FUNCTIONAL GENOMICS REVEALS THE MECHANISMS AND EVOLUTION OF EXTREME PHYSIOLOGICAL ADAPTATIONS IN SNAKES

by

BLAIR WILLIAM PERRY

Presented to the Faculty of the Graduate School of
The University of Texas at Arlington in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON
MAY 2021

ACKNOWLEDGEMENTS

I am extremely fortunate to have worked with many incredible people during my dissertation. Thank you to the past and present members of the Castoe Lab that have been great colleagues, collaborators, and friends over the years: Drew Schield, Giulia Pasquesi, Richard Adams, Daren Card, Audra Andrew, Zach Nikolakis, Aundrea Westfall, Sid Gopalan, Kat Ivey, and Shana Pau. You have all played important roles in my development as a scientist and growth as a person, and I hope that we can continue to work together as our careers in science progress. I am also grateful to the many collaborators that have been integral to the success of my research endeavors, particularly Steven Secor, Steve Mackessy, and Jesse Meik. I am grateful for the support provided by friends, colleagues, and staff within the Department of Biology at UTA, and would like to thank all of the office staff for all of their patience and hard work. Thank you to my committee – Esther Betran, Jeff Demuth, Mark Pellegrino, and Saiful Chowdhury – for your guidance during my time at UTA and for always being willing to discuss new ideas and provide constructive feedback. I owe a particularly large thank you to all of my family and friends that have provided endless support, love, and encouragement. Thank you to my mother, Doreen, my sister, Devon, and my fiancée, Jess. Without you three, I would not have achieved what I have today, and I am so grateful to have such incredible sources of strength, positivity, and love in my life. A special thank you to my cat, Olive. I also want to thank the Miller family, Darren Baun, Uriel Avila, Justin Stettner, and Jill Castoe for their tireless support and friendship. And of course, thank you to my PhD advisor, Todd Castoe - you have been an outstanding teacher, mentor, motivator, and friend. Thank you for being so generous with your time and so invested in helping me to achieve success. I look forward to our continued collaboration and friendship long into the future.

February 21st, 2021

DEDICATION

This dissertation is dedicated to my mother, Doreen Close, my sister, Devon Perry, and my uncle, Dennis Close. The three of you taught me the value of positivity in the face of adversity, and your strength never fails to amaze and inspire. I cannot express how important your love, support, and encouragement has been throughout my life. Thank you for always believing in me.

ABSTRACT

FUNCTIONAL GENOMICS REVEALS THE MECHANISMS AND EVOLUTION OF EXTREME PHYSIOLOGICAL ADAPTATIONS IN SNAKES

Blair W. Perry, PhD

The University of Texas at Arlington, 2021

Supervising Professor: Todd A. Castoe, PhD

Understanding the processes underlying the evolution and functional basis of novel adaptations has long been a central goal of biology, and recent advances in genomic sequencing and related technologies have enabled unprecedented opportunities to investigate these processes in non-traditional model organisms. Focusing primarily on snakes and other reptiles, this dissertation uses an array of integrative functional genomics approaches to better understand the processes by which vertebrate signaling pathways and molecules can be co-opted and modified to achieve unique functions. Specifically, I identify conserved vertebrate regulatory pathways that together facilitate unique regenerative organ growth capacity in snakes, and similarly characterize a suite of signaling pathways and cellular processes that have been co-opted for the maintenance and function of a novel organ system, the snake venom gland. Further, I use an integrative functional genomics dataset to build the first comprehensive model for snake venom gene regulation and evolution. Finally, I use comparative functional genomic analyses of multiple non-mammalian vertebrate lineages to highlight unique features and evolutionary implications of microchromosomes. Collectively, the diverse work presented in this dissertation provides insight into the evolution and regulation of novel physiological adaptations in vertebrates, and showcases the value of applying integrative functional genomics approaches to study unique adaptations outside of traditional model systems.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

The impressive diversity of adaptations found across the tree of life represent numerous evolutionary solutions to a largely shared set of challenges (i.e., to survive, obtain nutrients, and reproduce). By studying the genomic basis and evolution of unique adaptations otherwise absent in traditional model systems (i.e., *Drosophila* and mouse), we can gain new perspectives into the processes and constraints of vertebrate genome biology and the evolution of novelty. Within the last decade, the increasing feasibility and affordability of genomic sequencing and related techniques has resulted in an accumulation of high-quality genomic resources (i.e., high-quality reference genome assemblies and annotations) for non-traditional model organisms. Notably, such resources unlock the ability to use a variety of functional genomics techniques to study the genomic basis of adaptation in non-traditional model organisms that have previously been limited to only model organisms. These advances therefore enable unprecedented opportunities to study the genomic basis and evolution of unique adaptations from diverse lineages and to broaden our understanding of both fundamental features and idiosyncratic complexities of genome biology and evolution.

This dissertation focuses on leveraging functional genomics techniques in non-traditional model organisms (primarily snakes and other reptiles) to better understand the molecular basis of novel physiological adaptations and the processes by which vertebrate signaling pathways and molecules can be co-opted and modified to achieve unique functions. This work spans multiple biological scales, ranging from broad regulatory pathways underlying complex physiological phenotypes (Chapters 2, 3, and 4), to specific regulatory mechanisms driving the evolution of novel genes and organ systems (Chapter 5), to unique features of non-mammalian vertebrate genome structure, function, and evolution (Chapter 6).

The first two chapters of this dissertation (Chapters 2 and 3) use comparative studies of differential gene expression and protein abundance across multiple tissues and snake lineages to identify conserved vertebrate regulatory pathways that underlie the unique and extreme capacity for regenerative organ growth in snakes. These studies culminate in a detailed model of regeneration signaling in snakes and identify particular stress response mechanisms that likely facilitate, and even enable, such high-magnitude regenerative responses. Chapter 4 uses similar methodology (i.e., analyses of differential gene expression) to dissect cellular processes and regulatory pathways associated with the maintenance and function of the snake venom gland. This study provides an important understanding of the physiological context of snake venom production that has been historically overlooked in the literature despite a long-enduring interest in snake venom systems. Based on this same system, Chapter 5 represents a more comprehensive study to identify the mechanisms of venom gene regulation, and their evolutionary origins, through the generation and analysis of a large integrative functional genomics dataset. Results of this study provide broad insight into how regulatory networks can be co-opted through diverse genomic mechanisms to produce novel polygenic traits. Lastly, Chapter 6 uses comparative studies of genome structure and organization across multiple vertebrate lineages to identify unique features of microchromosomes that may have broad and important ramifications for genome structure, function, and evolution in non-mammalian vertebrates.

This dissertation collectively provides a diverse suite of examples for how the application of functional genomics techniques previously reserved only for traditional model systems can broaden our understanding of genome biology and evolution when used to interrogate unique and extreme adaptations in non-traditional model systems. The resources generated herein and the broad implications of these studies provide key foundations for understanding vertebrate biology and functional diversity, and inspire extensions of this work in snakes and other non-traditional model systems.

Chapter 2

GROWTH AND STRESS RESPONSE MECHANISMS UNDERLYING POST-FEEDING REGENERATIVE ORGAN GROWTH IN THE BURMESE PYTHON

Audra L. Andrew[a]*, Blair W. Perry[a]*, Daren C. Card[a], Drew R. Schield[a], Robert P. Reggiero[b], Suzanne McGaugh[c], Amit Choudhary[d,e], Stephen M. Secor[f], and Todd A. Castoe[a]

*** Authors contributed equally**

[a]Department of Biology & Amphibian and Reptile Diversity Research Center, 501 S. Nedderman Drive, University of Texas at Arlington, Arlington, TX 76019 USA

[b]Department of Biology, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, United Arab Emirates

[c]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, 55108 USA

[d]Harvard Medical School, Renal Division, Brigham and Woman's Hospital, Cambridge, MA 02142 USA

[e]Center for the Science of Therapeutics, Broad Institute, Cambridge, MA, 02142 USA

[f]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, 35487, USA

Abstract

Previous studies examining post-feeding organ regeneration in the Burmese python (*Python molurus bivittatus*) have identified thousands of genes that are significantly differentially regulated during this process. However, substantial gaps remain in our understanding of coherent mechanisms and specific growth pathways that underlie these rapid and extensive shifts in organ form and function. Here we addressed these gaps by comparing gene expression in the Burmese python heart, liver, kidney, and small intestine across pre- and post-feeding time points (fasted, one day post-feeding, and four days post-feeding), and by conducting detailed analyses of molecular pathways and predictions of upstream regulatory molecules across these organ systems. Identified enriched canonical pathways and upstream regulators indicate that while downstream transcriptional responses are fairly tissue specific, a suite of core pathways and upstream regulator molecules are shared among responsive tissues. Pathways such as mTOR signaling, PPAR/LXR/RXR signaling, and NRF2-mediated oxidative stress response are significantly differentially regulated in multiple tissues, indicative of cell growth and proliferation along with coordinated cell-protective stress responses. Upstream regulatory molecule analyses identify multiple growth factors, kinase receptors, and transmembrane receptors, both within individual organs and across separate tissues. Downstream transcription factors MYC and SREBF are induced in all tissues. These results suggest that largely divergent patterns of post-feeding gene regulation across tissues are mediated by a core set of higher-level signaling molecules. Consistent enrichment of the NRF2-mediated oxidative stress response indicates this pathway may be particularly important in mediating cellular stress during such extreme regenerative growth.

Introduction

Multiple species of snakes have evolved the ability to massively downregulate metabolic and physiological functions during extended periods of fasting, including the atrophy of organs such as the heart, liver, kidney, and intestine. Upon feeding, the size and function of these organs, along with oxidative metabolism, is massively upregulated to accommodate digestion (Ott and Secor, 2007; Secor, 2008; Secor and Diamond, 1998, 2000). Of the snake species that experience these large fluctuations in physiology, the Burmese python (*Python bivittatus*) is the most well-studied (Andrew et al., 2015). Within 48 hours of feeding, Burmese pythons can undergo up to a 44-fold increase in metabolic rate and >100-fold increases in plasma triglyceride content (Secor and Diamond, 1997, 1998). Organ-specific changes also occur, including 40-100% increases in the mass of the heart, liver, pancreas, kidneys, and small intestine (Cox and Secor, 2008; Lignot et al., 2005; Secor, 2008; Secor and Diamond, 1995; Secor and White, 2010). This extreme organ regenerative growth and atrophy is unparalleled across vertebrates, and studies indicate that this organ growth is driven by multiple cellular processes, including cellular hypertrophy in the heart and mixtures of hypertrophy and hyperplasia in the kidney, liver, and small intestine (Andrew et al., 2015; Helmstetter et al., 2009; Riquelme et al., 2011; Secor and Diamond, 1998). Organ growth peaks around 1-2 days post-feeding (DPF), and by 10-14DPF, organ form and function, as well as gene expression patterns, have completely reversed back to fasted levels (Andrew et al., 2015; Castoe et al., 2013; Cox and Secor, 2008; Lignot et al., 2005; Secor, 2008; Secor and Diamond, 1995; Secor and Ott, 2007).

Previous studies have examined aspects of this post-feeding response using morphological and physiological assays (Cox and Secor, 2008; Secor, 2003, 2008; Secor and Diamond, 1998; Secor et al., 2002; Wall et al., 2011), analyses of gene expression (Andrew et al., 2015; Castoe et al., 2013), and combinations of the two (Andrew et al., 2015; Wall et al., 2011). Together, these studies have demonstrated that transcriptional responses following feeding are extremely rapid and massive, both in the magnitude of expression changes and in the number of genes with significant differential expression. Genes important in a number of developmental, metabolic, proliferative, apoptotic, and growth processes have been shown to be involved in these major shifts in organ form and function (Andrew et al., 2015; Castoe et al., 2013; Riquelme et al., 2011). Previous studies have shown that mammalian cells respond to the growth signals in post-fed python serum, which likely indicates a conserved response to core signaling molecules (Riquelme et al., 2011; Secor et al., 2014). We therefore hypothesize that a relatively small number of core molecular regulatory molecules and signaling pathways may underlie these responses. However, the identification of a core set of upstream regulatory molecules and mechanisms has been hindered by the large number of genes that are significantly differentially expressed during this response, making manual interpretation of this gene expression data difficult. Additionally, the lack of comparable replicated sampling across multiple organs has further prevented meaningful across-organ comparisons of changes in gene expression in previous studies (Castoe et al., 2013). Accordingly, major gaps remain in our understanding of the specific mechanisms and growth pathways that are responsible for driving these extreme shifts in Burmese python organ size and function, as well as how these mechanisms may vary across different organ systems.

Our previous study of the Burmese python feeding response addressed some gaps through the use of increased replicates and more frequent time point sampling for one organ, the small intestine (Andrew et al., 2015). We identified over 1,700 genes that were significantly differentially expressed during post-feeding regeneration in the small intestine with many of these genes being functionally linked to cellular processes such as WNT signaling, cell cycling, and apoptosis. This study also linked changes in gene expression with functional and phenotypic shifts by comparing RNAseq data with physiological and histological data. This detailed analysis was only conducted on the small intestine, however, and failed to address any upper-level signaling mechanisms and pathways.

Here, we leverage fully replicated organ-specific time courses detailing gene-level responses to infer canonical pathways and regulatory molecules driving post-feeding organ growth in the Burmese python. We examined gene expression across four major organ systems – the heart, liver, kidney, and small intestine. We combined increased replicated sampling with statistical inferences of pathway activation and regulatory molecule prediction to identify the mechanistic drivers of cross-tissue, post-feeding organ regeneration. Despite highly organ-specific gene expression responses associated with organ regenerative growth, we found evidence for high degrees of overlap in predicted pathways and regulatory molecules underlying these growth processes between organs. Pathways predicted to be involved in regulating this physiological response include LXR/RXR activation, PI3K/AKT, and mTOR signaling. Interestingly, we also found strong and consistent evidence for the involvement of NRF2-mediated oxidative stress response and other stress-response pathways in this extreme example of rapid organ growth. Our results suggest that post-feeding, regenerative organ growth in the Burmese python may stem from small numbers of key effector molecules mediating a core set of growth and stress-response pathways, which in turn activate diverse, tissue-specific signaling cascades.

Materials and Methods

Feeding experiments

Burmese pythons were obtained from commercial breeders. All animal care and tissue sampling was conducted using protocols approved by the University of Alabama Institutional Animal Care and Use Committee (14-06-0075). Burmese pythons were sampled at three physiological states: fasted (30 days since last meal), 1 day post-feeding (1DPF) and 4DPF, with the meal consumed equaling at least 25% of their body mass. Previous studies have shown that organ masses and functional phenotypes climax between 1 and 3 DPF (Andrew et al., 2015; Secor, 2008; Secor and Diamond, 1995; Secor and Ott, 2007) and that phenotypes begin to decline by 4DPF (Cox and Secor, 2008; Secor, 2008; Secor and Diamond, 1995, 1998). We therefore chose sampling time points here to capture gene expression patters during the period before phenotypes climax (1DPF) and early in their regression (4DPF). Snakes were humanely euthanized by severing the spinal cord immediately behind the head, and organs were immediately extracted, snap frozen in liquid nitrogen, and stored at -80°C. Between three and five biological replicates (i.e., animals) were sampled for each time point.

Transcriptome library generation

Total RNA was extracted from ~50mg of snap-frozen tissue using Trizol Reagent (Invitrogen), followed by mechanical cell disruption using a TissueLyzer for 10 minutes at 20 strokes/minute, and precipitation of RNA using isopropanol. Individual Illumina mRNAseq libraries were constructed using either the Illumina TruSeq RNAseq kit or the NEB Next RNAseq kit, both of which included poly-A selection, RNA fragmentation, cDNA synthesis, and indexed Illumina adapter ligation. Completed RNAseq libraries were quantified on a BioAnalyzer (Agilent), pooled in equal molar ratios in various multiplex arrangements, and sequenced on either an Illumina GAIIx or Illumina HiSeq2000 (Supp. Table S1).

Quantifying and visualizing gene expression

Raw demultiplexed Illumina RNAseq reads were quality filtered and trimmed with Trimmomatic v. 0.32 (Bolger et al., 2014). In instances where the same library was sequenced in multiple different runs, reads were combined and mapped for each individual and time point. Mapping of reads to the reference transcriptome of the Burmese python (Castoe et al., 2013) was conducted using BWA v. 0.6.1 (Li and Durbin, 2011) with the following parameters: mismatch penalty=2, gap open penalty=3, and alignment score minimum=20. Expression was determined using SAMtools v. 0.1.19 (Li et al., 2009) by counting the number of unique gene reads that mapped to an annotated transcript, while excluding reads that mapped to multiple positions. New RNAseq data for various time points and replicates was analyzed together with previously published data from other individuals and replicates (Andrew et al., 2015; Castoe et al., 2013). Newly-generated sequencing data were archived on the NCBI Short Read Archive (NCBI: SRP051827).

Raw expression counts were normalized using TMM normalization in edgeR (Robinson et al., 2010) and all statistical analyses of gene expression were conducted using normalized data. We identified genes that were significantly differentially expressed between time points using two approaches. First, we estimated significant changes in gene expression between pairs of time points using pairwise exact tests for the binomial distribution calculated in edgeR, integrating both common and tagwise dispersion (Robinson et al., 2010). Second, to accommodate the time-series nature of the experimental design, we also conducted step-wise regression analysis of gene expression in maSigPro (Conesa et al., 2006). Regression analysis enabled the detection of genes with significant patterns of differential expression across all three time points. Gene expression heatmaps were generated in R and clustered with the package vegan (Dixon, 2003), with gene clustering calculated using average linkage hierarchical clustering based on a Bray-Curtis dissimilarity matrix. We used the program STEM (Ernst and Bar-Joseph, 2006) to identify and visualize significant expression profiles for all genes in our RNAseq data.

Assigning homology for functional analyses

To facilitate the use of various pathway activation and regulatory molecule predictions, we annotated the full Burmese python transcript set (Castoe et al., 2013) with orthologous human gene Ensembl (Aken et al., 2016) identifiers. Reciprocal tblastx was first conducted between *Anolis carolinensis* and Burmese python, and *Anolis* gene IDs identified as orthologous to python genes were converted to human Ensembl identifiers using homology tables from Ensembl's Biomart (Cunningham et al., 2015). The same process of reciprocal best blast using tblastx was performed between Burmese python and *Gallus gallus*, followed by conversion of chicken Ensembl identifiers to human Ensembl identifiers using homology tables from Ensembl's Biomart (Cunningham et al., 2015). We also performed reciprocal best blast of the python with *Homo sapiens*. Finally, we used one-way tblastx with *anolis*, chicken, and human to annotate python genes that were not assigned an ortholog from reciprocal best blast. Using this annotation approach, we were able to assign human Ensembl IDs to 22,393 of 25,385 total python reference transcripts.

Pathway and upstream regulatory molecule analysis

To infer the involvement of upstream regulatory molecules and pathways, we performed Core Analysis in Ingenuity Pathway Analysis (IPA; Qiagen), using default parameters. IPA uses gene identifiers and the fold-change value for each differentially expressed gene to identify enrichment patterns for Canonical Pathway Analysis (CPA) and Upstream Regulatory Molecule Analysis (URMA), and to infer the activation direction (activated versus inhibited) between particular time points. These two analyses both use observed gene expression data to infer unobserved features (e.g., activation state of key signaling molecules), but differ fundamentally in how they use expression data to make inferences. CPA predicts the involvement and activation/inhibition of canonical pathways based on observed evidence from gene expression data, specifically for genes that participate as higher-level regulatory molecules within a given pathway; analysis of observed gene expression data incorporates information from the Ingenuity Knowledge Base (including genes known to be involved within a given pathway) to provide both a statistical value of enrichment and a prediction of the biological involvement for the pathway as a whole (i.e. activated or inhibited; IPA documentation, Qiagen). In contrast, URMA uses observed changes in gene expression specifically for genes at lower levels within pathways (e.g., low level effectors) to predict activation or inhibition of regulatory molecules upstream of these genes (Krämer et al., 2013). Due to differences in these approaches, together these two methods provide a well-rounded set of comparable inferences for dissecting molecular mechanisms (Fig. 1).

For IPA analyses, we used only genes identified as significant in pairwise differential expression analyses between time intervals (per organ), and we input fold changes per gene averaged across biological replicates, along with our estimate of the orthologous human Ensembl ID for each gene. Pathways important to cross-tissue physiological responses were isolated using the IPA CPA (included with Core Analysis), with a right-tailed Fisher's exact test p-value of less than 0.01. We examined only those pathways that were significant, based on a predicted activation z-score, in at least one of the four organs for at least one of the post-feeding time points. For IPA analyses, the z-score is used to determine the statistical significance of the number of activated and inhibited predictions, and the sign of the value indicates the overall activation state (i.e., positive versus negative activation). We used a p-value cutoff of 0.01 for the CPA in IPA to reduce potentially spurious inferences. Upstream regulators and hypotheses for global signaling molecules were identified using URMA in IPA, with a Fisher's exact test overlap p-value threshold of 0.05. Pathway network figures were modified manually from predicted network figures generated in IPA. For analysis of specific pathways (mTOR signaling and NRF2-mediated oxidative stress response), we also determined the number of genes involved in each pathway that were assigned python orthologs by our orthology analyses, and how many of these genes were expressed at some level in our dataset (Supp. Table S2).

