 67 DE cholesterol metabolic genes from the *Regeneration Rosetta*, AME motif analysis [21] revealed that 18.9% of proximal open chromatin and 7% of distal open chromatin was enriched in the SREBF2 motif. We found a number of genes with temporal expression profiles similar to SREBF2 that are proximal to these accessible binding sites, including *dhcr24*, *hmgcs1*, *hsd17b7*, *insig1*, *sqlea*, and *idi1* (Figure 1B). Interestingly, the proximal and distal peaks of *srebf1*, which exhibited peak expression during brain reinnervation (7-12 dpi), were also enriched for SREBF2 motifs; SREBF1 is a transcription factor related to SREBF2 that is known to regulate genes involved in fatty acid synthesis and lipid homeostasis [21,22,23], both of which have been shown to be important for axon growth and myelination during neurogenesis [23–25]. This suggests that SREBF2 could potentially regulate the transcriptional activity of SREBF1, which, in turn, promotes the expression of genes associated with later regenerative processes.

Regeneration Rosetta provides insight into evolutionary conserved regenerative mechanisms

Nervous system function is dependent on the development of highly specific connections between neurons and their direct targets. The molecular mechanisms regulating this network of connections are highly conserved across evolution[26, 27]. Unlike mammals, vertebrates such as fish and amphibians exhibit regenerative abilities of complex tissues and structures. Therefore, comparing regenerative capabilities across species will enable researchers to identify genes and transcriptional networks that are critical to the regenerative program.

To facilitate a cross-species comparison, the *Regeneration Rosetta* app enables queries of patterns of gene expression across injury models with different regenerative capacities. To illustrate, we compared CNS regeneration in lamprey and zebrafish. Following [27,28], we used the

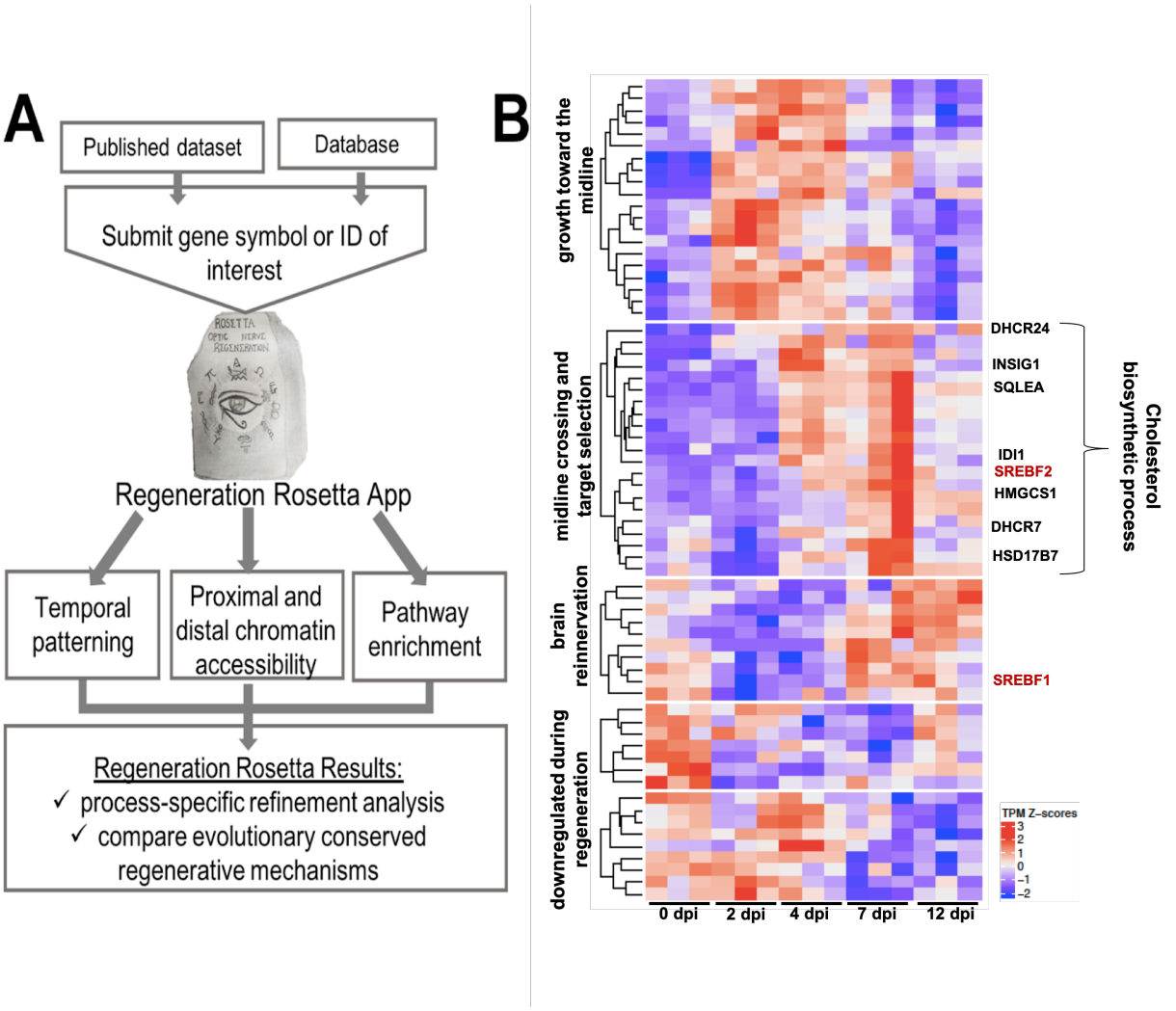

Fig.1 *Regeneration Rosetta* app identifies process-specific analysis after optic nerve injury. (A) Workflow for investigating temporal patterning of regeneration-associated genes classified within specific biological processes and/or comparative evolutionary analysis of the conserved mechanism among regenerative species, using the *Regeneration Rosetta* app. (B) Temporal transcript profiles of genes in the cholesterol metabolic pathway. Relative transcript counts from retinas dissected 2-, 4-, 7- and 12-days post injury (dpi) were compared with those from uninjured animals (0 dpi). Transcript expression is presented as TPM Z-scores; putative SREBF2 target genes are indicated to the right of the heatmap (biosynthetic enzymes in black; transcription factors are in red).

regenerating lamprey transcriptional profiles from cell bodies located in the brain and spinal cord following spinal cord injury over a course of 12 weeks. This study identified 3,664 and 3,999 differentially expressed regeneration-associated genes at one or more post-injury time points in lamprey brain and spinal cord, respectively. After removing duplicates, we filtered these lists to 2,325 (brain) and 2,520 (spinal cord) differentially expressed genes. We looked for an overlap of genes that were differentially regulated after injury to the zebrafish optic nerve and the lamprey spinal cord. We found 3,298 transcripts (corresponding to 1,971 genes) and 3,722 transcripts (corresponding to 2,151 genes) in the zebrafish optic regeneration data corresponding to the lamprey spinal cord and brain, respectively. After subsetting to those that were differential in the