Results

Trends in gene expression across organs

We used our expression data to examine the degree to which different organ systems 'turn on' upon feeding and then experience 'regression' towards pre-feeding patterns of expression at 4DPF. We found that for each organ, the majority of differentially expressed genes showed immediate up- or downregulation from fasting to 1DPF. Interestingly, each of the four organs examined appeared to experience regression towards fasting levels of expression by 4DPF to widely different extents, indicating that each organ may have its own unique temporal program of growth followed by atrophy. Across organs, the heart appeared to shift towards regression the fastest. Other organs experienced reversals of fasted to 1DPF expression shifts to varying degrees by 4DPF, ranging from the moderately paced small intestine and kidney, to the slow-paced liver (Table 1). STEM analysis further supported these temporal patterns of up-regulation and regression across organs (Supp. Fig. S1).

Regression analysis across time points, which tends to be conservative, identified hundreds of genes that were significantly differentially expressed across all three time points with 722 genes in the heart, 750 genes in the kidney, 711 genes in the liver, and 1,284 genes in the small intestine. Of the 2,922 total genes differentially expressed across all four organs, 21% are unique to the heart, 16% are unique to the kidney, 15% are unique to the liver, and 32% are unique to the small intestine (Fig. 2). Only a single gene was identified as significant in all four organs across all time points: *coagulation factor X* (F10).

To further dissect patterns of expression change following feeding, we conducted pairwise analyses of gene expression between time points for each organ. In the heart, pairwise analyses identified 436 significantly differentially expressed genes between fasted and 1DPF (208 upregulated and 228 downregulated; Table 1), and 76 genes were significantly differentially expressed between 1DPF and 4DPF (36 upregulated and 40 downregulated). In the kidney, 344 genes were significantly differentially expressed between the fasted state and 1DPF (244 upregulated and 100 downregulated), while only 8 genes were significantly differentially expressed from 1DPF to 4DPF (5 upregulated and 3 downregulated). In contrast to the heart, we found many genes (147) significantly differentially expressed between fasted and 4DPF in the kidney. In the liver, 461 genes were differentially expressed within 1DPF (335 upregulated and 126 downregulated), while only 41 genes were significantly differentially expressed from 1DPF to 4DPF (29 upregulated and 12 downregulated). With 371 genes significantly differentially expressed between fasted and 4DPF, among all four organs, the liver was the least 'reset' to the fasting condition by 4DPF. Finally, the small intestine showed higher levels of differential expression than the other three organs. Within 1DPF, 2,313 genes were significantly differentially expressed (1,271 upregulated and 1,042 downregulated). From 1DPF to 4DPF, 268 genes were upregulated and 146 genes were downregulated, and 892 genes were differentially expressed between fasted and 4DPF (Table 1).

Genes and pathways implicated in differential gene expression in individual tissues

To move beyond gene-specific responses and towards deciphering the mechanisms that may underlie growth responses across different organs, we identified pathways that were significantly activated/repressed between fasting and 1DPF (Fig. 3). We found consistent evidence that the NRF2 stress-response pathway is activated in all tissues, except in the heart, where there was insufficient data to determine the direction of activation. We also found relatively consistent evidence for activation of the related growth pathways mTOR and PI3K/AKT across organs, although this inference was most significant in the heart and small intestine. We also inferred the involvement of the related pathways: LXR/RXR, LPS/IL-1-mediated inhibition of RXR function, PPAR/RXR, and PPAR signaling in multiple organs; the direction of stimulation of these pathways was both variable across organs and inconclusive in some organs. Substantial involvement of cytoskeletal pathways, including Actin cytoskeleton signaling and Actin nucleation by ARP-WASP complex, was also inferred across organs and positive in the kidney and small intestine, yet negative or inconclusive for the heart and liver, respectively.

In addition to pathway activation/repression patterns shared across organs, a number of pathways showed substantial organ-specific directionality of response. Examples of this pattern include the growth-related AMPK signaling pathway (which was activated in the heart, repressed in the kidney and small intestine, and ambiguous in the liver), ERK5 signaling (activated in the heart and repressed in the small intestine), and Integrin signaling (stimulated in the heart and repressed in the small intestine). Lastly, a number of pathways appeared to be organ-specific, including p38 MAPK and ERK5 signaling in the heart and 14-3-3-mediated signaling in the small intestine (Fig. 3).

Upstream regulatory molecule analysis of 1DPF responses

Our inferences of upstream regulatory molecules (URMs) between the fasted and 1 DPF time points supported many of the same molecular mechanisms underlying organ growth identified via CPA, such as stress response, growth, and lipid signaling pathways. We explored URM predictions for all classes of URMs except biological drugs, chemicals, and microRNAs. We found that many predicted URMs were shared among organs, with 51 shared among all four organs. Predicted URMs also showed substantial organ-specific patterns, with a large number of URMs uniquely predicted for each organ. The heart showed the largest number of unique URM predictions (269), while only 123, 167, and 137 unique URMs were predicted in the kidney, liver, and small intestine, respectively (Fig. 4A).

To identify regulators with broadly relevant patterns across multiple organs, we focused on URMs predicted significantly in at least three organs and with moderate to high activation z-score ($z > |1.5|$) in at least one organ. A subset of the URMs meeting these criteria is shown in Fig. 4B, and the full set is shown in Additional file 1, Fig. S2. Many of these URM predictions coincided directly with predicted canonical pathways. NFE2L2 and ATF4, key regulators within the NRF2-mediated oxidative stress response pathway, were predicted to be strongly activated in the small intestine, liver, and kidney, consistent with the canonical pathway analysis predictions of activation of the overall NRF2 pathway in these three organs. We also predicted involvement of NFkB and NFkBIA, two key regulators within the NFkB signaling response pathway – this inflammatory response pathway is thought to be inhibited by activation of the NRF2-mediated oxidative stress response pathway (Cuadrado et al., 2014; Wardyn et al., 2015). NFkB was predicted to be inhibited in the liver and heart, weakly activated in the kidney, and absent in the small intestine, while NFkBIA was predicted to be inhibited in the liver, weakly activated in the heart and kidney, and again absent in the small intestine. Activation of the growth pathways mTOR and PI3K/AKT were additionally supported by activation of predicted regulators such as mTORC1 and RAF1, respectively, and the inhibition of PTEN. Lipid signaling pathways such as LXR/RXR signaling, LPS/IL-1-mediated inhibition of RXR function, PPAR/RXR, and PPAR signaling were supported by several predicted URMs such as PXR ligand, NR1H3, NR1I2, NR1I3, SREBF1, SREBF2, PPARA, PPARG, RXRA, PPARGC1A, and PPARGC1B (Fig. 4). These URMs were consistently predicted as activated in the small intestine, liver, and kidney and either absent or predicted as inhibited in the heart.

It is notable that while the inferences of activation directions of lipid signaling pathways across organs were largely ambiguous and sometimes inconsistent in our CPA (Fig. 3), the associated URMs display a consistent trend of predicted activation in the small intestine, liver, and kidney, and either predicted inhibition or absence in the heart (Fig. 4). Additionally, several URMs, particularly for the mTOR pathway, were predicted as inconsistent or even contradictory to the results of CPA or our experimental data. For example, while mTORC1 is predicted as significantly activated in the small intestine by URMA, this molecule is downregulated in our experimental data (see Discussion for details). Additionally, the mTOR protein that is involved in forming both of the main complexes of the mTOR signaling pathway (mTORC1 and mTORC2) is predicted to be strongly inhibited in the small intestine and weakly inhibited in the kidney and heart. Both of these URM predictions appear to contradict the positive activation of the mTOR signaling pathway inferred for the small intestine and heart as inferred from the CPA.

In addition to URMs involved in key predicted canonical pathways, upstream regulatory analysis predicted several other notable URMs with strong activation or informative trends across organs. Insulin and INSR were both predicted as strongly activated regulators in the kidney, liver, and small intestine, suggesting a possible role of insulin receptor signaling in facilitating this regenerative response, which is also consistent with activation of the mTOR pathway. Myc, a regulator within the ERK5 and p38 MAPK signaling pathways, was predicted as activated in all four organs, although strongest in the liver. Several regulators within the MAPK signaling pathway were also predicted in URMA, with ATF4 and ATF6 predicted as activated in the kidney, liver, and small intestine, and ERK predicted as activated in the kidney and inhibited in the heart and liver. These URMs suggest the involvement of the ERK and MAPK signaling pathways in this response, even though CPA predictions for these two pathways were not substantially strong (Figs. 3 & 4).

Detailed dissection of NRF2 and mTOR pathway responses to feeding

We were particularly interested in our findings that the NRF2 stress response and the mTOR growth pathways appear to be involved in post-feeding growth in multiple organs. To investigate these inferences further, we fully dissected evidence from our gene expression data for activation of these pathways by visualizing observed and inferred evidence for activation of these pathways in the context of IPA-generated pathway maps (Figs. 5-6; Supp. Figs. S3-S6). Specifically, we generated pathway predictions that integrate both observed shifts in gene expression from our data (from fasting - 1DPF), and estimates of activation/inhibition of molecules downstream of these observed genes that are inferred based on canonical signaling patterns in these pathways. Relevant to our power to detect pathway-wide signals of activity, we were able to associate over 70% of human genes within the mTOR and NRF2 pathways with python orthologs that were expressed at some level in our dataset (Supp. Table S2); thus, we expect that our power and degree of resolution of pathway activation for these particular pathways is quite good.

Pathways maps for mTOR responsiveness between fasted and 1DPF show both common and divergent patterns of pathway activation among organs (Figs. 5; Supp. Fig. S3). The heart (Fig. 5A) and kidney (Supp. Fig. S3) both show similar patterns of mTOR activation, including the activation of both the mTORC1 and mTORC2 complexes. Major differences in mTOR activation between these two organs includes strong evidence for downregulation of AMPK and the eIF4 complex in the heart, yet, no direct and/or clear evidence for up- or downregulation of these complexes in the kidney. In the small intestine, the mTOR pathway was inferred to be strongly downregulated, as is AKT; AMPK and the eIF4 complex showed mixed signs of activation (both positive and negative; Fig. 5B). It is also notable that different organs showed different levels of internal consistency in the integration of results with the known functionality within the mTOR pathway. For example, the heart and kidney have either zero or one pathway connection in which gene expression results contradict the direction of activation of the pathway (pink arrows in Figs. 5A and Supp. Fig. S3) – for the kidney this disagreement occurs in the relationship between RSK and inhibition of TSC1 (Supp. Fig. S3). In the small intestine, eight such disagreements occur (Fig. 5B), and most of these occur at the steps immediately above or below activation of mTORC1 and mTORC2 complexes. The liver was the only organ that contained no signal for the activation or repression of mTOR pathway (i.e., no differentially expressed genes in this pathway were observed). It should be noted that inferences for mTOR activation from CPA are at times contradictory to those identified via URMA (Figs. 3-5; Supp. Fig. S3). While predictions based on the pathway maps indicate downregulation of mTOR in the small intestine, the z-score suggests slight upregulation of this pathway during regenerative growth in this tissue. URMA predicts inhibition of the mTOR molecule in the heart, kidney, and small intestine, while mTORC1 activation is predicted in both the kidney and small intestine, and undefined in the heart. Thus, while mTOR involvement in organ regenerative growth is clear across organs, the relationships between pathway scores, molecule-level inferences, and URMs are complex.

Pathway maps for the NRF2-mediated oxidative stress response between fasted and 1DPF indicate consistent activation of this pathway in the kidney, liver, and small intestine (Fig. 6; Supp. Figs. S4-S6). In addition to predicted responses inferred from CPA (Figs. 2 & 5; Supp. Figs. S4-S6), multiple observed genes in our dataset downstream of NRF2 are upregulated in these three organs, including *thioredoxin* (TXN), *glutathione s-transferase mu 1* (GST), and *peroxiredoxin 1* (PRDX1), providing confirmatory evidence of NRF2 activation. The response of this pathway in the heart is, however, less clear (see Supp. Fig. S4). In the heart, NRF2 responses were predicted based on the observed fold-change values of only four genes, and predictions suggest inhibition of this pathway in the heart (Supp. Fig. S4) although the direction (activation versus inhibition) was not statistically significant (Fig. 3). It is also notable that we observed differences in the inferred consistency of integrated gene expression results and activation/inhibition inferences across organs (Fig. 6; Supp. Figs. S4-S6): in the heart, only two inconsistencies are observed while the kidney, liver, and intestine have one, two, or four inconsistencies, respectively. Inferences from URMA for the activation of NRF2 are highly consistent with activation inferences from CPA, including significant URM activation predicted for NFE2L1 in the liver and intestine and significant activation of NFE2L2 in kidney, liver, and small intestine (Fig. 4). In contrast, upstream regulators of this pathway were not predicted to be significantly activated or inhibited in the heart, inconsistent with the predictions given in the pathway figure (Figs. 4; Supp. Fig. S4).

Expression response between 1 and 4 DPF

In comparison to expression between fasting and 1DPF, the IPA analyses conducted on genes differentially expressed between 1DPF and 4DPF across organs predicted a substantially smaller number of pathways as significantly enriched, the majority of which were predicted with ambiguous directions of activation. This is likely due to the substantially smaller number of significantly differentially expressed genes identified in all organs between 1DPF and 4DPF, which is expected because 4DPF represents a sampling time intermediate between the peaking of organ growth and the regression of these phenotypes. This time interval (1DPF - 4DPF) aimed to capture the early stages of organs shifting expression towards organ atrophy and towards a reversion to the fasted state, and we expected to observe partial reversals in pathways predicted to be active between fasted and 1DPF, and perhaps additional new pathways involved in apoptosis and atrophy. However, we found few consistent or clear patterns of interpretable pathway involvement between the 1DPF and 4DPF time points (Supp. Fig. S7). Pathways predicted for this time interval include various pathways related to biosynthesis and stress response, such as unfolded protein response. We also inferred inconsistent involvement of these pathways across organs, and none were predicted with a direction of activation (Supp. Fig. S7). Only one pathway, mitotic roles of polo-like kinase, was predicted as significant and with a direction of activation between 1DPF and 4DPF, and was predicted only in the small intestine. While we did infer a single lipid signaling pathway that also was indicated by CPA predictions from the fasted to 1DPF interval (LPS/IL-1 mediated inhibition of RXR function), the lack of predicted directions of activation and unclear involvement across organs prevents informative interpretation of the activity of this pathway between 1DPF and 4DPF. Collectively, these results suggest that the 4DPF time point may not be sufficient to capture shifts in gene expression that elucidate the mechanisms involved in the early stages of regression of organ phenotypes.

Discussion

A detailed understanding of the molecular mechanisms capable of driving regenerative growth in vertebrates may provide important insights into the treatment of diverse human diseases. Because traditional vertebrate model systems offer limited insight into natural organ regenerative processes, non-traditional model systems, including snakes in general and Burmese pythons in particular, hold great potential for providing unique insights into vertebrate regenerative organ growth processes. In this study we have found that multiple integrated growth pathways, in addition to multiple stress-response pathways, appear to underlie the coordinated organ regenerative process in Burmese pythons upon feeding. Despite distinct patterns of gene expression associated with growth for each organ, pathway and upstream regulatory molecule analyses reveal substantial similarities in pathways associated with post-feeding, extreme-growth responses across multiple organs. Specifically, we found evidence for a consistent interactive role of three major types of pathways underlying growth responses in python organs following feeding, including the related growth pathways mTOR and PI3K/AKT, lipid-signaling pathways such as PPAR and LXR/RXR, and stress-response/cell-protective pathways including NRF2.

mTOR and other growth pathways underlying organ growth

Across the four organs examined, we found evidence for the involvement of the mTOR signaling pathway as a key integrator of growth signals underlying post-feeding regenerative organ growth. This pathway integrates processes for the use of energy and nutrients to regulate growth and homeostasis (Laplante and Sabatini, 2012). mTOR interacts with multiple other pathways, including PI3K/AKT, several lipid metabolism and signaling pathways (Laplante and Sabatini, 2009, 2012), and the NRF2-mediated oxidative stress response (Lee et al., 2012; Okouchi et al., 2006) – all of which are also active in multiple organs during growth (Figs. 3-5). mTOR complex 1 (mTORC1) is the most well-characterized of the two mTOR complexes and integrates signaling from growth factors, energy status, oxygen, and amino acids to promote cell growth when activated (Laplante and Sabatini, 2009). The TSC1/2 complex transmits upstream signals from growth factor and insulin signaling to modulate the activity of mTORC1 and its interaction with other pathways including PI3K/AKT (Laplante and Sabatini, 2009, 2012; LoPiccolo et al., 2008). The effector kinases of these external pathways inactivate the TSC complex through phosphorylation, thus, indirectly activating mTORC1 (Laplante and Sabatini, 2009, 2012). AKT can also directly activate mTORC1 through phosphorylation of an mTORC1 inhibitor. In a low energy state, AMPK inhibits mTORC1 by phosphorylating regulatory associated protein of mTORC1 (RAPTOR) (Laplante and Sabatini, 2009, 2012). mTORC2 signaling is less well-understood, but is known to respond to growth factors through PI3K signaling (Laplante and Sabatini, 2012).

CPA of gene expression in the first 24 hours after feeding indicate that involvement of the mTOR signaling pathway is significant in the small intestine (predicted activation), but insignificant in both the heart (predicted activation) and kidney (activation state undetermined). The liver lacked evidence of involvement of the mTOR signaling pathway from CPA (Figs. 3-4). In URM analysis, the mTOR molecule itself was predicted to be downregulated in the heart, liver, and intestine with no presence in the kidney, which contrasts our CPA results (Figs. 3-4). However, URMA-predicted activation of the mTORC1 complex is supported in both the kidney and small intestine with undefined involvement in the heart, and the liver shows no signal for mTORC1 (Fig. 4). Interestingly, CPA indicate mTORC1 is downregulated in the small intestine at 0-1DPF (Fig. 6), yet this downregulated state of mTORC1 is based only on the downregulation of a single gene, G protein subunit beta 1 like (GNB1L), which IPA identifies as a subunit of the mTORC1 complex. In contrast, AMPK signaling is predicted to be downregulated in the kidney and small intestine, indicative of elevated ATP levels and active mTORC1 (Laplante and Sabatini, 2009, 2012) (Fig.3). It is notable that nearly all genes in the mTOR pathway were associated with python orthologs that were observed as expressed across our dataset (see Supp. Table S2), which suggests that our inferences of non-responsive genes within the mTOR pathway are biologically meaningful (e.g., true negatives), rather than representative of a lack of data. Thus, mTOR signaling in python tissues during regenerative organ growth may include non-canonical features compared to typical models of mTOR signaling that account for the partial responsiveness of genes and targets inferred from our CPA.

Our results identify mTOR as a central regulator and integrator of a number of diverse growth signals that drive post-feeding regenerative organ growth in Burmese pythons. Insulin signaling represents a key-regulating factor of the mTOR pathway (Laplante and Sabatini, 2009), and we found multiple lines of evidence indicating roles of insulin signaling in post-feeding growth responses. Specifically, 0-1DPF URMA inferred the activation of INSR and insulin, and the inhibition of INSIG1 and INSIG2, in the kidney, small intestine, and liver, and the inverse of these activation patterns in the heart. INSIG1 and INSIG2 are negative regulators of SCAP (Espenshade, 2006; Yang et al., 2002), which in turn regulates SREBP activity. Consistent with inferences of inhibition of INSIG1-2, URMA predicted the upregulation of SREBF1 and SREBF2, which provide evidence of an increase in sterol-regulatory element activity coincident with organ growth (Espenshade, 2006; Yang et al., 2002) (Fig. 4). In addition to the interaction of insulin signaling and mTOR activity, we also found multiple lines of evidence for PI3K/AKT signaling that would interact with mTOR. Our URMA indicates significant downregulation of PTEN, an upstream regulator of the PI3K/AKT pathway, across all four organs, and CPA predicts activation of the PI3K/AKT signaling pathway in the small intestine and liver.

Evidence from previous studies also support the role of mTOR, PI3K/AKT, and AMPK signaling mechanisms in python post-feeding growth, at least in the heart. Western blots of python cardiac tissue post-feeding support the inference of early activation of mTOR and PI3K/AKT pathways by demonstrating that phosphorylated AKT and MTOR proteins increase significantly in abundance between 12 and 24 hours post-feeding (Riquelme et al., 2011). These western blots also demonstrated phosphorylated AMPK protein was upregulated within 24 hours post-feeding, but lagging temporally behind the peak in phosphorylated MTOR and AKT (Riquelme et al., 2011)., consistent with the antagonistic relationship between AMPK and MTOR/AKT (Laplante and Sabatini, 2012). These independent lines of evidence for the roles of mTOR, PI3K/AKT, and AMPK signaling in python post-feeding organ growth confirm our inferences of the central roles of these pathways, and support the power of pathway and URM inferences for inferring signaling mechanisms.