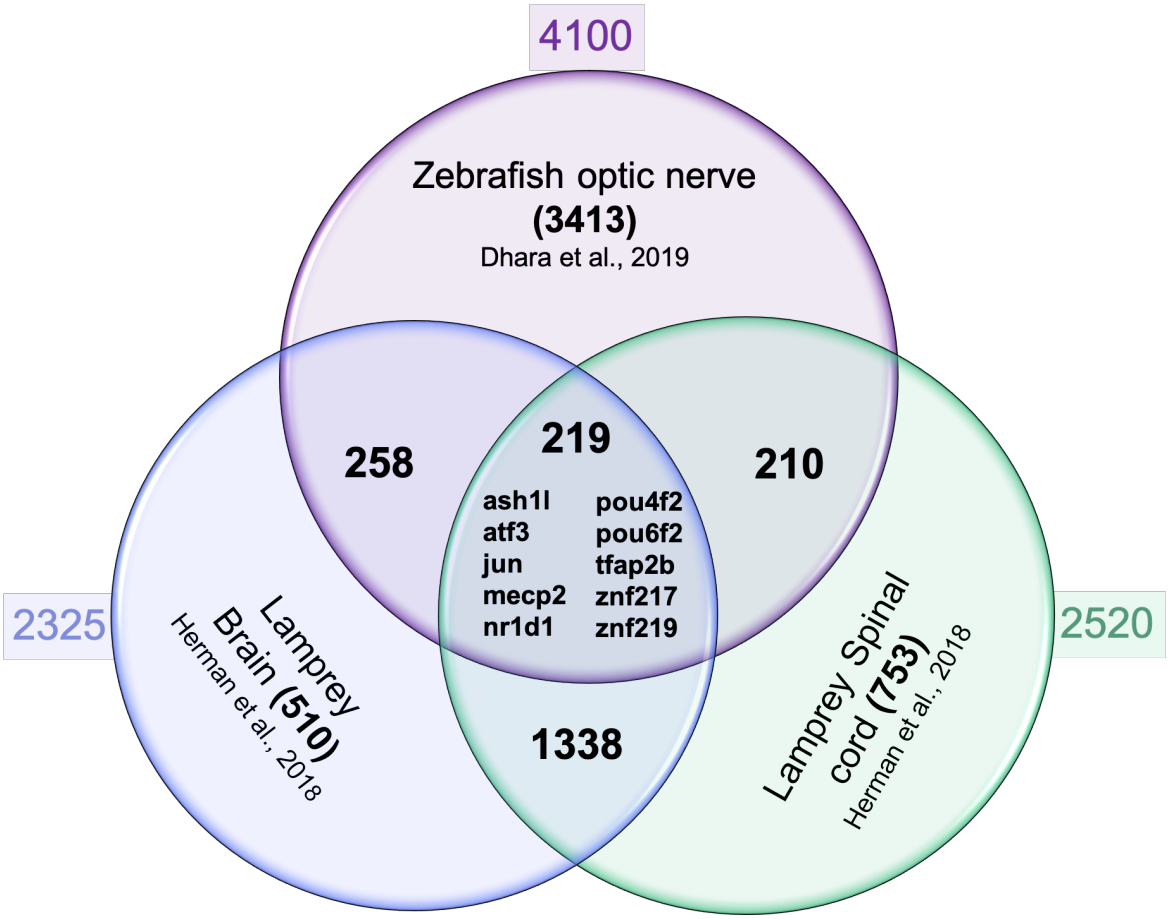

Fig. 2. *Regeneration Rosetta* app identifies conserved core regulators of CNS axon regeneration. Venn diagram of axon growth-associated genes from regenerating CNS neurons from zebrafish and lamprey. Approximately 9-10% of regeneration-associated genes are shared between neurons regenerating axons in brain, spinal cord and optic nerve, including a core set of 10 regeneration-associated transcription factors.

zebrafish (FDR < 1%), we identified 429 (spinal cord) and 477 (brain) differentially expressed genes (nearly 20% of genes identified as differentially expressed in any one study or more) common to the zebrafish optic regeneration, lamprey brain, and lamprey spinal cord (Fig. 2), suggesting a considerable overlap between the two regenerative models and three different tissue types.

To identify the core transcription factors that could potentially regulate stage-specific regeneration-associated gene transcription in both injury models, we cross referenced the lamprey list of DE transcripts to a recently compiled list of human transcription factors[29]. We identified 105 brain- and 131 spinal- transcription factor encoding genes that were differentially expressed at one or more post-injury time points (Table S4 and S5). Assessing the combined list of DE transcription factors, the *Regeneration Rosetta* revealed that 17 transcripts corresponding to 10 transcription factor encoding genes are common among neurons in the lamprey spinal cord and brain, and zebrafish retina (Fig 2). Thus, the *Regeneration Rosetta* highlights potential

regulatory factors driving regeneration-associated gene expression among regenerative organisms.

Discussion

The *Regeneration Rosetta* interactive web app represents a rich resource of fully processed, analyzed, and queryable data from a unique study of regeneration-associated gene expression and chromatin accessibility during optic nerve regeneration in Dhara et al. (2109). The app was a crucial component for generating and interpreting results in Dhara et al. (2019), as it facilitated a deep interrogation of the data that would have otherwise only been possible with extensive bioinformatic expertise. In addition, we have illustrated the broad utility of the *Regeneration Rosetta* app through examples focusing on time-dependent regulation during regeneration for specific biological processes of interest and regenerative mechanisms that are evolutionarily conserved across species and tissue types. The *Regeneration Rosetta* app will be widely useful, both for further investigation and interpretation of the data from Dhara et al. (2019) and for hypothesis generation in other organisms. We expect these use cases of the app to inform the design of future functional studies that are crucial for translating basic biological insights into new therapeutics for optic nerve injury.