MAPK and related pathways also appear to be prominently involved in organ growth responses post-feeding, which is sensible given their known interactions with multiple growth pathways, including PI3K/AKT signaling and mTOR (Aksamitiene et al., 2012; Pappalardo et al., 2016; Wong et al., 2016). Our data reveal the involvement of MAPK signaling most clearly in the heart, with significant enrichment and predicted inhibition of p38 MAPK signaling and significant activation of ERK5 signaling (Fig. 3). ERK5 is a member of the Mitogen-activated protein kinases (MAPKs) that is crucial to cell proliferation and activated in response to growth factors and oxidative stress (Gomez et al., 2016; Kato et al., 2000). MYC is a downstream transcription factor regulated by the MAPK pathway and ERK5 specifically (English et al., 1998; Wang and Tournier, 2006), and an essential regulator of development and cell proliferation (Davis et al., 1993; Mateyak et al., 1997; Shao et al., 2013). Our URMA predict significant activation of MYC in all four organs, indicating a broad role of active MAPK signaling in post-feeding organ growth in the python.

NRF2 – protective function and interaction with growth pathways

One of the strongest and most consistent signals in the canonical pathway and upstream regulatory molecule analyses was the involvement of the NRF2-mediated oxidative stress response pathway. Commonly associated with anti-aging and longevity (Lewis et al., 2010, 2015; Sykiotis and Bohmann, 2008), injury repair, and mitigation of inflammation (Reddy et al., 2009), evidence for the central involvement of the NRF2-mediated oxidative stress response pathway in the small intestine, liver, and kidney begs the question of whether there is an important yet largely unappreciated role for stress-response signaling pathways in growth responses, and regenerative organ growth in particular.

The NRF2 pathway was significantly upregulated in small intestine, kidney, and liver within the first day following feeding (Fig. 3), and the NRF2 transcription factor (NFE2L2) was one of the most significant and highest in magnitude URMs predicted in these three organs (p-values < $1.55e^{-10}$, z-scores > 3.0) (Fig. 4). The 24 hour period following feeding in Burmese pythons involves unparalleled rates and magnitudes of organ growth, and also includes massive upregulation of metabolism – up to 44-fold increases in aerobic metabolism, which is among the highest fluctuation known for any vertebrate (Secor and Diamond, 1998)=. It is, therefore, sensible that activation of NRF2 is related to these major shifts in oxidative metabolism, and associated generation of reactive oxygen species (Secor, 2008; Secor and Diamond, 1995, 1997; Secor and Ott, 2007). An open question, however, is what broader role the activation of NRF2 may play in facilitating the extraordinary growth responses associated with feeding in pythons. For example, post-fed Burmese python blood plasma has been shown to convey resistance to apoptosis to mammalian cells, even with exposure to high fatty acid concentrations that would otherwise cause cell death (Riquelme et al., 2011; Secor et al., 2014); such cell-protective qualities may be related to signals that activate NRF2 and/or other stress-response pathways. Interestingly, in addition to cell-protective roles of NRF2, this pathway also contains multiple points of integration with various growth pathways, including those activated in python organ regenerative growth.

The NRF2-mediated oxidative stress response pathway interacts with multiple pathways predicted in our canonical pathway analysis (Beyer and Werner, 2008; Braun et al., 2004; Hayes and Ashford, 2012; Kannan et al., 2013; Kensler et al., 2007; Shibata et al., 2010) (Figs. 3-4). The PI3K/AKT signaling pathway, predicted to be upregulated upon feeding in both the liver and small intestine, is essential for regulating the antioxidant functions of NRF2, and studies have shown that inhibition of this signaling pathway leads to attenuation of NRF2 activities (Papaiahgari et al., 2006; Wang et al., 2008). This interaction is evident when examining the role of NRF2 in the proliferation of cancer cells. Studies have shown that NRF2 is able to redirect glucose and glutamine into anabolic pathways through activation of PI3K/AKT signaling (Mitsuishi et al., 2012). The activated PI3K/AKT pathway leads to greater accumulation of NRF2 in the nucleus, which allows NRF2 to enhance metabolic activities as well as promote cell proliferation and cytoprotection (Mitsuishi et al., 2012). The PI3K/AKT signaling pathway activates mTOR activity in response to growth factors, and this and previous studies (Riquelme et al., 2011) have shown that PI3K/AKT and mTOR signaling are key growth pathways underlying organ regenerative growth in the Burmese python. Therefore, there appears to be strong and coordinated links between growth signaling (via PI3k/AKT and mTOR) and stress response signaling via NRF2 underlying organ growth in pythons following feeding. Like mTOR, a large majority of genes in the NRF2 pathway were associated with python orthologs and were observed as expressed across our dataset (Supp. Table S2), which indicates that our inferences of non-responsive genes within the NRF2 pathway are likely true negatives, rather than artifacts due to a lack of ortholog identification in the python. Accordingly, predicted but unobserved expression responses in the NRF2 pathway in pythons suggest that the absence of expected responses may represent novel or non-canonical aspects of python biology or of the organ regeneration response in pythons.

In addition to NRF2-mediated oxidative stress response, evidence for the involvement of other stress response signaling mechanisms in python post-feeding organ growth was also observed. EIF2 signaling, important in translational control and responsiveness to conditions of environmental stress (Boyce et al., 2005; Wek et al., 2006), is strongly downregulated in the intestine, yet, absent in the other three organs (Fig. 3). Acute phase response signaling, which is involved in restoring homeostasis following inflammation or injury (Moshage, 1997), is predicted to be strongly downregulated in the liver and moderately upregulated (but non-significant in the heart; Fig. 3). The precise roles of these additional stress response mechanisms in regenerative organ growth in the python remains an open question, although there is strong and consistent signal for the involvement of multiple stress response pathways overall in python post-feeding organ growth.

Role of lipid signaling in driving growth

Previous studies have shown evidence that molecules responsible for triggering python post-feeding organ growth circulate in the blood of the Burmese python (Riquelme et al., 2011; Secor et al., 2001). Riquelme et al. demonstrated that post-feeding python plasma was capable of inducing cardiomyocyte growth in pythons and mice, and that fasted python plasma supplemented with three particular fatty acids successfully stimulated cardiomyocyte growth in mice (Riquelme et al., 2011). Because these fatty acids only facilitated a growth response in the presence of fasted Burmese python serum, it is likely that python plasma contains additional factors required for successful post-feeding regenerative growth and that fatty acids are only partially responsible for stimulating growth responses. In the heart, we found significant enrichment and predicted activation for the LXR/RXR activation pathway as well as predicted activation of this pathway (although insignificant enrichment with $P > 0.01$) in the small intestine (Fig. 3). LXR is a potent activator of the SREBP-1c gene (Raghow et al., 2008), and our data predict clear and significant activation of both SREBF1 and SREBF2 upon feeding in the kidney, liver, and small intestine with significant down-regulation and undefined direction for SREBF1 and SREBF2 in the heart, respectively (Fig. 4). When activated, these proteins directly enhance genes important for the uptake and synthesis of various lipids.

SCAP, important for the activation of these SREB molecules, is also predicted to be strongly activated in the kidney, liver, and small intestine (Fig. 4) (Espenshade, 2006; Matsuda et al., 2001; Yang et al., 2002).

We also examined PPAR signaling as a potential pathway for lipid signaling during this regenerative growth, given the central role of PPAR in mediating fatty acid signaling as well as its effects on gene expression (Schoonjans et al., 1996). PPAR has also been identified as an important regulator of cell survival during wound repair and regeneration (Gurtner et al., 2008). Although CPA did not detect significant PPAR signaling activation, URMA significantly predicted PPARA, PPARG, PPARGC1A, and PPARGC1b involvement across organs, typically inhibited in the heart and activated in the other three organs in 0-1DPF comparisons (Fig. 4). Given the variations in pathway and URM inferences between the heart and the other three organs, the question of whether fatty acids also play a similar stimulatory role in regenerative growth in the small intestine, liver, and kidney as they do in the heart remains. Our results do, however, argue for a poorly understood yet central role of lipid-signaling in these growth responses, and suggest that the unusually strong bioactivity of fatty acids may elicit growth through conserved canonical pathway signaling mechanisms.

Early phases of organ regression following digestion

Physiological studies have shown that python post-feeding organ growth peaks between 1DPF and 3DPF (Andrew et al., 2015; Secor, 2008; Secor and Diamond, 1995, 1998; Secor and Ott, 2007) and that phenotypes begin to decline by 4DPF (Cox and Secor, 2008; Secor, 2008; Secor and Diamond, 1995, 1998). Thus, as post-feeding growth phenotypes reverse from 1DPF to 4DPF, we expected to observe shifts towards the fasted state, such as the reversal or inhibition of growth-associated pathways. Relative to comparisons between fasting and 1DPF, comparisons between 1DPF and 4DPF yielded nearly an order of magnitude fewer significantly differentially expressed genes (Table 1). Accordingly, expression heatmaps (Fig. 2) and expression profile summaries (Supp. Fig. S1) show that expression profiles of many genes at 4DPF tend to remain elevated (i.e., similar to levels at 1DPF), or exist at intermediate levels (between fasted and 1DPF levels of expression). We did not observe any particularly informative trends in canonical pathways and upstream regulator molecule predictions (Supp. Fig. S7) associated with shifts in gene expression from 1DPF to 4DPF, and this result is not surprising given the relatively small number of genes that significantly change between these time points. Among the predicted pathways were several that are related to stress response and biosynthesis (Supp. Fig. S7), although a lack of predicted direction of activation prevents detailed interpretation of the involvement of nearly all pathways predicted between 1DPF and 4DPF. The only pathway predicted as significant and with a direction of activation between 1DPF and 4DPF was the mitotic roles of polo-like kinase pathway, which was activated in the small intestine (Supp. Fig. S7). It therefore remains an open question whether atrophy and other processes involved in reverting to the fasting state are controlled actively (via a new signal that stimulates the apoptotic and atrophy processes), passively (the signal(s) that stimulates the initial cascade of responses fades or stops), or some combination of the two mechanisms. Collectively, our results suggest that comparisons between the 1DPF to 4DPF time points may not be sufficient to predict the physiological mechanisms involved in phenotypic regression with adequate power. Further experiments, possibly with multiple later-stage time point sampling, may be required to address outstanding questions about how these growth phenotypes are reversed.

Comparisons of python organ regeneration to other regenerative model systems

Organ regeneration in snakes represents an extreme and unique phenotype among vertebrates. However, other examples of regenerative growth do exist among vertebrates, such as limb regeneration in salamanders (Brockes and Kumar, 2002), fin regeneration in fish (Poss et al., 2003), and regenerative heart growth in zebrafish (Jopling et al., 2010; Kikuchi, 2014) and prenatal mammals (Porrello et al., 2011). This raises the question of whether or not these regenerative responses share common mechanisms, and as we continue to better understand the mechanisms driving regenerative growth in snakes, such key comparisons can begin to be made. While none of these other vertebrate regenerative growth systems directly parallel regenerative organ growth in snakes, regeneration of heart tissue in zebrafish is the most analogous comparison, as it occurs in adult organisms and represents regenerative growth of organ tissue specifically. Following injury or amputation of cardiac tissue, zebrafish hearts grow primarily by dedifferentiation and subsequent proliferation of cardiomyocytes (Jopling et al., 2010). Conversely, python hearts grow only by hypertrophy (Riquelme et al., 2011; Secor and Diamond, 1998), and therefore may be driven by largely different regenerative mechanisms. The python small intestine, liver, and kidney, however, do grow via by hypertrophy and hyperplasia (Andrew et al., 2015; Helmstetter et al., 2009; Riquelme et al., 2011; Secor and Diamond, 1998); while they represent different organ systems than the zebrafish heart, they may be driven by similar pathways that regulate cell proliferation in general. Indeed, there are parallels between zebrafish and python responses in the shared involvement of p38 MAPK signaling, a negative regulator of cardiomyocyte proliferation in zebrafish (Kikuchi, 2014) that we infer to be inhibited in the Burmese python heart between fasting and 1DPF (Fig. 3). Additionally the mitotic roles of polo-like kinase pathway, which was the only pathway we predicted as significant and with a direction of activation between 1DPF and 4DPF (activated in the small intestine; Supp. Fig. S7) is also involved in zebrafish regenerative heart growth. Cell-cycle regulation by polo-like kinase 1 is an important component of cardiomyocyte proliferation in zebrafish (Jopling et al., 2010), and therefore may be playing a similar role in the python small intestine, although it is notable that it was not predicted as significant between fasting and 1DPF, when growth is presumably greatest in this organ (Andrew et al., 2015; Secor and Diamond, 1998). Other pathways involved in zebrafish regenerative growth, such as IGF signaling, FGF signaling, HIPPO signaling, and TGF-Beta signaling (Kikuchi, 2014), were not inferred as significant based on canonical pathway analyses of either post-feeding time interval in our study of the Burmese python. TGFB1 and IGF1 growth factors were, however, inferred in our URMA analysis of the fasting to 1DPF interval (Supp. Fig. S2), suggesting that there may still be some involvement of these growth factors in the regulation of regenerative growth in the Burmese python. A key conclusion based on our study is that, to our knowledge, mTOR signaling and NRF2-mediated oxidative stress response pathways have not been implicated in zebrafish regenerative growth. Thus, regenerative organ growth in the Burmese python appears to remain quite unique among vertebrates, both in the nature of the phenotype, and now in the molecular mechanisms underlying growth.

Conclusions

Multiple coordinated growth pathways appear to play an important role in facilitating regenerative organ growth in multiple tissues of the Burmese python, and the overlap of pathways across organs suggests common signaling molecules may drive this response – consistent with evidence that common factors circulating in the plasma of pythons are capable of eliciting growth (Riquelme et al., 2011; Secor et al., 2001). Our analyses provide strong evidence for the involvement of particular growth and stress response pathways in post-feeding organ growth responses in multiple organs, although it is notable that our inferences of the activation versus inhibition of mechanisms was not always consistent across analyses (e.g., CPA versus URMA). As discussed above, such conflicting inferences could be due to the fundamental differences in CPA and URMA (e.g., Fig. 1), in that they are integrating very different sources of evidence, coupled with the possibility that the continuous nature of this response may survey various mechanisms during an inflection point of activity that can confound inferences of directionality. However, contradictory inferences of mechanistic activation may also suggest that some of these core signaling pathways function differentially in snakes, or that some molecules or pathways are signaling via non-canonical mechanisms. Experiments have demonstrated that exposure to Burmese python 2DPF blood serum elicits significant growth of rat cardiomyocytes (Riquelme et al., 2011), as well as increases in size and insulin production of human pancreatic beta cells (Secor et al., 2014). These findings suggest that even if regenerative organ growth in snakes is achieved in part by non-canonical pathway or regulator activity, core aspects of signaling underlying organ growth in pythons is conserved across vertebrates. Among the most intriguing results of this study is the consistent predicted activation of the NRF2-mediated oxidative stress response pathway, and NRF2-related signaling molecules, during regenerative organ growth. The integration of NRF2 signaling with other growth pathways, including mTOR, provide an exciting and novel mechanistic hypothesis for how NRF2 and other stress-response pathways may play an important yet largely unappreciated role in regenerative growth responses in vertebrates.

Abbreviations

DPF: days post-feeding; **IPA**: Ingenuity Pathway Analysis; **CPA**: Canonical pathway analysis; **URMA**: Upstream regulatory analysis

Acknowledgments

We thank J. Castoe, R. Wostl, and D. Polanco for assistance with laboratory aspects of this work. Support was provided from startup funds from the University of Texas at Arlington to TAC, and grants from the University of Texas at Arlington Phi Sigma Society to ALA. Additional funding was provided by the National Science Foundation (IOB-0466139) to SMS.

Data Availability

New sequencing data are archived on the NCBI Short Read Archive (NCBI: SRP051827). Previously generated data are accessioned at NCBI: SRP051827.

Figures

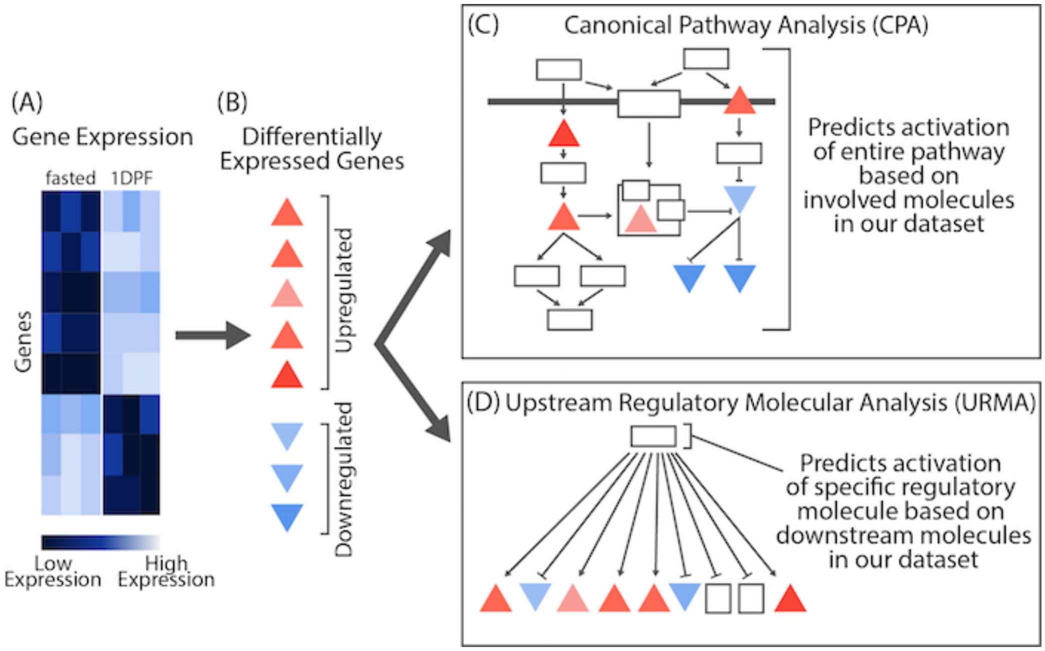

(A)
Gene Expression
fasted
1DPF
Genes
Low Expression
High Expression
(B)
Differentially Expressed Genes
Upregulated
Downregulated
(C) Canonical Pathway Analysis (CPA)
Predicts activation of entire pathway based on involved molecules in our dataset
(D) Upstream Regulatory Molecular Analysis (URMA)
Predicts activation of specific regulatory molecule based on downstream molecules in our dataset

Figure 1. Conceptual overview of differences between Canonical Pathway Analysis (CPA) and Upstream Regulatory Molecule Analysis (URMA). Pairwise analyses on experimental gene expression data (A) identify significantly upregulated and downregulated genes (B). Significantly differentially expressed genes are then analyzed in two distinct IPA analyses (CPA and URMA) (C) Canonical Pathway Analysis predicts pathway activation based on overlap of gene expression data with molecules within the pathway. (D) Upstream Regulatory Molecule Analysis predicts activation of specific regulatory molecules based on downstream molecules in our gene expression dataset.

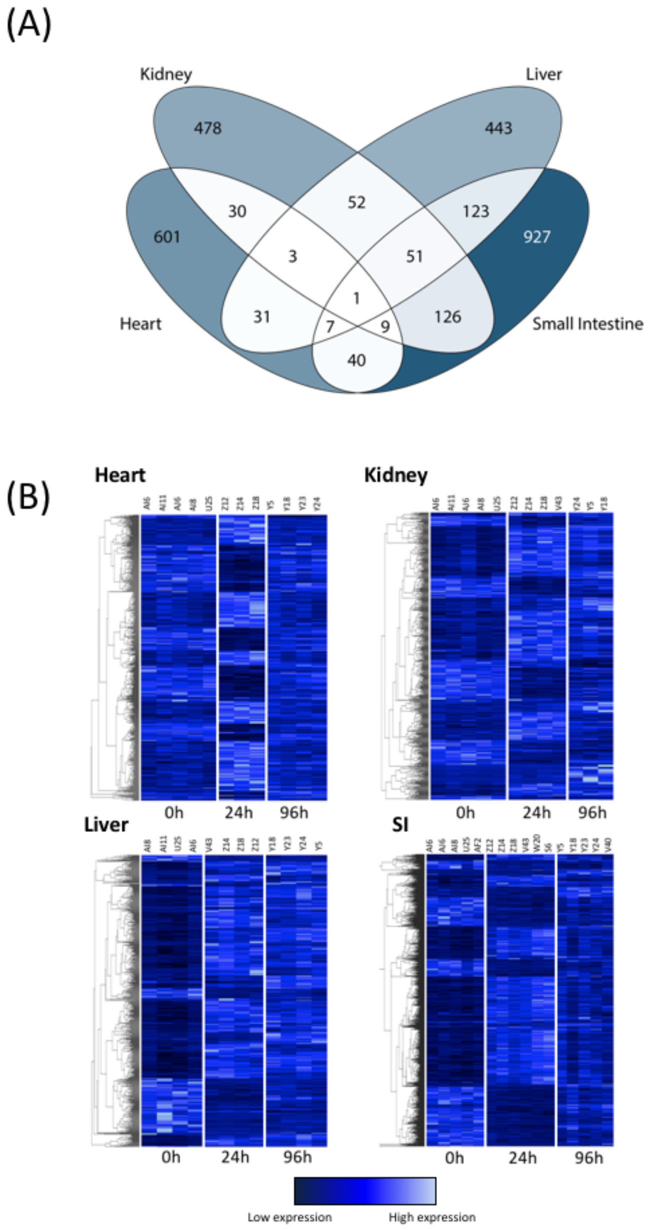
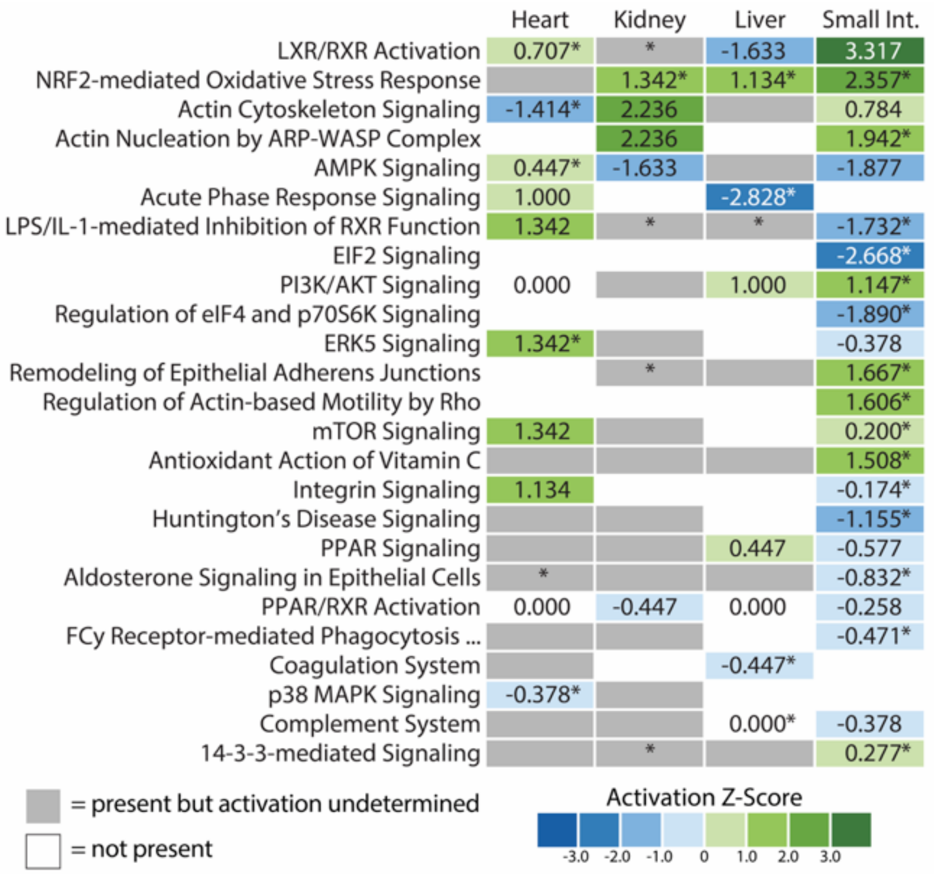


Figure 2. Summary of significantly differentially expressed genes for all four organs identified via regression analysis. (**A**) Venn diagram depicting the numbers of genes significantly differentially expressed across time points. Darker colors indicate a large number of genes and lighter colors indicate a smaller number of genes. (**B**) Heatmaps depicting all significantly differentially expressed genes across all time points in each organ. 722 genes were significantly differentially expressed in the heart. There were 750 genes significantly differentially expressed in the kidney. 711 genes were significantly differentially expressed in the liver and 1,284 genes showed significant differential expression in the small intestine.



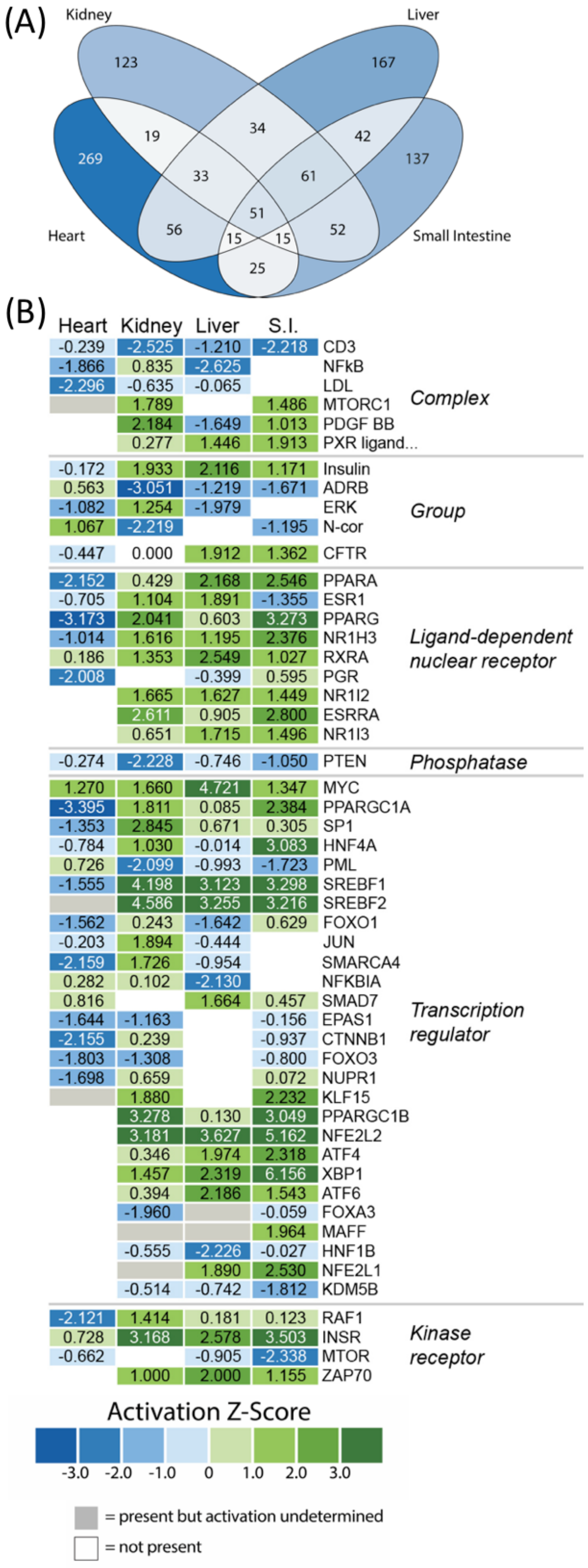
	Heart	Kidney	Liver	Small Int.
LXR/RXR Activation	0.707*	▨ *	-1.633	3.317
NRF2-mediated Oxidative Stress Response	▨	1.342*	1.134*	2.357*
Actin Cytoskeleton Signaling	-1.414*	2.236	▨	0.784
Actin Nucleation by ARP-WASP Complex		2.236		1.942*
AMPK Signaling	0.447*	-1.633	▨	-1.877
Acute Phase Response Signaling	1.000		-2.828*	
LPS/IL-1-mediated Inhibition of RXR Function	1.342	▨ *	▨ *	-1.732*
EIF2 Signaling				-2.668*
PI3K/AKT Signaling	0.000	▨	1.000	1.147*
Regulation of eIF4 and p70S6K Signaling				-1.890*
ERK5 Signaling	1.342*	▨		-0.378
Remodeling of Epithelial Adherens Junctions		▨ *	▨	1.667*
Regulation of Actin-based Motility by Rho				1.606*
mTOR Signaling	1.342	▨		0.200*
Antioxidant Action of Vitamin C	▨	▨	▨	1.508*
Integrin Signaling	1.134			-0.174*
Huntington's Disease Signaling	▨	▨		-1.155*
PPAR Signaling	▨	▨	0.447	-0.577
Aldosterone Signaling in Epithelial Cells	▨ *	▨	▨	-0.832*
PPAR/RXR Activation	0.000	-0.447	0.000	-0.258
FCy Receptor-mediated Phagocytosis ...	▨	▨		-0.471*
Coagulation System	▨		-0.447*	
p38 MAPK Signaling	-0.378*	▨		
Complement System	▨	▨	0.000*	-0.378
14-3-3-mediated Signaling	▨	▨ *	▨	0.277*

▨ = present but activation undetermined

□ = not present

Figure 3. Canonical pathways predicted to be activated or inhibited from gene expression data. Each pathway shown is significantly enriched for our genes with a Fisher's Exact test p-value less than 0.01 (depicted with an asterisk). Pathways were shown only if they met our criteria for significance and had a predicted activation state in at least one organ. Z-scores of 0.000 indicate pathway predictions that lack a bias in the direction of gene regulation observed in our dataset. PPAR signaling (P<0.05) was also included.

Figure 4. Predicted upstream regulators from IPA analysis of gene expression changes from fasted to 1DPF. (**A**) Venn diagram of all upstream regulatory molecules analyzed. (**B**) Heatmap of predicted activation z-scores for selected classes of upstream regulatory molecules. Green indicates predicted activation, blue indicates predicted inhibition, white indicates the regulator is not predicted to function in that organ, and grey indicates that the upstream regulator is predicted to have significant involvement but the activation state cannot be determined based on the gene expression data. Regulators shown in this heatmap were filtered by three conditions: 1) were present in at least three of the four organs, 2) are significantly predicted (p-value < 0.05), and 3) have activation z-scores greater than |1.5| in at least one organ. Biological drug, chemical, and microRNA categories were excluded from URM analyses.

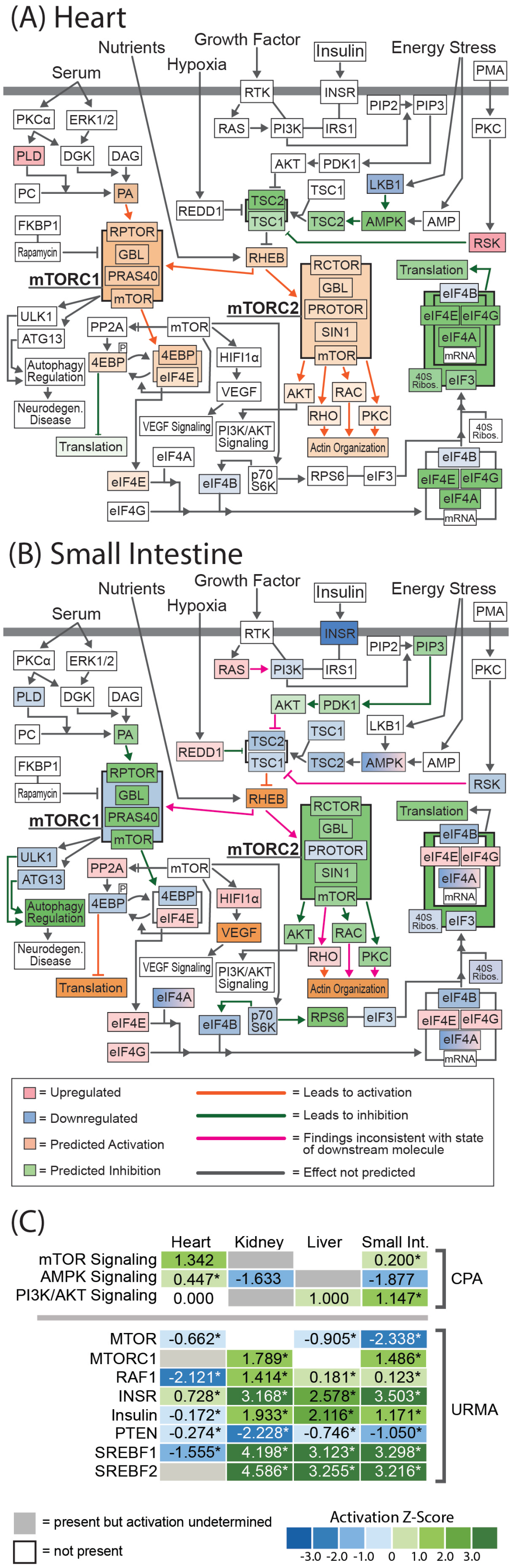


Heart	Kidney	Liver	S.I.		
-0.239	-2.525	-1.210	-2.218	CD3	
-1.866	0.835	-2.625		NFkB	
-2.296	-0.635	-0.065		LDL	
▨	1.789		1.486	MTORC1	*Complex*
	2.184	-1.649	1.013	PDGF BB	
	0.277	1.446	1.913	PXR ligand...	
-0.172	1.933	2.116	1.171	Insulin	
0.563	-3.051	-1.219	-1.671	ADRB	
-1.082	1.254	-1.979		ERK	*Group*
1.067	-2.219		-1.195	N-cor	
-0.447	0.000	1.912	1.362	CFTR	
-2.152	0.429	2.168	2.546	PPARA	
-0.705	1.104	1.891	-1.355	ESR1	
-3.173	2.041	0.603	3.273	PPARG	
-1.014	1.616	1.195	2.376	NR1H3	*Ligand-dependent*
0.186	1.353	2.549	1.027	RXRA	*nuclear receptor*
-2.008		-0.399	0.595	PGR	
	1.665	1.627	1.449	NR1I2	
	2.611	0.905	2.800	ESRRA	
	0.651	1.715	1.496	NR1I3	
-0.274	-2.228	-0.746	-1.050	PTEN	*Phosphatase*
1.270	1.660	4.721	1.347	MYC	
-3.395	1.811	0.085	2.384	PPARGC1A	
-1.353	2.845	0.671	0.305	SP1	
-0.784	1.030	-0.014	3.083	HNF4A	
0.726	-2.099	-0.993	-1.723	PML	
-1.555	4.198	3.123	3.298	SREBF1	
▨	4.586	3.255	3.216	SREBF2	
-1.562	0.243	-1.642	0.629	FOXO1	
-0.203	1.894	-0.444		JUN	
-2.159	1.726	-0.954		SMARCA4	
0.282	0.102	-2.130		NFKBIA	
0.816		1.664	0.457	SMAD7	*Transcription*
-1.644	-1.163		-0.156	EPAS1	*regulator*
-2.155	0.239		-0.937	CTNNB1	
-1.803	-1.308		-0.800	FOXO3	
-1.698	0.659		0.072	NUPR1	
▨	1.880		2.232	KLF15	
	3.278	0.130	3.049	PPARGC1B	
	3.181	3.627	5.162	NFE2L2	
	0.346	1.974	2.318	ATF4	
	1.457	2.319	6.156	XBP1	
	0.394	2.186	1.543	ATF6	
	-1.960	▨	-0.059	FOXA3	
	▨	▨	1.964	MAFF	
	-0.555	-2.226	-0.027	HNF1B	
	▨	1.890	2.530	NFE2L1	
	-0.514	-0.742	-1.812	KDM5B	
-2.121	1.414	0.181	0.123	RAF1	
0.728	3.168	2.578	3.503	INSR	*Kinase*
-0.662		-0.905	-2.338	MTOR	*receptor*
	1.000	2.000	1.155	ZAP70	

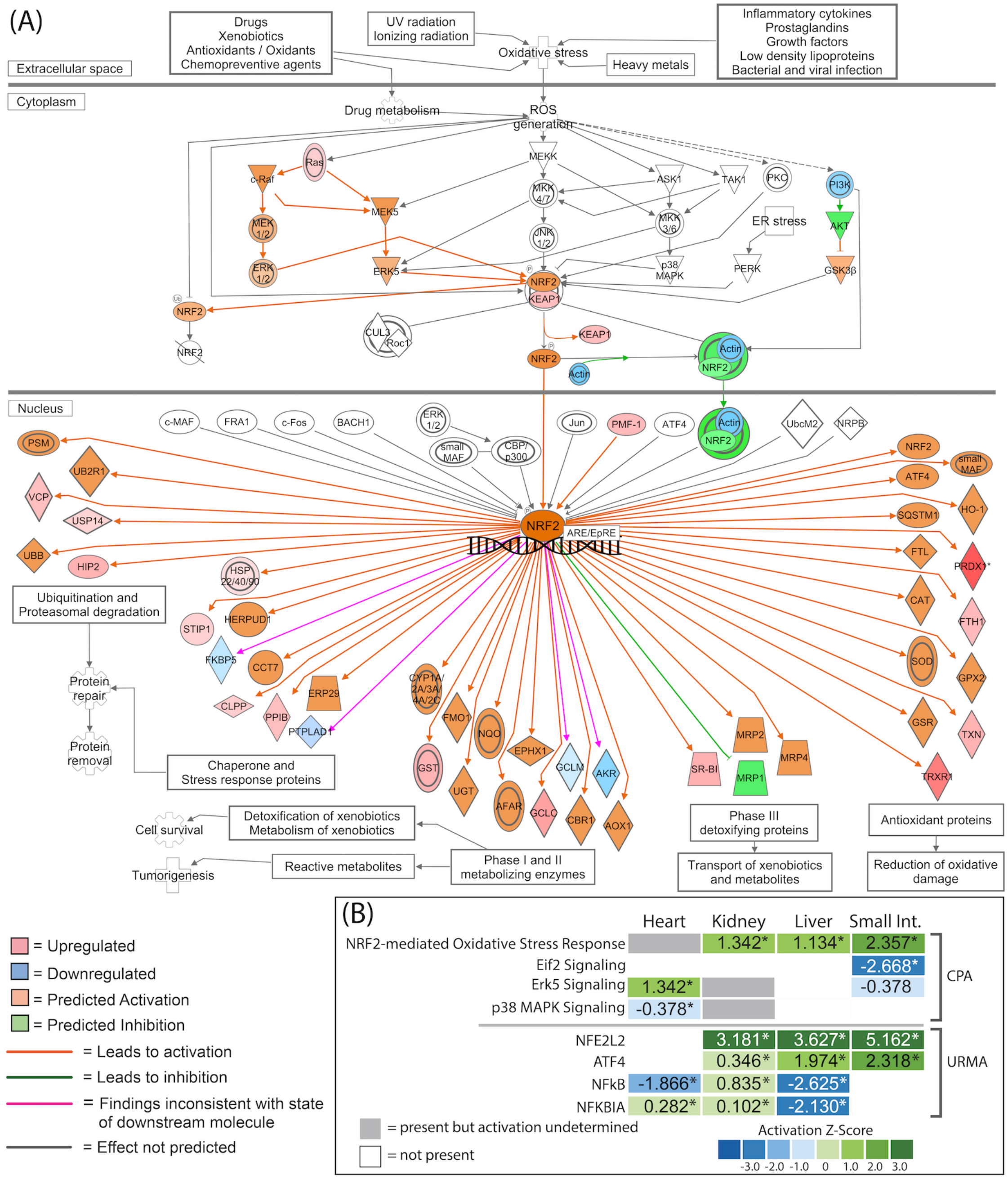
▨ = present but activation undetermined

□ = not present

Figure 5. Combined gene expression and predicted activation information for the mTOR pathway in the heart and small intestine. (**A**) Gene expression and predicted activity for the mTOR pathway in the heart. (**B**) Gene expression and predicted activity for the mTOR pathway in the small intestine. Differentially expressed genes identified in our RNAseq data set are highlighted in red (upregulated) and blue (downregulated) while predicted activation states are highlighted in orange (activation) and green (inhibition). (**C**) CPA and URMA results for pathways and upstream regulatory molecules involved in mTOR signaling and other relevant growth pathways.



	Heart	Kidney	Liver	Small Int.
mTOR Signaling	1.342			0.200*
AMPK Signaling	0.447*	-1.633		-1.877
PI3K/AKT Signaling	0.000		1.000	1.147*
MTOR	-0.662*		-0.905*	-2.338*
MTORC1		1.789*		1.486*
RAF1	-2.121*	1.414*	0.181*	0.123*
INSR	0.728*	3.168*	2.578*	3.503*
Insulin	-0.172*	1.933*	2.116*	1.171*
PTEN	-0.274*	-2.228*	-0.746*	-1.050*
SREBF1	-1.555*	4.198*	3.123*	3.298*
SREBF2		4.586*	3.255*	3.216*



	Heart	Kidney	Liver	Small Int.
NRF2-mediated Oxidative Stress Response	▨	1.342*	1.134*	2.357*
Eif2 Signaling				-2.668*
Erk5 Signaling	1.342*	▨		-0.378
p38 MAPK Signaling	-0.378*	▨		
NFE2L2		3.181*	3.627*	5.162*
ATF4		0.346*	1.974*	2.318*
NFkB	-1.866*	0.835*	-2.625*	
NFKBIA	0.282*	0.102*	-2.130*	

Figure 6. IPA generated pathway prediction for the NRF2-mediated oxidative stress response in the small intestine. Predicted activation state of the pathway was estimated using genes identified as significantly differentially expressed from our RNAseq data set.

Table 1. Numbers of differentially expressed genes between pre- and post-feeding time points for the four organs studied. For each comparison, the numbers of up and downregulated genes were inferred using pairwise analysis with a Benjamini-Hochberg corrected p-value <0.05.