Web resources:

- The *Regeneration Rosetta* R/Shiny application is available at http://ls-shiny-prod.uwm.edu/rosetta/. A FAQ page is available directly on the app website.
- Source code for the *Regeneration Rosetta* app is available from GitHub: https://github.com/andreamrau/rosetta. The processed data used within the app are directly located in https://github.com/andreamrau/rosetta/tree/master/data; scripts used to process the raw data from Dhara et al. (2019) may be found at https://github.com/andreamrau/OpticRegen_2019.
- Archived source code at the time of publication can be found at https://doi.org/10.5281/zenodo.2658771.
- Software licence (GPL-3)

Acknowledgments

We thank the UWM High Performance Computing (HPC) Service for computing resources used in this work. We are grateful to Maria Replogle for comments on the manuscript. We wish to thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility, Madison for providing RNA and ATAC sample sequencing facilities. We gratefully acknowledge support from the Research Growth Initiative Grant - RGI (to P.L.A. and A.J.U.), University of Wisconsin-Milwaukee. A.R. was supported by the AgreenSkills+ fellowship program, which received funding from the EU's Seventh Framework Program under grant agreement number FP7-609398 (AgreenSkills+ contract).

Literature Cited

1. Crair, M. C. & Mason, C. A. Reconnecting Eye to Brain. *J. Neurosci.* **36**, 10707–10722 (2016).
2. Diekmann, H., Kalbhen, P. & Fischer, D. Characterization of optic nerve regeneration using transgenic zebrafish. *Front. Cell. Neurosci.* **9**, 118 (2015).
3. Erskine, L. & Herrera, E. Connecting the retina to the brain. *ASN Neuro* **6**, (2014).
4. Skene, J. H. Axonal growth-associated proteins. *Annu. Rev. Neurosci.* **12**, 127–156 (1989).
5. Moore, D. L. & Goldberg, J. L. Multiple transcription factor families regulate axon growth and regeneration. *Dev. Neurobiol.* **71**, 1186–1211 (2011).
6. Smith, D. S. & Skene, J. H. A transcription-dependent switch controls competence of adult neurons for distinct modes of axon growth. *J. Neurosci.* 17, 646–658 (1997).
7. Dhara, S. *et al*. Regulation of CNS regeneration-associated genes is driven by a temporally changing cast of transcription factors. DRAFT (2019).
8. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**, W202–W208 (2009).
9. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
10. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **14**, 687–690 (2017).
11. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
12. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
13. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
14. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

15. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
16. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
17. Wang, Z. *et al.* KLF6 and STAT3 co-occupy regulatory DNA and functionally synergize to promote axon growth in CNS neurons. *Sci. Rep.* **8**, 12565 (2018).
18. Smith, C. L. *et al.* Mouse Genome Database (MGD)-2018: knowledgebase for the laboratory mouse. *Nucleic Acids Res.* **46**, D836–D842 (2018).
19. Roux, P. & Fisher, R. M. Chest injuries in children: an analysis of 100 cases of blunt chest trauma from motor vehicle accidents. *J. Pediatr. Surg.* **27**, 551–555 (1992).
20. Madison, B. B. Srebp2: A master regulator of sterol and fatty acid synthesis. *J. Lipid Res.* **57**, 333–335 (2016).
21. McLeay, R. C. & Bailey, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).
22. Ye, J. & DeBose-Boyd, R. A. Regulation of cholesterol and fatty acid synthesis. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
23. Ferré, P. & Foufelle, F. SREBP-1c transcription factor and lipid homeostasis: clinical perspective. *Horm. Res.* **68**, 72–82 (2007).
24. Salles, J. *et al.* Acetyl-CoA carboxylase and SREBP expression during peripheral nervous system myelination. *Biochim. Biophys. Acta* **1631**, 229–238 (2003).
25. Dietschy, J. M. & Turley, S. D. Thematic review series: brain Lipids. Cholesterol metabolism in the central nervous system during early development and in the mature animal. *J. Lipid Res.* **45**, 1375–1397 (2004).
26. Jiang, X. & Nardelli, J. Cellular and molecular introduction to brain development. *Neurobiol. Dis.* **92**, 3–17 (2016).

27. Kaprielian, Z., Imondi, R. & Runko, E. Axon guidance at the midline of the developing CNS. *Anat. Rec.* **261**, 176–197 (2000).

28. Herman, P. E. *et al.* Highly conserved molecular pathways, including Wnt signaling, promote functional recovery from spinal cord injury in lampreys. *Sci. Rep.* **8**, 742 (2018).

29. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).