	Fasted vs. 1DPF		**1DPF vs. 4DPF**		**Fasted vs. 4DPF**	
	Up	**Down**	**Up**	**Down**	**Up**	**Down**
Heart	208	228	36	40	5	3
Kidney	244	100	5	3	125	22
Liver	335	126	29	12	295	76
Small Intestine	1,271	1,042	268	146	547	345

Supplementary Methods

Feeding Experiments

The following information pertains to snakes that were sampled and sequenced for this study (see also Table S2.1, Andrew et al., 2015, and Castoe et al., 2013 for details regarding previously sequenced data incorporated in this study). Burmese pythons (*Python molurus bivittatus*) were purchased within 1-2 months of hatching from commercial vendors. All snakes included in this study originated from captive colonies, were phenotypically normal in coloration (i.e., no albino animals), and ranged in age from 9 months to 6 years (mean = 1.9 years) and in mass from 406 to 5,776 grams (mean = 1,036 g). Snakes were housed individually in 12L plastic bins that slide into customized racks in the Central Animal Care Facility at the University of Alabama. Each bin featured a floor substrate of newspaper and contained a water bowl. All pythons were maintained on a light/dark cycle of 14 hours of light followed by 10 hours of dark. Room temperature was maintained at 26- 28°C and was constantly monitored by the Central Animal Care Facility. Prior to experimentation, pythons were fed weekly a meal of 1-2 rodents (adult mice or small rats) and water was provided ad libitum. Pythons were monitored daily by the Animal Care staff and personnel of the laboratory of Dr. Stephen Secor prior to and during experimentation. There were no interventions in snake care prior to or during experimentation. All experimentation and dissection was performed by Secor lab personnel. No special attention was given to selecting animals randomly from a research colony, however there was an attempt for matching in sexes (7 males: 6 females), so that there would be no bias due to sex in any treatment or the experiment overall. At the time of sampling, all animals were in good health and had not been subjected to any previous procedures or drug administration. Fasted snakes had been fasted for a minimum of 30 days prior to sampling. Snakes of the 1 and 4 days post-feeding treatments had been fasted for 30 days and then fed a rodent meal equal in mass to 25% of the snake body mass, and sampled 1 and 4 days after feeding, respectively. The mean mass of snakes in each treatment were: fasted (1,504 g), 1DPF (892 g), and 4DPF (593 g). At the time of sampling, snakes were sacrificed by humanely severing the spinal cord immediately behind the head; this provided the most efficient and rapid means to obtain organ samples for storage and study without compromising physiological responses of interest. Organ tissues were immediately extracted, snap frozen in liquid nitrogen, and stored at -80°C. Feeding experiments and subsequent sampling of snakes were completed over a span of several years, with fasted snakes sampled in 2005 and 2009, and 1 and 4DPF snakes sampled in 2005 and 2006. There was no particular order to the sampling of tissues from animals. No adverse events occurred during animal care or experimentation, and thus no modifications to the experimental protocol were undertaken as a result.

The Burmese python has become an outstanding animal model (compared to traditional mammal model systems) to explore the cellular and molecular mechanisms underlying regenerative organ growth and physiology, and therefore serves as an excellent replacement for exploring such systems in typical mammalian models. We made efforts to minimize the number of animals used overall in this study, as evident in the relatively small sample sizes for each treatment (3-6 individuals). Additionally, dissection of snakes included the removal and storage of all organs and other tissues (muscle, blood, etc.) so that subsequent studies can utilize these tissues to study this regenerative phenotype in other organ systems without the need for the sacrifice of additional animals.

Supplementary Figures

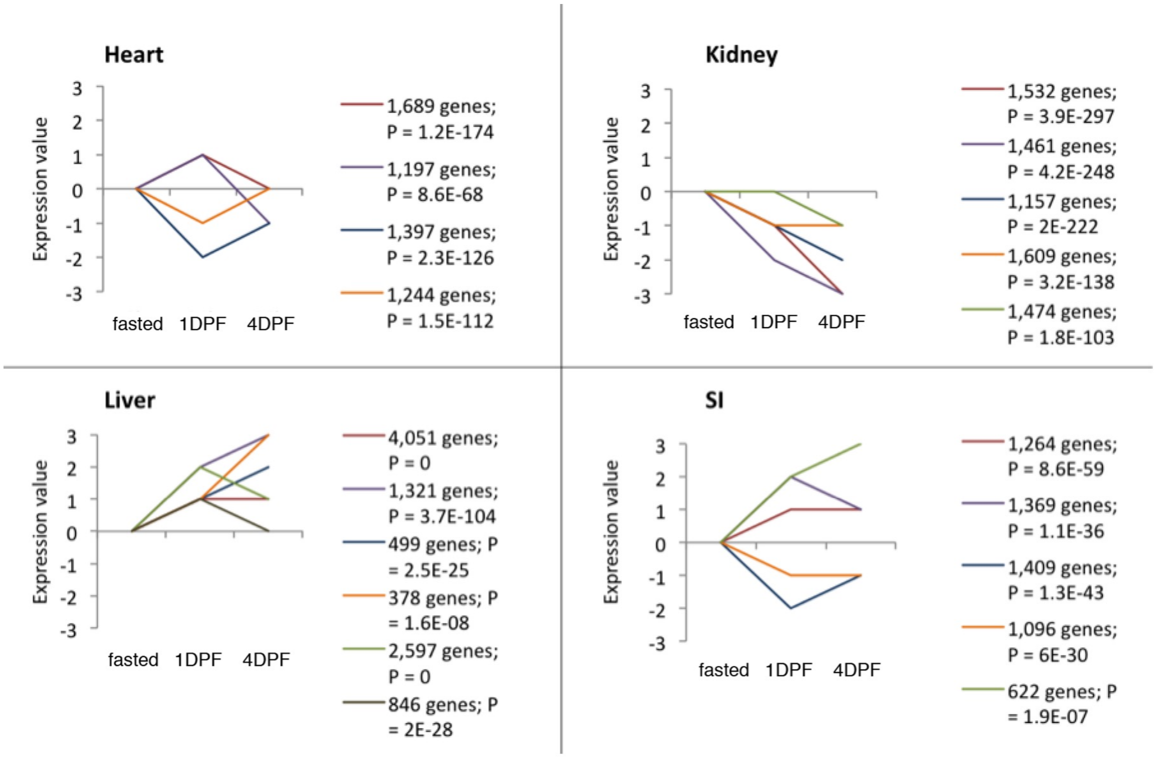
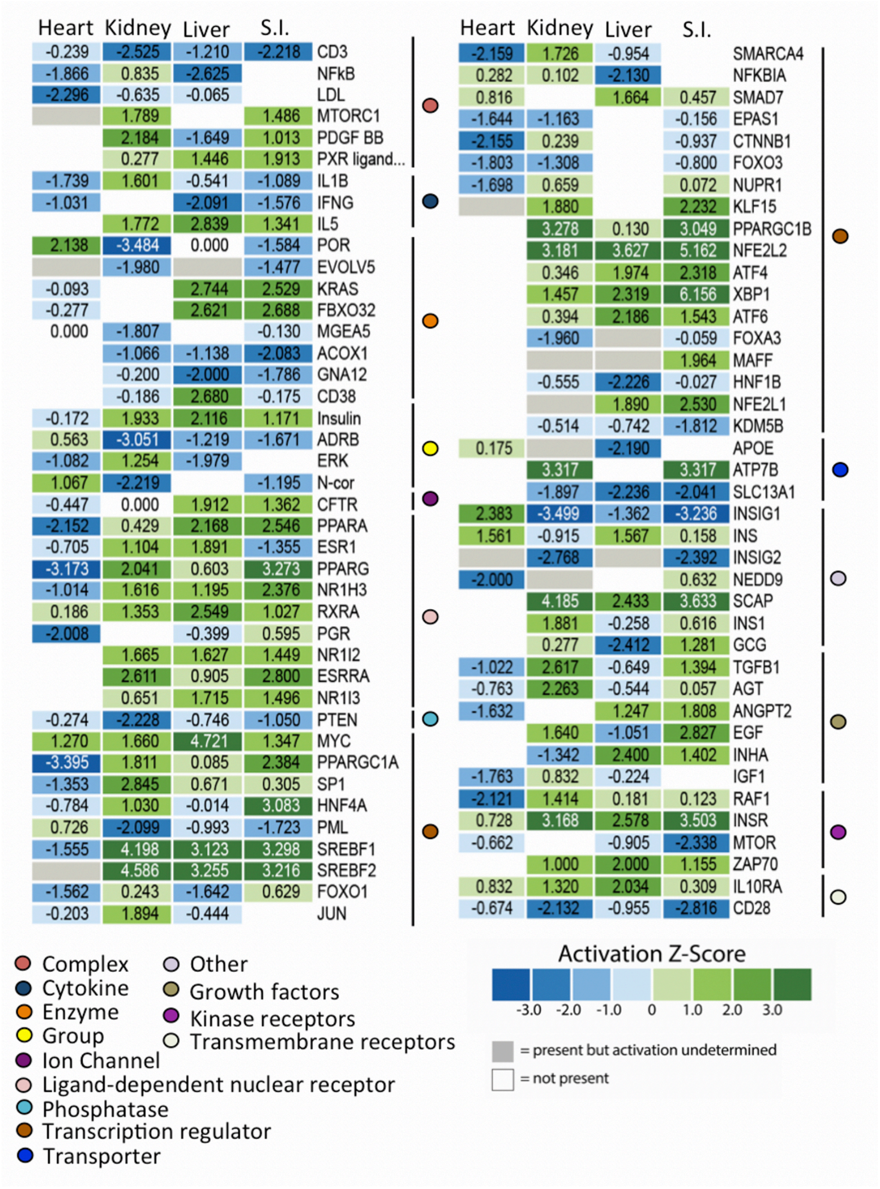


Figure S1. STEM analysis of all genes differentially expressed across all time points (fasted – 4DPF). All significant expression profiles are shown with P-value and number of genes following that profile.



Heart	Kidney	Liver	S.I.	
-0.239	-2.525	-1.210	-2.218	CD3
-1.866	0.835	-2.625		NFkB
-2.296	-0.635	-0.065		LDL
▨	1.789		1.486	MTORC1
	2.184	-1.649	1.013	PDGF BB
	0.277	1.446	1.913	PXR ligand...
-1.739	1.601	-0.541	-1.089	IL1B
-1.031		-2.091	-1.576	IFNG
	1.772	2.839	1.341	IL5
2.138	-3.484	0.000	-1.584	POR
▨	-1.980	▨	-1.477	EVOLV5
-0.093		2.744	2.529	KRAS
-0.277		2.621	2.688	FBXO32
0.000	-1.807		-0.130	MGEA5
	-1.066	-1.138	-2.083	ACOX1
	-0.200	-2.000	-1.786	GNA12
	-0.186	2.680	-0.175	CD38
-0.172	1.933	2.116	1.171	Insulin
0.563	-3.051	-1.219	-1.671	ADRB
-1.082	1.254	-1.979		ERK
1.067	-2.219		-1.195	N-cor
-0.447	0.000	1.912	1.362	CFTR
-2.152	0.429	2.168	2.546	PPARA
-0.705	1.104	1.891	-1.355	ESR1
-3.173	2.041	0.603	3.273	PPARG
-1.014	1.616	1.195	2.376	NR1H3
0.186	1.353	2.549	1.027	RXRA
-2.008		-0.399	0.595	PGR
	1.665	1.627	1.449	NR1I2
	2.611	0.905	2.800	ESRRA
	0.651	1.715	1.496	NR1I3
-0.274	-2.228	-0.746	-1.050	PTEN
1.270	1.660	4.721	1.347	MYC
-3.395	1.811	0.085	2.384	PPARGC1A
-1.353	2.845	0.671	0.305	SP1
-0.784	1.030	-0.014	3.083	HNF4A
0.726	-2.099	-0.993	-1.723	PML
-1.555	4.198	3.123	3.298	SREBF1
▨	4.586	3.255	3.216	SREBF2
-1.562	0.243	-1.642	0.629	FOXO1
-0.203	1.894	-0.444		JUN

Heart	Kidney	Liver	S.I.	
-2.159	1.726	-0.954		SMARCA4
0.282	0.102	-2.130		NFKBIA
0.816		1.664	0.457	SMAD7
-1.644	-1.163		-0.156	EPAS1
-2.155	0.239		-0.937	CTNNB1
-1.803	-1.308		-0.800	FOXO3
-1.698	0.659		0.072	NUPR1
▨	1.880		2.232	KLF15
	3.278	0.130	3.049	PPARGC1B
	3.181	3.627	5.162	NFE2L2
	0.346	1.974	2.318	ATF4
	1.457	2.319	6.156	XBP1
	0.394	2.186	1.543	ATF6
	-1.960	▨	-0.059	FOXA3
	▨	▨	1.964	MAFF
	-0.555	-2.226	-0.027	HNF1B
	▨	1.890	2.530	NFE2L1
	-0.514	-0.742	-1.812	KDM5B
0.175	▨	-2.190		APOE
	3.317		3.317	ATP7B
	-1.897	-2.236	-2.041	SLC13A1
2.383	-3.499	-1.362	-3.236	INSIG1
1.561	-0.915	1.567	0.158	INS
▨	-2.768	▨	-2.392	INSIG2
-2.000	▨		0.632	NEDD9
	4.185	2.433	3.633	SCAP
	1.881	-0.258	0.616	INS1
	0.277	-2.412	1.281	GCG
-1.022	2.617	-0.649	1.394	TGFB1
-0.763	2.263	-0.544	0.057	AGT
-1.632		1.247	1.808	ANGPT2
	1.640	-1.051	2.827	EGF
	-1.342	2.400	1.402	INHA
-1.763	0.832	-0.224		IGF1
-2.121	1.414	0.181	0.123	RAF1
0.728	3.168	2.578	3.503	INSR
-0.662		-0.905	-2.338	MTOR
	1.000	2.000	1.155	ZAP70
0.832	1.320	2.034	0.309	IL10RA
-0.674	-2.132	-0.955	-2.816	CD28

Figure S2. Heat maps depicting activation z-scores for classes of upstream regulator molecules significant between fasted and 1DPF. Green indicates predicted activation, blue indicates predicted inhibition, white indicates that the regulator is not predicted to function in that organ, and grey indicates that the upstream regulator is predicted to have significant involvement but the activation state cannot be determined based on the gene expression data. Regulators shown on the heat maps were filtered by activation z-scores greater than |1.5| in at least one tissue.

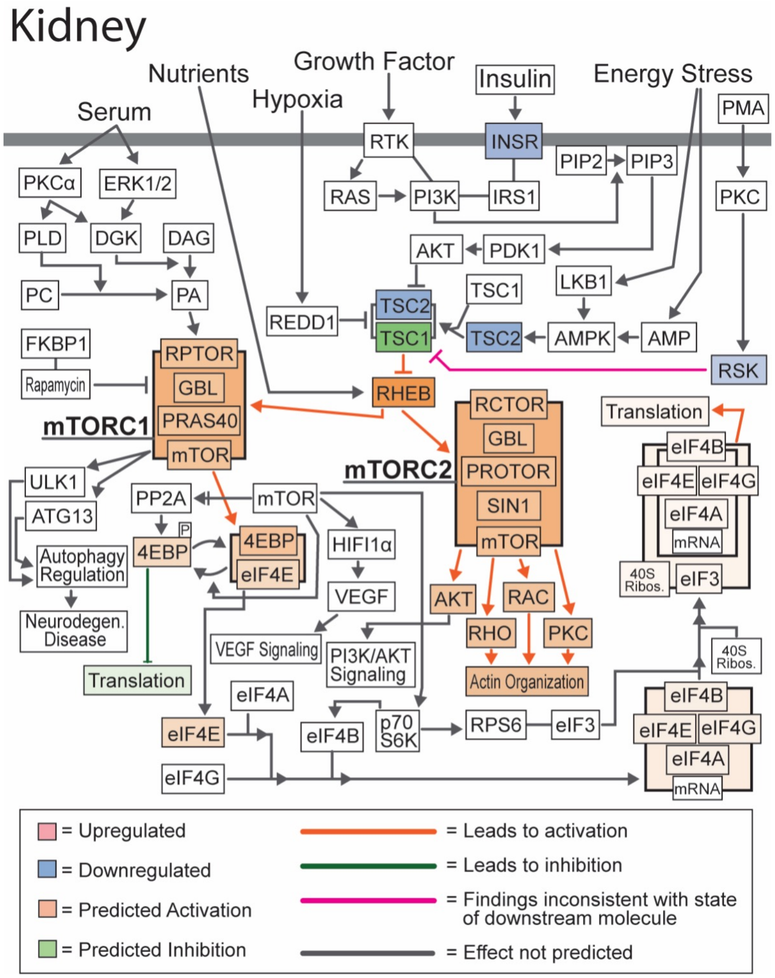


Figure S3. Combined gene expression and predicted activation information for the mTOR pathway in the kidney. Differentially expressed genes identified in our RNA-seq data are highlighted in red (upregulated) and blue (downregulated) while predicted activation states are highlighted in orange (activation) and green (inhibition).

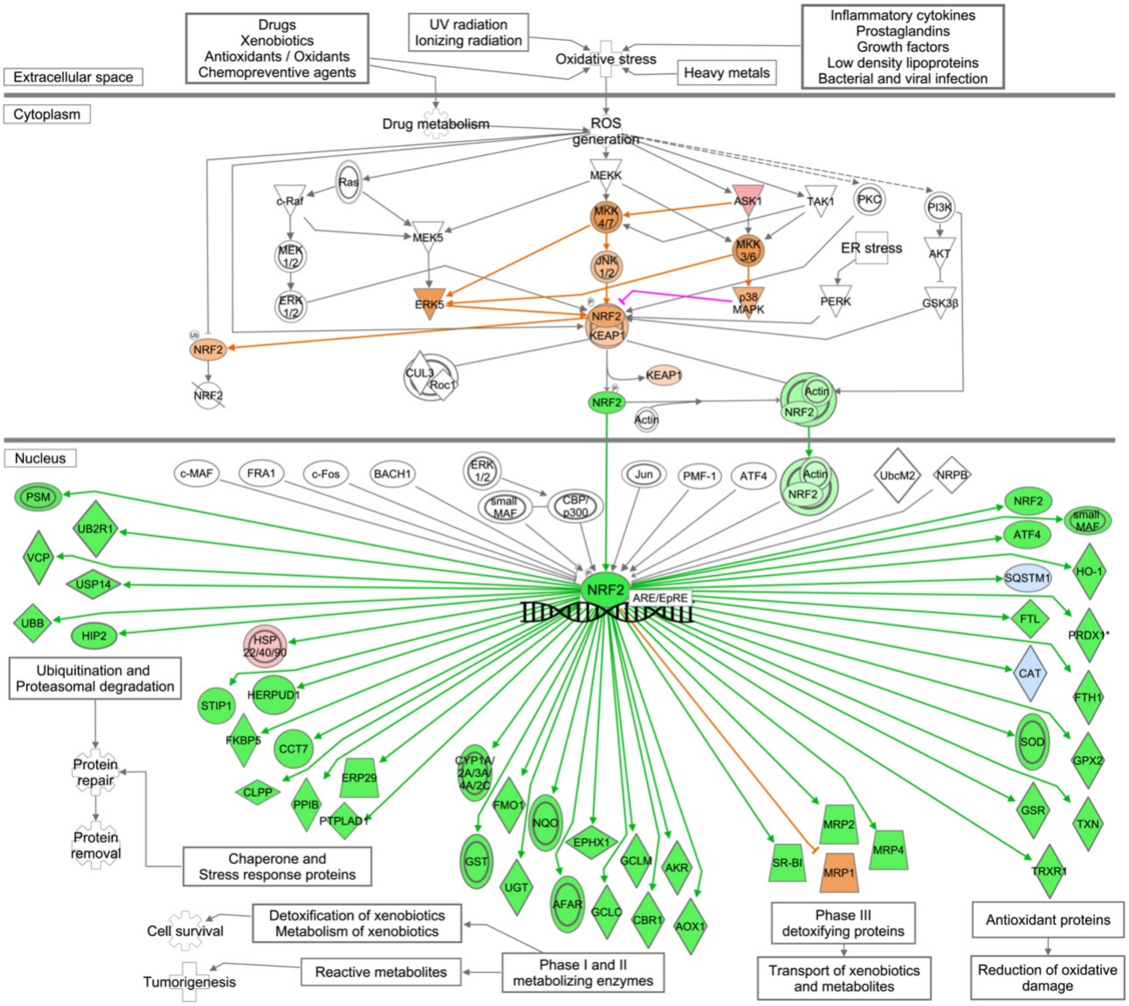


Figure S4. Pathway prediction for the NRF2-mediated oxidative stress response in the heart. Predicted activation state of the pathway was estimated using genes identified as significantly differentially expressed form our RNA-seq data set.

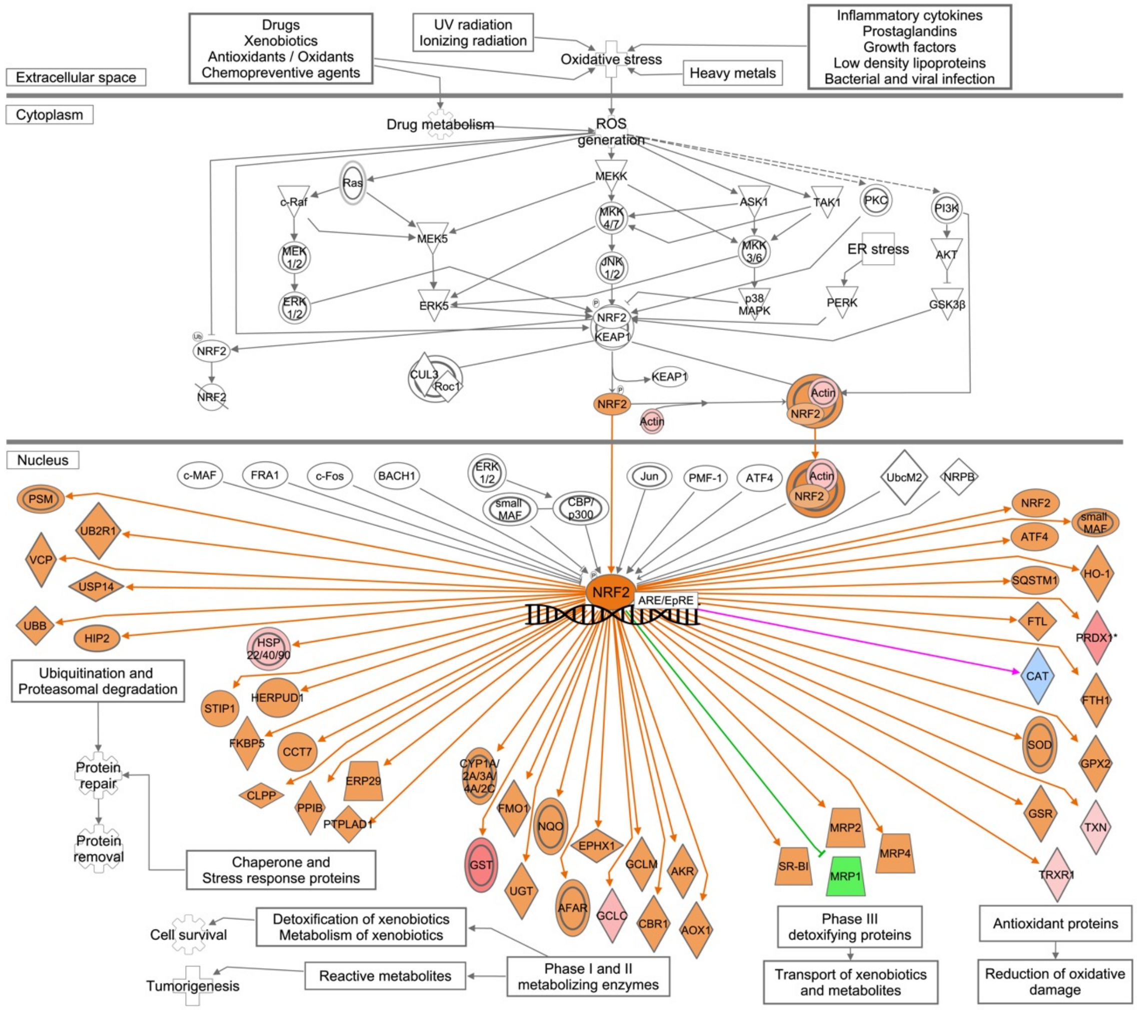


Figure S5. Pathway prediction for the NRF2-mediated oxidative stress response in the kidney. Predicted activation state of the pathway was estimated using genes identified as significantly differentially expressed from our RNA-seq data set.

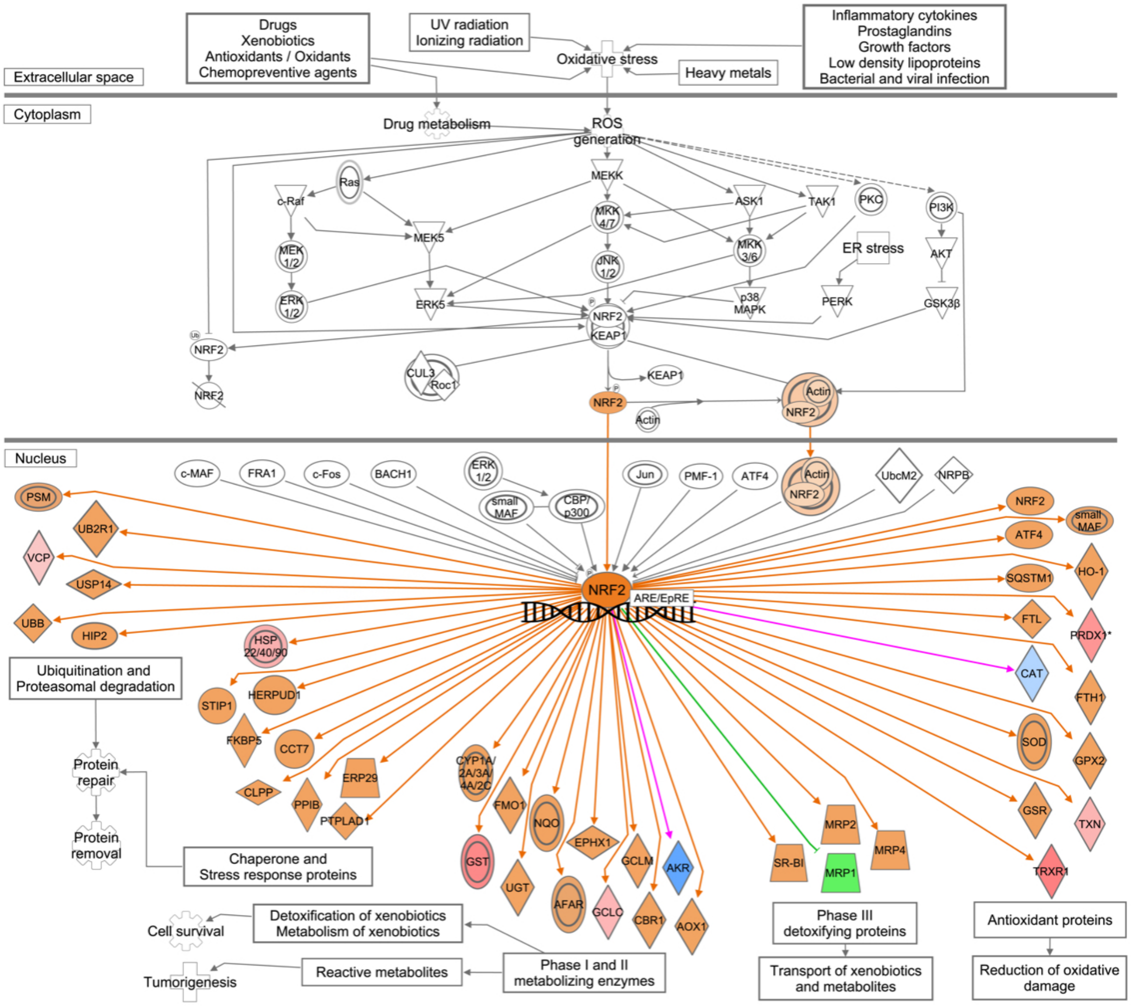


Figure S6. Pathway prediction for the NRF2-mediated oxidative stress response in the liver. Predicted activation state of the pathway was estimated using genes identified as significantly differentially expressed from our RNA-seq data set.

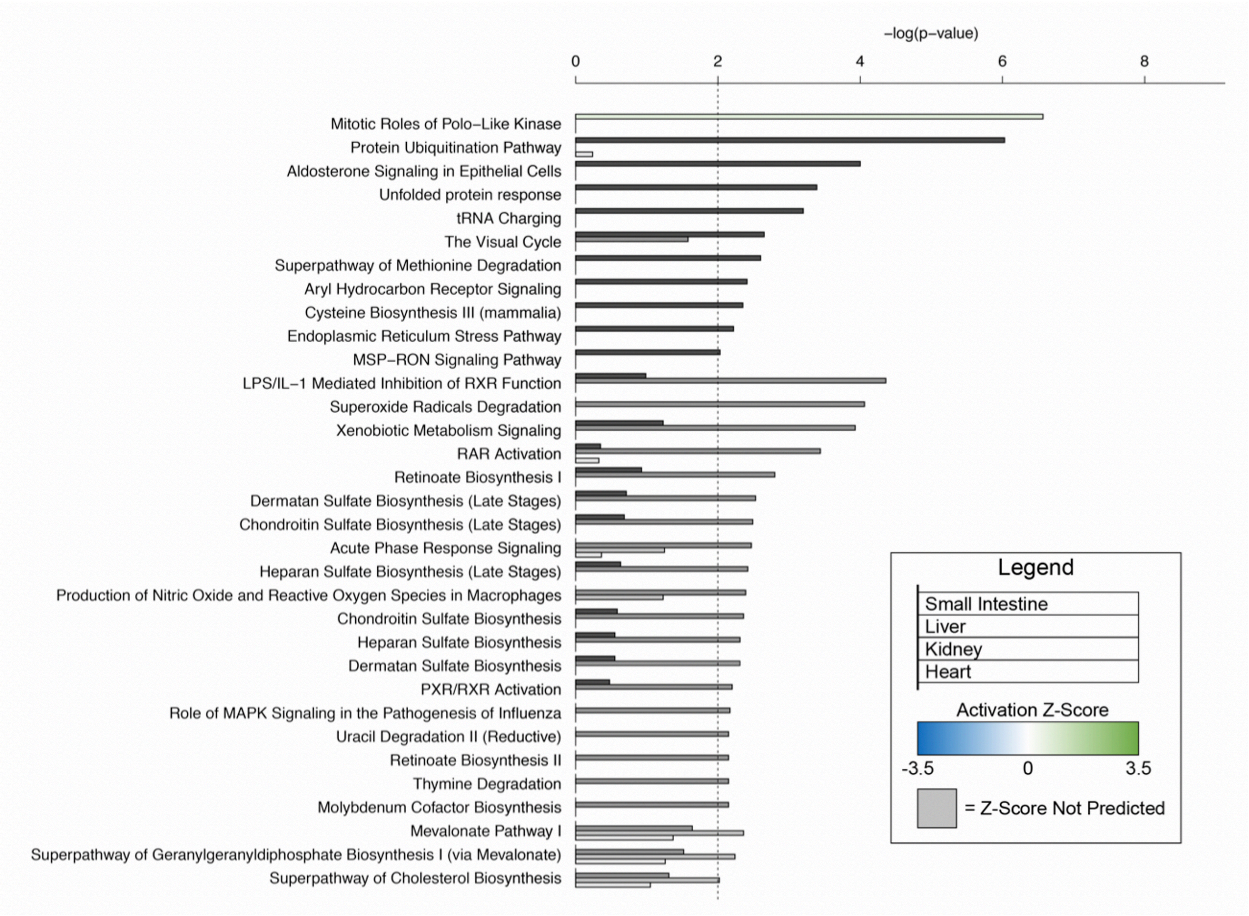


Figure S7. Pathway analysis of all genes significantly differentially expressed from 1DPF to 4DPF in the four organs. Bar graph showing significant canonical pathways (Fisher's Exact test P<0.01) enriched for genes differentially expressed at these time points. Pathways were filtered to include those with at least one significant p-value in one of the four organs. Bars are colored based on the predicted activation Z-score for that pathway.

Supplementary Tables

Table S1. Sequencing information for all included python samples. PE76 and PE120 stand for the sequence read type (e.g., Paired-end 76bp). The year provided represents the year in which the sample was sequenced.

Tissue	Timepoint	Animal ID	Instrument	cDNA prep kit	Year	Sequence type	Library Name
Heart	fasted	AI6_1	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	fasted	AI6_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Heart	fasted	AI11	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	fasted	AI8	HiSeq	NEB Next	2013	SE50	pRNA-A
Heart	fasted	U25	HiSeq	NEB Next	2013	SE50	pRNA-B
Heart	1DPF	Z12	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	1DPF	Z14_1	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	1DPF	Z14_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Heart	1DPF	Z18	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	4DPF	Y5_1	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	4DPF	Y5_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Heart	4DPF	Y18	GAIIx	Illumina Truseq	2010	PE76	TC01
Heart	4DPF	Y23	GAIIx	Illumina Truseq	2010	PE76	TC01
Kidney	fasted	AI8	HiSeq	NEB Next	2013	SE50	pRNA-A
Kidney	fasted	U25	HiSeq	NEB Next	2013	SE50	pRNA-B
Kidney	fasted	AI6_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	fasted	AI6_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	fasted	AI11_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	fasted	AI11_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	fasted	AJ6_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	fasted	AJ6_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	fasted	AJ6_3	GAIIx	Illumina Truseq	2011	PE120	TC05
Kidney	1DPF	Z12_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	1DPF	Z12_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	1DPF	Z14_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	1DPF	Z14_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	1DPF	Z18_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	1DPF	Z18_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	1DPF	Z18_3	GAIIx	Illumina Truseq	2011	PE120	TC05
Kidney	1DPF	V43	HiSeq	NEB Next	2013	SE50	pRNA-B
Kidney	1DPF	Z14_3	HiSeq	NEB Next	2013	SE50	pRNA-B
Kidney	4DPF	Y18_1	HiSeq	NEB Next	2013	SE50	pRNA-B
Kidney	4DPF	Y24	HiSeq	NEB Next	2013	SE50	pRNA-A
Kidney	4DPF	Y5_1	HiSeq	Illumina Truseq	2011	SE50	s1

Kidney	4DPF	Y5_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	4DPF	Y5_3	GAIIx	Illumina Truseq	2011	PE120	TC05
Kidney	4DPF	Y18_2	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	4DPF	Y18_3	GAIIx	Illumina Truseq	2011	PE120	SP03
Kidney	4DPF	Y23_1	HiSeq	Illumina Truseq	2011	SE50	s1
Kidney	4DPF	Y23_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Liver	fasted	AI6_1	GAIIx	Illumina Truseq	2010	PE76	TC01
Liver	fasted	AI6_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Liver	fasted	AI8	HiSeq	NEB Next	2013	SE50	pRNA-A
Liver	fasted	AI11	HiSeq	NEB Next	2013	SE50	pRNA-B
Liver	fasted	U25	HiSeq	NEB Next	2013	SE50	pRNA-B
Liver	1DPF	V43	HiSeq	NEB Next	2013	SE50	pRNA-B
Liver	1DPF	Z14	HiSeq	NEB Next	2013	SE50	pRNA-A
Liver	1DPF	Z18	HiSeq	NEB Next	2013	SE50	pRNA-B
Liver	1DPF	Z12_1	GAIIx	Illumina Truseq	2010	PE76	TC01
Liver	1DPF	Z12_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Liver	4DPF	Y5_1	GAIIx	Illumina Truseq	2010	PE76	TC01
Liver	4DPF	Y5_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Liver	4DPF	Y18	HiSeq	NEB Next	2013	SE50	pRNA-B
Liver	4DPF	Y23	HiSeq	NEB Next	2013	SE50	pRNA-B
Liver	4DPF	Y24	HiSeq	NEB Next	2013	SE50	pRNA-A
Small intestine	fasted	AI8	HiSeq	NEB Next	2013	SE50	pRNA-A
Small intestine	fasted	AI11	HiSeq	NEB Next	2013	SE50	pRNA-B
Small intestine	fasted	U25	HiSeq	NEB Next	2013	SE50	pRNA-A
Small intestine	fasted	AI6_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	fasted	AI6_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	fasted	AI11_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	fasted	AI11_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	fasted	AJ6_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	fasted	AJ6_2	GAIIx	Illumina Truseq	2011	PE120	TC05
Small intestine	fasted	AJ6_3	GAIIx	Illumina Truseq	2011	PE120	TC05
Small intestine	1DPF	Z12_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	1DPF	Z12_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	1DPF	Z14_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	1DPF	Z14_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	1DPF	Z14_3	GAIIx	Illumina Truseq	2011	PE120	TC05
Small intestine	1DPF	Z18_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	1DPF	Z18_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	1DPF	V43	HiSeq	NEB Next	2013	SE50	pRNA-B
Small intestine	1DPF	Z18_3	HiSeq	NEB Next	2013	SE50	pRNA-B
Small intestine	4DPF	Y24	HiSeq	NEB Next	2013	SE50	pRNA-B
Small intestine	4DPF	Y5_1	HiSeq	Illumina Truseq	2011	SE50	s1

Small intestine	4DPF	Y5_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	4DPF	Y18_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	4DPF	Y18_2	GAIIx	Illumina Truseq	2011	PE120	SP03
Small intestine	4DPF	Y18_3	GAIIx	Illumina Truseq	2011	PE120	TC05
Small intestine	4DPF	Y23_1	HiSeq	Illumina Truseq	2011	SE50	s1
Small intestine	4DPF	Y23_2	GAIIx	Illumina Truseq	2011	PE120	SP03

Table S2. The number of genes involved in each pathway as defined by IPA, the number of genes in the pathway that were assigned python orthologs via tblastx, and the number of those python orthologs observed with a non-zero level of expression in our dataset.

Pathway	Organ	Number of Genes	Number of genes assigned an orthologous python gene	Number of genes assigned an orthologous python gene and observed as expressed in dataset
mTOR	Heart	199	172	169
	Kidney			170
	Liver			167
	Small Int.			171
NRF2	Heart	292	223	220
	Kidney			222
	Liver			218
	Small Int.			220

Chapter 3

MULTI-SPECIES COMPARISONS OF SNAKES IDENTIFY COORDINATED SIGNALING NETWORKS UNDERLYING POST-FEEDING INTESTINAL REGENERATION

Blair W. Perry[1], Audra L. Andrew[1], Abu Hena Mostafa Kamal[2], Daren C. Card[1], Drew R. Schield[1], Giulia I.M. Pasquesi[1], Mark W. Pellegrino[1], Stephen P. Mackessy[3], Saiful M. Chowdhury[2], Stephen M. Secor[4], and Todd A. Castoe[1]

[1]Department of Biology, 501 S. Nedderman Dr., The University of Texas at Arlington, Arlington, TX 76010, USA

[2]Department of Ecology and Evolutionary Biology, 1041 E. Lowell St., University of Arizona, Tucson, AZ 85721 USA

[3]Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720 USA

[4]National Natural Toxins Research Center and Department of Chemistry, 975 W. Ave. B., Texas A&M University Kingsville, Kingsville, TX 78363

[5]School of Biological Sciences, 501 20th St., University of Northern Colorado, Greeley, CO 80639, USA

Abstract

Several snake species that feed infrequently in nature have evolved the ability to massively upregulate intestinal form and function with each meal. While fasting, these snakes downregulate intestinal form and function and upon feeding restore intestinal structure and function through major increases in cell growth and proliferation, metabolism, and upregulation of digestive function. Previous studies have identified changes in gene expression that underlie this regenerative growth of the python intestine, but the unique features that differentiate this extreme regenerative growth from non-regenerative post-feeding responses exhibited by snakes that feed more frequently remain unclear. Here, we leveraged variation in regenerative capacity across three snake species – two distantly-related lineages (*Crotalus* and *Python*) that experience regenerative growth, and one (*Nerodia*) that does not – to infer molecular mechanisms underlying intestinal regeneration using transcriptomic and proteomic approaches. Utilizing a comparative approach, we identify a suite of growth, stress response, and DNA damage response signaling pathways with inferred activity specifically in regenerating species, and propose a hypothesis model of interactivity between these pathways that may drive regenerative intestinal growth in snakes.

Introduction

Snakes have emerged as a model system in which to study the regulation of intestinal form and function due to the extreme degree of intestinal modulation (5 to 10-fold) and regenerative growth that some heavy-bodied, infrequently feeding species experience upon feeding after a prolonged fast. At the completion of digestion, these snakes exhibit intestinal atrophy through decreases in cell proliferation and increases in apoptosis of enterocytes, reductions in microvillus length, and downregulation of metabolism and digestive function (Ott and Secor, 2007; Secor, 2008; Secor and Diamond, 1998, 2000). Immediately following the ingestion of a meal, the intestine is rapidly restored, resulting in up to 100% increases in intestinal wet mass, 5-fold increases in microvillus length, 44-fold increases in metabolic rate, and the upregulation of intestinal function within 24 hours (Ott and Secor, 2007; Secor, 2008; Secor and Diamond, 1998, 2000). This extreme regenerative response has been primarily studied in the Burmese Python (*Python bivittatus*) (Andrew et al., 2015, 2017; Lignot et al., 2005; Secor, 2005; Secor and Diamond, 1997, 1998) but has also been identified in other infrequently feeding snake species including other python species and several rattlesnake and boa species (Menzel et al., 2012; Ott and Secor, 2007; Reif et al., 2015; Secor and Diamond, 2000; Secor et al., 2001). In contrast, frequently feeding snakes do not regulate intestinal form and function to this degree and instead exhibit relatively narrow regulation (~ 2-fold) similar to that of most vertebrates, including humans (Cox and Secor, 2008; Lignot et al., 2005; Reif et al., 2015). This extreme regenerative phenotype appears to be highly correlated with feeding ecology rather than phylogeny: distantly related snake species with similar feeding ecologies possess comparable extreme regenerative capacity upon feeding, yet some closely related species with divergent feeding ecologies exhibit divergent phenotypes (extreme versus minimal intestinal remodeling) (Secor and Diamond, 1995; Secor and Ott, 2007). Notably, while examples of tissue and limb regeneration have been investigated in other vertebrates, and reptiles specifically (Brockes, 1997; Delorme et al., 2012; Jopling et al., 2010; Lozito and Tuan, 2017; Seifert et al., 2012; Vinarsky et al., 2005), post-feeding regenerative growth in snakes is unique given that it is not a response to tissue loss or damage through injury (i.e., as in limb or tail regeneration), but instead occurs with every meal ingested following a period of prolonged fasting and in the absence of injury.

Recent studies have revealed the apparent role of cellular growth, metabolic, lipid signaling, and stress response pathways during regenerative intestinal growth in the Burmese Python (Andrew et al., 2015, 2017; Riquelme et al., 2011), but have not compared these responses to other species. It is therefore unknown what distinguishes the regenerative response from the modest regulatory response associated with feeding, and thus a number of crucial questions about regenerative mechanisms remain. For example, is regeneration achieved through the activation of unique signaling pathways in infrequently feeding species, or instead through greater magnitude activity of “normal” post-feeding cellular responses associated with feeding? Do species with similar regenerative phenotypes achieve these responses through the same mechanisms, or have different species evolved different mechanistic solutions for intestinal regeneration? And, can we leverage the natural variation in regenerative phenotypes present across different snake species as a comparative framework for dissecting essential mechanisms underlying intestinal regeneration?

This study begins to address these questions by incorporating an additional regenerating species that possesses the capacity for regenerative growth, the Prairie Rattlesnake (*Crotalus viridis*), as well as a non-regenerating species, the Diamondback Watersnake (*Nerodia rhombifer*) with previous and new data from the Burmese Python. We use analyses of transcriptomic and proteomic data across fasted and post-feeding time points to comparatively dissect and characterize molecular responses associated with regenerative intestinal growth in these species of snakes. Our results demonstrate that the rattlesnake and python intestinal responses to feeding are characterized by an suite of growth and stress response signaling mechanisms, some of which appear to be unique to these two regenerating species, while others are shared with the non-regenerating watersnake species but likely active at different magnitudes in the two regenerative species. Based on our comparative analyses, we develop a hypothetical model to explain potential signaling networks that may underlie regenerative growth following feeding in the snake intestine.

Materials and Methods

Feeding experiments

Prairie Rattlesnakes (*Crotalus viridis viridis*; Colorado: 17HP0974 to SPM) and Diamondback Watersnakes (*Nerodia rhombifer*; Mississippi Scientific Collecting Permit for 2015; No. 0508152 – to SMS) were wild-caught under state collecting permits. Animal care and tissue sampling was conducted under protocols approved by the Institutional Animal Care and Use Committee at the University of Alabama (14-06-0075) and the University of Northern Colorado (1701D-SM-S-20). Individuals from both species were sampled at three time points to target three distinct physiological states: fasted (30 days since last meal), 1 day post-feeding (DPF), and 4 DPF. Watersnakes were fed catfish filets and rattlesnake were fed adult mice in order to approximate their natural primary prey type (fish and mammals, respectively). Consumed meals were equal in mass to 25% of individual snake body mass. Snakes were humanely euthanized by severing the spinal cord immediately posterior to the head. Intestinal tissue was immediately extracted, snap frozen in liquid nitrogen, and stored at -80°C. Between three and four individual animals per species were sampled for each time point.

Transcriptomic data generation

Transcriptomic data for *C. viridis* and *N. rhombifer* were generated for this study. Transcriptomic data for the Burmese Python (*P. bivittatus*) was generated previously using protocols described in (Andrew et al. 2015; Andrew et al. 2017). Total RNA was extracted by placing ~50mg of snap-frozen tissue into 1mL of Trizol Reagent (Invitrogen), mechanically lysing cells using a TissueLyzer for 10 minutes at 20 strokes/minute, and precipitating RNA from the aqueous phase using isopropanol. Individual Illumina mRNA-seq libraries were constructed using either a NEB Next RNAseq kit with poly-A selection, RNA fragmentation, cDNA synthesis, and indexed Illumina adapter ligation. RNAseq libraries were quantified using a BioAnalyzer (Agilent), pooled in equal molar ratios in various multiplex arrangements, and sequenced on an Illumina HiSeq 2500 using 100bp paired-end reads (Supp. Table. S1). Newly-generated transcriptomic data for *C. v. viridis* and *N. rhombifer* is archived at the NCBI Short Read Archive (NCBI: SRP200900) as well as previously-generated transcriptomic data for *P. bivittatus* (NCBI: SRP051827).

Identifying homologous genes between snakes and human

Downstream pathway and regulatory molecule analyses require gene expression data with human gene identifiers. To identify homologous genes between the snake species and human and ultimately assign human gene identifiers to snake genes for downstream analysis, OrthoMCL (Li et al., 2003) was run via the Orthomcl-pipeline (https://github.com/apetkau/orthomcl-pipeline) using species-specific protein fasta files as input. In cases where multiple isoforms were annotated for a given gene, the protein sequence corresponding to the longest coding sequence was used. For resulting OrthoMCL homolog groups containing a single gene from each of the four species, the identifier of the human gene was assigned to each snake gene within the group. For groups containing a single human gene identifier but more than one orthologous gene in one or more snake species, the human gene identifier was assigned to all snake genes in the group. Lastly, for groups containing multiple human orthologs, each snake gene in the group was assigned the human identifier that produced the best one-way BLAST hit to that gene during the orthoMCL pipeline.

Quantifying and visualizing gene expression

Raw demultiplexed Illumina RNAseq reads were quality filtered and trimmed with Trimmomatic v. 0.32 (Bolger et al., 2014). All reads were mapped using STAR v.2.5.3a (Dobin et al., 2013) in basic two pass mode with --outFilterMultimapNmax set to 1 to exclude reads that mapped to multiple positions. Reads for *P. bivittatus* and *C. v. viridis* were mapped to their respective genome assemblies (Castoe et al., 2013; Schield et al., 2019); as no genome assembly is currently available for *N. rhombifer*, these reads were mapped to genome of the closely-related eastern garter snake (*Thamnophis sirtalis*; Perry et al., 2018). Raw expression counts were determined using featureCounts v1.6.3 (Liao et al., 2013).

Count normalization and pairwise exact tests of differential expression between fasted versus 1DPF and 1DPF versus 4DPF were performed using DEseq2 v. 1.12.4 (Love et al., 2014). As transcriptomic data for *P. bivittatus* included both single-end 50bp and paired-end 120bp reads, read type was included as a factor in DEseq2 pairwise comparisons to account for any potential batch effects. Independent hypothesis weighting was applied to pairwise test results using the IHW package (Ignatiadis et al., 2016) with baseMean as a covariate.

The results of pairwise comparisons in each species were filtered to exclude genes that were not assigned a homologous gene identifier via OrthoMCL. We further filtered pairwise results to only include genes for which there was detectable expression (>= 1 raw expression count) in at least one time point in all three snake species.

Heatmaps were generated using the pHeatmap package (Kolde, 2012) and alluvial plots of patterns of gene expression were generated using the ggalluvial package (Brunson, 2018) in R (R Core Team, 2014).

Proteomic data generation and analysis

Proteins were extracted using T-PER Tissue Protein Extraction Reagent (Thermo Fisher, 78510) from multiple fasted, 1 DPF, and 4DPF proximal small intestine tissue samples for the python and watersnake (Supp. Table. S4). Proteins were quantified using a BCA assay, purified, and digested with trypsin. The dried pellet was resuspended in 50 mM NH4CO3. Following Kamal et al. 2018 (Kamal et al. 2018), proteins were reduced and alkylated, then digested with Trypsin (MS Grade) at a 1:50 enzyme/protein concentration for 16 h at 37 °C. Formic acid (pH < 3) was added to the resulting peptides for acidifying the sample. A C18 desalting column (ThermoFisher Scientific, IL, USA) was used for desalting the samples. After drying by speed vacuum, peptides were dissolved in 0.1% formic acid, and stored at -20°C.

Digested peptides were analyzed by nano-LC-MS/MS using a Velos Pro Dual-Pressure Linear Ion Trap Mass Spectrometer (ThermoFisher Scientific, MA) coupled to an UltiMate 3000 UHPLC (ThermoFisher Scientific, MA). Peptides were loaded onto the analytical column and separated by reversed-phase chromatography using a 15-cm column (Acclaim PepMap RSLC) with an inner diameter of 75 µm and packed with 2 µm C18 particles (Thermo Fisher Scientific, MA). The peptide samples were eluted from the Nano column with multi-step gradients of 4-90% solvent B (A: 0.1% formic acid in water; B: 95% acetonitrile and 0.1% formic acid in water) over 70 min with a flow rate of 300 nL/min with a total run time of 90 min. The mass spectrometer was operated in positive ionization mode with nano spray voltage set at 2.50 kV and source temperature at 275°C. The three precursor ions with the most intense signal in a full MS scan were consecutively isolated and fragmented to acquire their corresponding MS2 scans. Full MS scans were performed with 1 micro scan at resolution of 3000, and a mass range of *m/z* 350-1500. Normalized collision energy (NCE) was set at 35%. Fragment ion spectra produced via high-energy collision-induced dissociation (CID) was acquired in the Linear Ion Trap with a resolution of 0.05 FWHM (full-width half maximum) with an Ultra Zoom-Scan between *m/z* 50-2000. A maximum injection volume of 5 µl was used during data acquisition with partial injection mode. The mass spectrometer was controlled in a data-dependent mode that toggled automatically between MS and MS/MS acquisition. MS/MS data acquisition and processing were performed by Xcalibur^{TM} software (ThermoFisher Scientific, MA).

Spectra were searched using Proteome Discoverer software (ver. 2.0, ThermoFisher Scientific) against species-specific protein databases generated from the genome of the Burmese Python (Castoe et al., 2013) and the genome of the garter snake, *Thamnophis sirtalis* (Perry et al., 2018), which was used as a reference set for the watersnake as it is the most closely-related snake species with a complete genome assembly and annotation. The considerations in SEQUEST searches for normal peptides were used with carbamidomethylation of cysteine as the static modification and oxidation of methionine as the dynamic modification. Trypsin was indicated as the proteolytic enzyme with two missed cleavages. Peptide and fragment mass tolerance were set at ± 1.6 and 0.6 Da and precursor mass range of 350-5000 Da, and peptide charges were set excluding +1. SEQUEST results were filtered with the target PSM validator to improve the sensitivity and accuracy of the peptide identification. Using a decoy search strategy, target false discovery rates for peptide identification of all searches were < 1% with at least two peptides per proteins, and the results were strictly filtered by ΔCn (< 0.01), Xcorr (≥1.5) for peptides, and peptide spectral matches (PSMs) with high confidence (q-value of ≤ 0.05). Protein quantification was conducted using the total spectrum count of identified proteins. Additional criteria were applied to increase confidence that PSMs must be present in all three biological replicates samples. Protein identifiers from the rattlesnake and garter snake genomes were converted to orthologous python identifiers using reciprocal best BLASTp, followed by reciprocal best tBLASTx and, finally, stringent one-way BLASTp. Peptide spectrum matches (PSM) were normalized and analyzed using DEseq2 (Love et al. 2014) in R to identify proteins that exhibited significant changes in abundance between time points (Benjamini-Hochberg corrected p-value < 0.1). Differentially expressed proteins were characterized using GO term overrepresentation analysis against all protein coding genes using ClueGO (Bindea et al., 2009) and WebGestalt (Zhang et al., 2005).

Inferences of canonical pathway and upstream regulatory molecule activity

To infer canonical pathways and regulatory molecules that may be driving observed patterns of gene expression in these three species, and specifically those that may be driving regenerative growth, we employed a multi-level comparative approach to analyze particular subsets of differentially expressed genes ($p < 0.05$) using Core Analysis in Ingenuity Pathway Analysis software (Krämer et al., 2013). First, we used all differentially expressed genes from pairwise analyses as input. Second, to distinguish between mechanisms that are uniquely shared between regenerating species and those that are shared between all three species, we compared inferences generated from genes that were differentially expressed in both the python and rattlesnake to inferences generated from genes that were differentially expressed in all three species. To compare evidence of pathway and regulatory activity inferred from gene expression and protein abundance, we also used Core Analysis in IPA to analyze proteins that demonstrated significant changes in abundance between Fasted and 4DPF ($p < 0.1$). Networks of overlapping genes between pathways was generated using the GGally (Schloerke et al., 2018) and network (Butts, 2008; Handcock et al., 2008) packages in R (Fig. 3.3). In these networks, pathways were connected if at least 50% of the genes underlying the prediction of one of the pathway overlap with the genes underlying the prediction of the other pathway. Pathways that did not exhibit this overlap with at least one additional pathway are shown as unconnected nodes. Networks were manually annotated to group pathways based on similar biological function.

Results

Rapid and massive changes in gene expression in regenerating species

Between fasting and 1 day post-feeding (DPF), the regenerating python exhibits the largest number of differentially expressed (DE; $p < 0.05$) genes (2,559), followed by the regenerating rattlesnake (1,439); the number of DE genes in the non-regenerating watersnake is substantially smaller (793; Fig. 1). The python and rattlesnake shared 767 DE genes between fasted and 1DPF, 563 of which were uniquely DE in these two regenerating species. Pairwise comparisons of 1DPF and 4DPF revealed a considerably larger response in the python (1,595 DE genes) compared to that observed in both the rattlesnake (376) and watersnake (194) during this interval (Fig. 1C). Across all three species, many genes with significant up- or downregulation between fasting and 1DPF showed no differential expression between 1DPF and 4DPF (Fig. 1D). A smaller number of genes showed a change in direction of differential expression, continued differential expression in the same direction (i.e. upregulated Fasted vs. 1DPF and upregulated 1DPF vs. 4DPF), or delayed regulation (i.e. not DE in Fasted vs. 1DPF, but DE in 1DPF vs. 4DPF; Fig. 2A).

Conserved regulatory molecule and pathway activity in regenerating species

To infer patterns of canonical pathway and regulatory molecule activity following feeding based on our transcriptomic data, we first performed Core Analysis in IPA (Qiagen) using all DE genes for each of the three species and identified pathways and upstream regulatory molecules (URMs) that showed one of two patterns predicted to be informative for dissecting mechanisms of regenerative growth: 1) those with significant inferences of regulatory activity in only the two regenerating species ($p < 0.05$), and 2) those with significant inferences of activity in all three species ($p < 0.05$). IPA analyses of all DE genes between fasted and 1DPF inferred significant regulatory activity of pathways associated with cellular growth, proliferation, and metabolism signaling in all three species, including the PI3K/AKT Signaling, ERK/MAPK Signaling, PDGF Signaling, Insulin Receptor Signaling, and JAK/Stat Signaling pathways (Supp. Fig. 1A). Additionally, multiple pathways associated with cellular stress responses were inferred to regulate DE genes between fasted and 1DPF, including the NRF2-mediated Oxidative Stress Response pathway and pathways associated with endoplasmic reticulum stress and the Unfolded Protein Response (Supp. Fig. 1A). Fewer pathways were inferred to drive gene activity in the regenerating python and rattlesnake alone in this analysis (Supp. Fig. 1A), several of which are associated with DNA damage repair and tumor suppression. URMs associated with growth, metabolism, and stress response signaling, including nuclear factor erythroid 2 like 2 (NFE2L2), insulin receptor (INSR), sterol regulatory element binding transcription factor 1 and 2 (SREBF1/2), several peroxisome proliferator-activated receptor (PPAR) molecules, and x-box binding protein 1 (XBP1) were inferred to be significantly activated in all three species between fasted and 1DPF, while few URMs were significant only in the python and rattlesnake (Supp. Fig. 1B). Several pathways with inferred activity between fasted and 1DPF, including the Protein Ubiquitination Pathway, Aldosterone Signaling in Epithelial Cells, and Endoplasmic Reticulum Stress Pathway, were also inferred to be actively regulating genes in all three species during the later 1DPF versus 4DPF interval (Supp. Fig. 1C). Additionally, multiple URMs including XBP1, activating transcription factor 4 (ATF4), and NFE2L2 were inferred to be inhibited or downregulated during this later time point compared to their inferred activation between fasted and 1DPF (Supp. Fig. 1D).

To further dissect regulatory mechanisms that may explain unique patterns of gene regulation in the two regenerating species, we performed separate IPA Core Analyses on targeted subsets of DE genes that were 1) DE only in the python and rattlesnake and 2) DE in all three species. The resulting IPA inferences of pathway and regulatory molecule were categorized based on patterns of overlap between the analyses of these two gene sets: regulatory mechanisms inferred only from analyses of DE genes shared between all three species were considered "feeding" mechanisms, those inferred only from analyses of DE genes shared between the python and rattlesnake were considered "regeneration unique," and mechanisms inferred from analyses of both gene sets were considered to be "shared" between the feeding and regenerative response (Fig. 2). "Feeding" pathways (Fig. 2A, "Feeding") included many of the same growth and stress response pathways inferred in the above analyses based on all DE genes (Supp. Fig. 1). "Shared" pathways inferred from analyses of both gene sets indicate that some pathways may respond with greater magnitude and/or breadth (in terms of the number of DE genes being regulated) in regenerating species (Fig. 2A, "Shared"), including the NRF2-mediated Oxidative Stress Pathway, which was previously implicated in regenerative growth in studies of the Burmese Python (Andrew et al., 2017). "Regeneration unique" pathways included many pathways associated with DNA damage repair and tumor suppression, as well as several growth and metabolism pathways including the Insulin Receptor and Insulin-like Growth Factor 1 (IGF-1) Signaling pathways, ERK5 Signaling, and JAK/Stat Signaling (Fig. 2A, "Regen Unique"). Inferences of "shared" URMs (Fig. 2B, "Shared") suggest that the two regenerating species respond to feeding by differentially expressing additional sets of genes potentially regulated by NFE2L2, XBP1, INSR, which are major regulators within the NRF2-mediated Oxidative Stress Response, Unfolded Protein Response, and Insulin Receptor Signaling pathways, respectively. This indicates that although these URMs show activity in all three species, they are potentially regulating a larger number of DE genes in species that show regenerative post-feeding responses (pythons and rattlesnakes), which are not DE in the non-regenerating watersnake. These URMs may therefore contribute to regeneration-specific signaling beyond a baseline level general feeding response signaling. "Regeneration unique" URMs included fibroblast growth factor 21 (FGF21), a known regulator of growth and metabolism (Fisher and Maratos-Flier, 2016), and matrix metallopeptidase 3 (MMP3), which is involved in the breakdown of the extracellular matrix during tissue remodeling and growth and has specifically been implicated in limb regeneration in newts (Vinarsky et al., 2005) (Fig. 2B).

To assess the potential interaction among inferred canonical pathways, networks of pathways were constructed based on the overlap of genes underlying inferred pathway activity (Fig. 3). In these networks, a connection between two pathways indicates that at least 50% of the genes underlying the inferred activity of one pathway were also underlying the inference of the other pathway. The feeding response network, generated from "feeding" and "shared" pathways described above, features a large interconnected group of pathways associated with cellular growth, metabolism, and homeostasis (Fig. 3A). The NRF2-mediated oxidative stress response pathway and hypoxia signaling in the cardiovascular system pathway overlap with pathways within this growth-related cluster, suggesting the potential integration of growth and oxidative stress response signaling during the feeding response. Other stress response pathways did not show direct overlap with this group; these include the Unfolded Protein Response and Endoplasmic Reticulum Stress Pathway, which are connected with the growth-related group via the Protein Ubiquitination Pathway, Aldosterone Signaling in Epithelial Cells pathway, and Glucocorticoid Receptor Signaling pathway.

The regenerative response network, generated from "regeneration unique" and "shared" pathways described above, also exhibited an interconnected group of growth-related pathways (Fig. 3B), although the pathways within this cluster were distinct from growth-related pathways in the feeding response network (Fig. 3A). In this regenerative response network, NRF2-mediated Oxidative Stress Response was again directly interconnected to this growth-related group, but here via the JAK/Stat signaling pathway. This network also features a group of cell junction signaling pathways and a group of DNA damage repair pathways, one of which (Role of BRCA1 in DNA Damage Response) connects directly with the growth-related group via AMPK signaling, with other DNA damage response pathways forming a separate cluster of interconnected pathways (Fig. 3B).

Based on interconnected pathways inferred from DE genes unique to the python and rattlesnake and known biological interactions and consequences of these pathways, we developed a hypothesis network for how growth and stress response mechanisms drive regenerative growth in snakes (Fig. 3C). Key features of this model are the stimulation of regenerative growth through growth factor signaling via cell junction signaling as well as cell surface receptor signaling (e.g., Insulin receptor), and the interaction and coordination of multiple growth pathways with stress response and DNA Damage response pathways (Fig. 3C).

Proteomic changes underlying regenerative growth

For comparison with inferences based on RNAseq data, we generated quantitative shotgun proteomic data for an overlapping set of samples. We successfully quantified intestinal protein abundance for 857 and 637 proteins for the python and watersnake, respectively. In both the python and watersnake, the number of proteins showing significant changes in abundance between time points (FDR < 0.1) was greatest between fasted and 4DPF (Fig. 4A). Of the 68 differentially abundant proteins in the watersnake, 53 were successfully matched to an orthologous python protein ID and were used in downstream characterization and analysis. The 12 differentially abundant proteins between fasted and 4DPF in the python and watersnake were enriched for GO terms relating to cell-cell adhesion and oxidation-reduction processes ($p < 0.05$; Supp. Fig. 3). GO term analysis of the 97 proteins differentially abundant only in the python revealed several significantly overrepresented terms relevant to regenerative growth, including RNA and unfolded protein binding, actin cytoskeleton regulation, and regulation of anatomical structure size and cellular component biogenesis (Fig. 4C). At each of the three sampled time points, a weak but significant positive correlation was found between RNA expression and protein abundance in the python and watersnake when excluding data points with low RNA expression or low protein abundance ($p < 0.05$; Supp. Fig. 8). In both species, the correlation between RNA and protein abundance was weakest at the fasted time points and strongest at 1DPF (Supp. Fig. 8).

IPA Core Analysis based on differentially abundant proteins between fasted and 4DPF was compared to pathway and URM activity predictions inferred from patterns of gene expression. Relatively few canonical pathways were inferred to have significant activity based on protein data alone (Supp. Fig. 4A), likely due to the small size of the input datasets. However, several pathways that were inferred to have significant regulatory activity in the python based on gene expression data were also inferred to be significant based on differential protein abundance, including the growth and metabolism-related VEGF signaling pathway. The NRF2-mediated oxidative stress response pathway was inferred to be significantly active in both protein and RNA-derived analysis in the python and watersnake (Supp. Fig. 4A). URM analysis showed more consistency between analyses based on protein and gene expression datasets (Fig. 4D). Many URMs inferred from gene expression to be activated between fasted and 1DPF in the python and watersnake showed similar activation patterns based on differential protein abundance, including URMs associated with key growth and stress response pathways such as NFE2L2, KRAS, PPARA and PPARG, and EGF (Fig. 4D).

Discussion

Interest in leveraging snakes to study the mechanisms underlying extremes of vertebrate organ regenerative growth, including intestinal regeneration, has steadily increased since the discovery of their extreme post-feeding regenerative capacities over 20 years ago (Secor and Diamond, 1995). Recent molecular studies that have made progress in understanding the signaling mechanism underlying these responses have focused on the Burmese Python (Andrew et al., 2015, 2017), but have lacked a cross-species comparative context that might differentiate post-feeding regeneration responses from general feeding responses. Here, we provide the first multi-species comparison of post-feeding organ regenerative responses in snakes by analyzing the response of two species that do, and a third that does not, regenerate upon feeding. Our results indicate that the regenerating python and rattlesnake exhibit significant differential expression of thousands of genes following feeding, including a large number of shared genes that do not respond in the non-regenerating watersnake. Responsive genes in the two regenerating species show greater overlap with one another than they do with the non-regenerating watersnake, indicating that some mechanisms of regenerative growth responses are shared between these two regenerating species despite ~90 million years of divergence (Castoe et al., 2009; Kumar et al., 2017; Zheng and Wiens, 2016).

Inferences of growth pathway activity between regenerating and non-regenerating snake species

To investigate signaling mechanisms that may differentiate regeneration versus feeding responses, we separately inferred pathway and regulatory molecule activity for differentially expressed genes shared between all three species (i.e., those likely associated with a general feeding response) and genes differentially expressed only in the two regenerating species (i.e., those uniquely associated with the regenerative response). In these targeted analyses, we found evidence for largely non-overlapping sets of canonical pathways and regulatory molecules regulating the core DE genes of the general feeding response in all three species versus the DE genes shared between the python and rattlesnake, suggesting that regenerative growth in these two species involves unique regulatory activity otherwise not active in regulating the feeding response. Our inferences suggest that the general feeding response is largely comprised of a distinct set of pathways associated with cellular growth, metabolism, and homeostasis (Fig. 2, "Feeding"), including PI3K/AKT signaling, which was previously suggested as a potential central regulator of regenerative growth in the Burmese Python (Andrew et al., 2017). In addition to growth signaling pathways, multiple stress-response pathways were inferred to be involved in the general feeding response, including the Unfolded Protein Response and Endoplasmic Reticulum Stress Pathways.

The regenerative response involves a distinct set of growth and metabolism pathways (Fig. 2, "Regen unique"), including known regulators of vertebrate growth, tissue repair, and regeneration such as the IGF-1 Signaling and Insulin Receptor Signaling pathways (Beyer and Werner, 2008; Desbois-Mouthon et al., 2006). Insulin Receptor Signaling and associated downstream pathways have been implicated in reptile longevity, growth, and stress response (Reding et al., 2016; Schwartz and Bronikowski, 2014, 2016), and have undergone rapid evolution in snakes (McGaugh et al., 2015). Insulin Receptor Signaling also interacts with stress response pathways that have been implicated by this and previous studies in the regenerative growth response (Andrew et al., 2017). Previous studies of the Burmese Python have demonstrated that the concentration of circulating insulin, one of the main initiators of the Insulin Receptor Signaling cascade, increases six-fold with 24 hours of feeding (Secor et al., 2001). Unique activity of Insulin Receptor Signaling is therefore a promising candidate for a high-level driver of regenerative growth in snakes.

Our analyses also identified pathways and URMs that were inferred in analyses of distinct DE gene sets associated with both the feeding and regenerative responses (Fig. 2, "Shared"), suggesting that more broad and/or higher magnitude stimulation of these signaling pathways that otherwise exhibit a baseline level of activity during the feeding response may also contribute to the regenerative response. Notably, this group of overlapping regulatory mechanisms included the NRF2-mediated oxidative stress pathway, which was previously suggested to be involved in regenerative growth in the Burmese Python (Andrew et al., 2017). The NRF2 pathway overlaps with distinct growth pathways in both the regenerative and feeding response networks, suggesting the potential for direct integration of growth and stress signaling responses during both feeding and regenerative responses. NFE2L2, the primary regulatory molecule within the NRF2 pathway, was also inferred as a major regulator in both the regenerative and feeding responses. Additionally, XBP1 and INSR, major regulatory molecules within the Unfolded Protein Response and Insulin Receptor Signaling pathways, respectively, were inferred as URMs in analyses of distinct DE gene sets associated with both the regenerative and feeding responses, indicating a potentially expanded regulatory role of these URMs during regenerative growth.

Broadly, our results highlight shared patterns of signaling activity between divergent regenerating species and raise questions about the number of times this regenerative response may have evolved in snakes and the degree to which aspects of the regenerative response may be driven by shared ancestral regulatory programs versus convergent evolution of regulatory programs in divergent snake lineages. While convergent evolution of complex signaling programs may seem unlikely, large-scale metabolic adaptation and convergent evolution has been demonstrated previously in snakes, and thus cannot be readily discounted as an explanation for the phylogenetic dispersion of regenerative growth phenotypes and the regulatory pathways that underlie these phenotypes (Castoe et al., 2008, 2009).

Activation of stress response signaling during regeneration

Oxidative and other cellular stresses are known to impair tissue repair and regeneration in vertebrates (Beyer and Werner, 2008; Schäfer and Werner, 2008; Sen and Roy, 2008), and links between regulation of stress-responses and regeneration are beginning to emerge in the literature (Puente et al., 2014; Rees et al., 2001; Roesner et al., 2006). In rats, the transition from an oxygen-poor pre-natal environment to an oxygen-rich post-natal environment corresponds with a cessation of regenerative capacity in heart tissue due to induced DNA damage inflicted by increased oxidative stress (Puente et al., 2014). Additionally, one of the most well studied vertebrate systems of tissue regeneration is the zebrafish, which inhabits a hypoxic aquatic environment and thus experiences a lesser degree of oxidative stress during regenerative growth (Puente et al., 2014; Rees et al., 2001; Roesner et al., 2006). A previous study focused on post-feeding organ regenerative response in the Burmese Python identified the NRF2-mediated oxidative stress response pathway as having the greatest upregulation in activity of all inferred pathways in the intestine, kidney, and liver (Andrew et al., 2017).

. Given the rapid increases in metabolism (up to 44-fold; Secor and Diamond, 1998) and cell proliferation following feeding in regenerating snake species and previous evidence for a role of stress responses in python organ regeneration, it is logical that a coordinated and highly-activated armada of stress response pathways may play a role in the extreme regenerative growth observed in some snakes.

The NRF2-mediated oxidative stress response pathway was inferred to be a regulator in both the general feeding and regenerative responses. Thus, the NRF2 pathway is likely associated with feeding regardless of regeneration phenotype, but may play a more broad and highly stimulated role during regeneration in these two species. Mitigation of oxidative stress by NRF2 has been shown to play a vital role in liver tissue repair in mice by preventing insulin and insulin growth factor 1 (IGF-1) resistance that occurs via the phosphorylation of insulin receptor substrates by serine/threonine kinases that are activated by oxidative stress (Aguirre et al., 2002; Beyer and Werner, 2008; Kamata et al., 2005). In NRF2-deficient mice, insulin resistance prevents the insulin signaling pathway from properly activating PI3K/AKT and MAPK signaling pathways, two major pathways of growth and anti-apoptotic signaling, thus impairing tissue growth and repair (Beyer and Werner, 2008). Our results, together with the emerging role of NRF2 in regeneration, suggest that the action of NRF2 may play an important role in facilitating regenerative growth by permitting activity of growth mechanisms that otherwise negatively respond to oxidative stress.

The Unfolded Protein Response (UPR), which senses and mitigates endoplasmic reticulum (ER) stress, was an inferred regulator of the general feeding response and is likely involved in mitigating ER stress associated with the high degree of cell turnover, exposure to metabolites and toxins, and general secretory nature of digesting intestine tissue (Kaser et al., 2013). While the entire UPR pathway was not inferred to be a regulator of the regeneration, XBP1, a major regulatory molecule within the IRE1-XBP1 signaling cascade of the UPR (Smith et al., 2011), was inferred to be an active regulator in both feeding and regenerative responses. XBP1 has been identified as an important factor in preventing tumor formation during regeneration of intestinal epithelial tissue following injury in mice (Niederreiter et al., 2013), and the broad activation of XBP1 signaling may serve a similar role during regeneration in the python and rattlesnake, although further study would be necessary to confirm the role of this regulatory cascade in the regenerative response.

Our analyses suggest that pathways associated with DNA damage responses are uniquely involved in the regenerative response in the python and rattlesnake. This apparent involvement of a DNA damage response is likely to play a role in facilitating the high degree of cell proliferation required for rapid tissue growth. The involvement of these DNA damage response mechanisms, and particularly those associated with tumor suppression, are intriguing given that snakes, and reptiles in general, exhibit lower incidences of cancer than mammals (Effron et al., 1977). Future studies into the specific means by which snakes activate DNA damage responses during regenerative growth may provide new insight into tumor suppression mechanisms in vertebrates.

Insight into the regulation of regeneration from proteomic analyses

Our integrated analysis of transcriptomic and proteomic data provides complementary support for a number of key inferences regarding mechanisms and activation of signaling networks. Core Analyses in IPA based on shifts in protein abundance between fasted and 4DPF produced broadly similar inferences of URM signaling as did analyses of DE genes between fasted and 1DPF, including consistent activity of stress and growth URMs such as NFE2L2, KRAS, EGF, and others. The lag time between transcriptomic and proteomic responses, together with the rapid response time of regenerative phenotypes also suggests that other means of regulation, such as post-translational modification of proteins, are likely also important in directing signaling that underlies the regenerative response. Future work to explore the role of post-translational modifications in the early phases of regenerative growth in snakes would provide an important dimension to our understanding of signaling that initiates regeneration.

A model for the regulation of regenerative growth in snakes

We generated a model for signaling underlying regenerative intestinal growth in snakes based on inferences of regulatory mechanisms from this study and documented interactions among these mechanisms in other vertebrates (Fig. 3C). In this model, growth signaling pathways are activated by circulating signal molecules, such as insulin or other growth factors, with some of these signals potentially integrated via cell junction signaling in intestinal epithelial cells (Perez-Moreno et al., 2003). As growth signaling promotes cellular growth and proliferation, the buildup of reactive oxygen species and ER stress activate stress response pathways including NRF2-mediated oxidative stress response and the Unfolded Protein Response (Beyer and Werner, 2008; Kaser et al., 2013), which in turn act to mitigate stress and prevent the cessation of growth signaling (Beyer and Werner, 2008; Hirosumi et al., 2002). In response to initial and/or constitutive increases in cellular stress, DNA damage response pathways are also activated to ensure proper replication of cells and promote cell survival during proliferation (Barash et al., 2010; Puente et al., 2014). Although preliminary, this model provides a hypothesis that can be further tested with additional analyses and experiments, and as such, presents a valuable step toward understanding how extreme bouts of regeneration might be accomplished in vertebrates.

Conclusions

Major advances in genomics have enabled the development of new vertebrate model systems that have traditionally lacked genomic resources but possess interesting phenotypes. Snakes are an example of such a system, and new genomic resources now allow for intensive study of their extreme and medically relevant phenotypes, including regenerative growth following feeding (Castoe et al., 2013; Perry et al., 2018; Schield et al., 2019). By studying multiple species of snakes that do and do not experience regenerative growth upon feeding, we were able to begin to identify signaling mechanisms that may underlie extreme intestinal regeneration in snakes and distinguish these from mechanisms that are instead associated with a feeding response. Our findings highlight the value of employing a comparative approach to dissect a complex physiological response, and suggest that a combination of mechanisms uniquely activated in regenerating species and mechanisms shared with a typical feeding response, but regulating a greater number or distinct set of genes, may drive regenerative intestinal growth in snakes. We developed a hypothesis for how growth and stress response pathways might coordinate extreme intestinal regenerative growth while managing cellular stress and DNA damage associated with the extreme nature of this growth (e.g., 100% increases in mass in 24 hours in pythons; Secor and Diamond, 1995). Our inference suggests that extreme regenerative growth in snake requires the coordination of stress response, DNA damage response, and pro-survival signaling in addition to growth signaling. Testing and validating the precise role of these pathways and interactions among them is a priority for future studies and may enable further insight into regenerative signaling mechanisms with therapeutic potential for treating human conditions ranging from digestive diseases to cancer. From an evolutionary perspective, our findings raise interesting questions regarding the evolution of the regenerative response among snakes and pose further questions about how this phenotype may have influenced (or been driven by) major features of snake ecology. Considering the diversity of snakes, our analyses also beg the question of how broadly the three study species characterize the dichotomy between those that undergo regeneration upon feeding and those that do not, and future studies incorporating a greater diversity of species will be valuable for testing the generalizability of our conclusions across different snake lineages.

Acknowledgments

We thank the Minnesota Supercomputing Institute and the Texas Advanced Computing Center for access to computational resources. Support for this work was provided by National Science Foundation (NSF) grants IOS-655735 to TAC and SMC, DEB-1655571 to TAC and SPM, and IOS-0466139 and IOS-1656138 to SMS.

Data Accessibility

Time series transcriptomic data for *C. viridis* and *N. rhombifer* intestine tissues are available from the NCBI Short Read Archive at NCBI: SRP200900, as are previously generated data for P. m. bivittatus (NCBI: SRP051827). Label-free quantitative proteomics data are available from the Dryad Digital Repository (doi:10.5061/dryad.db660b8). Relevant scripts and code are available from GitHub (https://github.com/blairperry/3snake-RegenerativeGrowth).

Figures

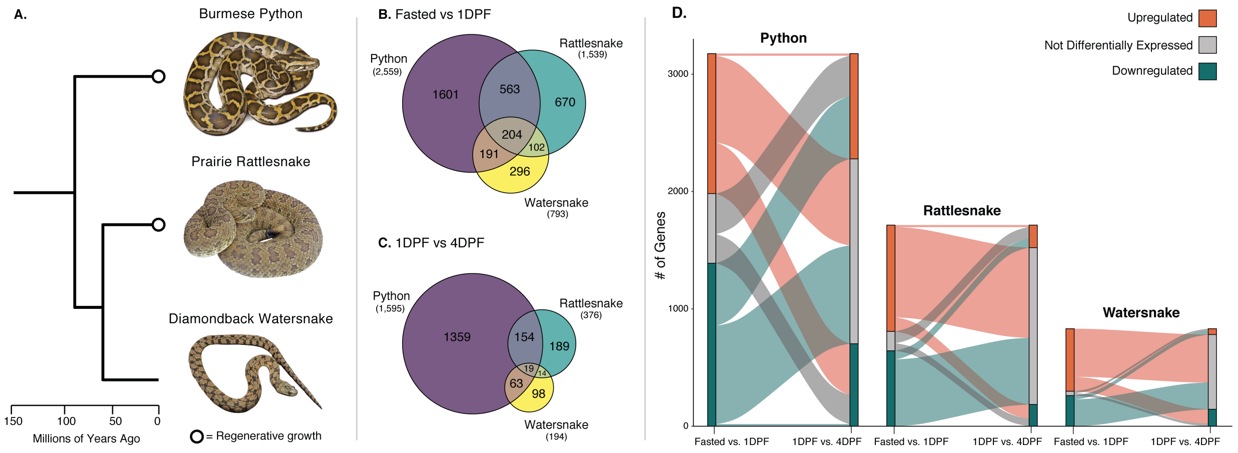


Figure 1. Divergent species that experience post-feeding regenerative growth exhibit similar gene expression responses. A) The Burmese Python and Prairie Rattlesnake both exhibit regenerative organ growth after feeding, despite being separate by roughly 90 million years of divergence. B-C) Venn diagrams of differentially expressed (DE) genes in the Burmese Python, Prairie Rattlesnake, and Diamondback Watersnake in pairwise comparisons between B) Fasted and 1DPF and C) 1DPF and 4DPF. D) Alluvial plots summarizing the number of upregulated ($p < 0.05$), downregulated ($p < 0.05$), and not differentially expressed ($p > 0.05$) genes for fasted vs. 1DPF and 1DPF vs. 4DPF pairwise comparisons in the Burmese Python, Prairie Rattlesnake, and Diamondback Watersnake. Ribbon width represents the number of genes exhibiting a specified pattern of expression across the two pairwise comparisons (i.e. upregulated in fasted vs. 1DPF and downregulated in 1DPF vs. 4DPF). Genes that were not DE in both pairwise comparisons are not shown.

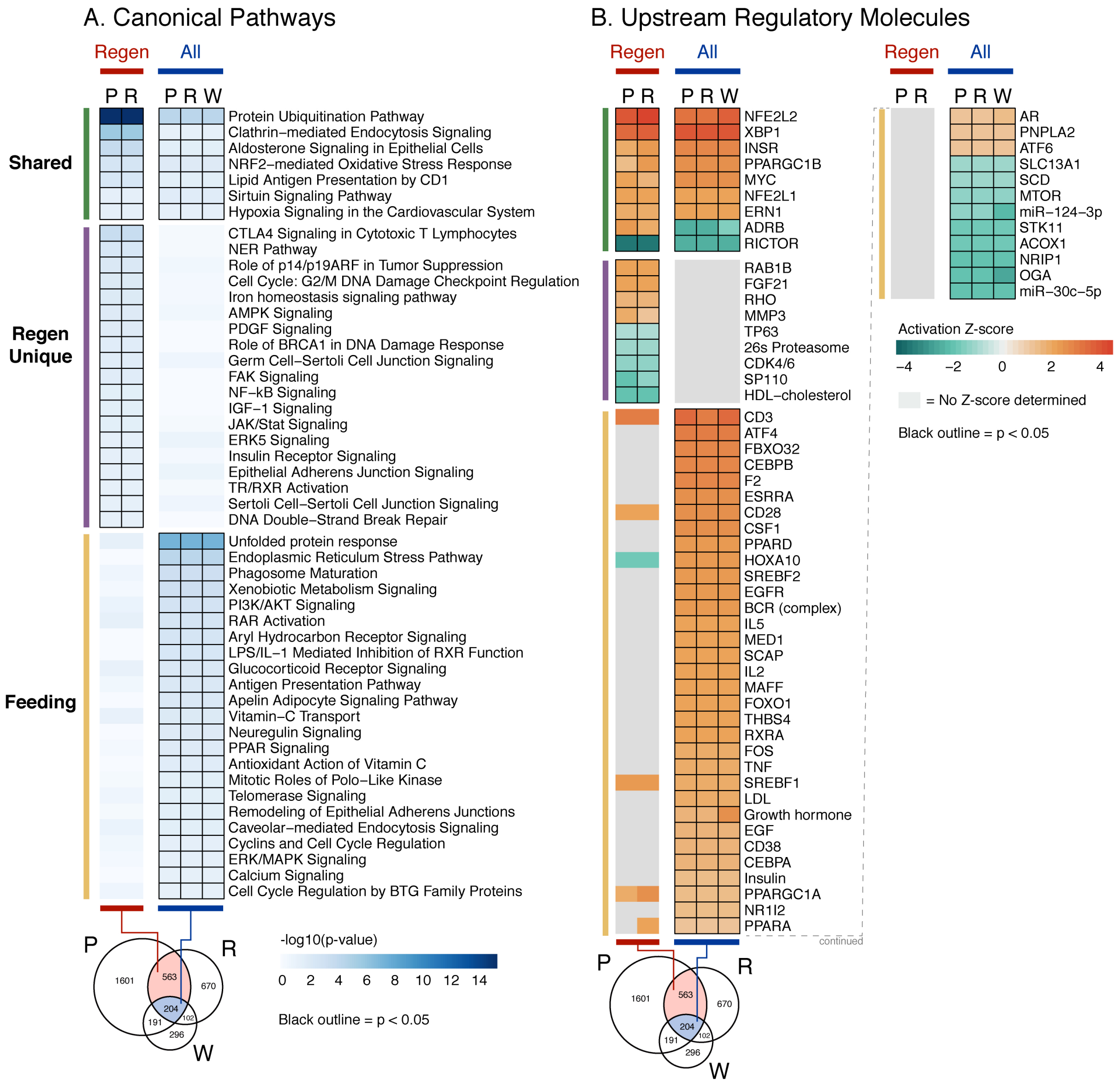


Figure 2. Canonical pathway and upstream regulatory molecule activation inferences based on comparisons of fasted versus 1DPF RNAseq data. A) Canonical pathways enrichment of differentially expressed genes between fasted and 1DPF based on genes shared uniquely between the two regenerating species ("Regen," left column) and genes shared between all three species ("All," right column). Black outlines denote a $p < 0.05$. B) Predicted upstream regulatory molecule activity based on Regen and All gene sets. Cells with a black outline indicate significant enrichment and predicted activity ($p < 0.05$ and $|z| > 1$).

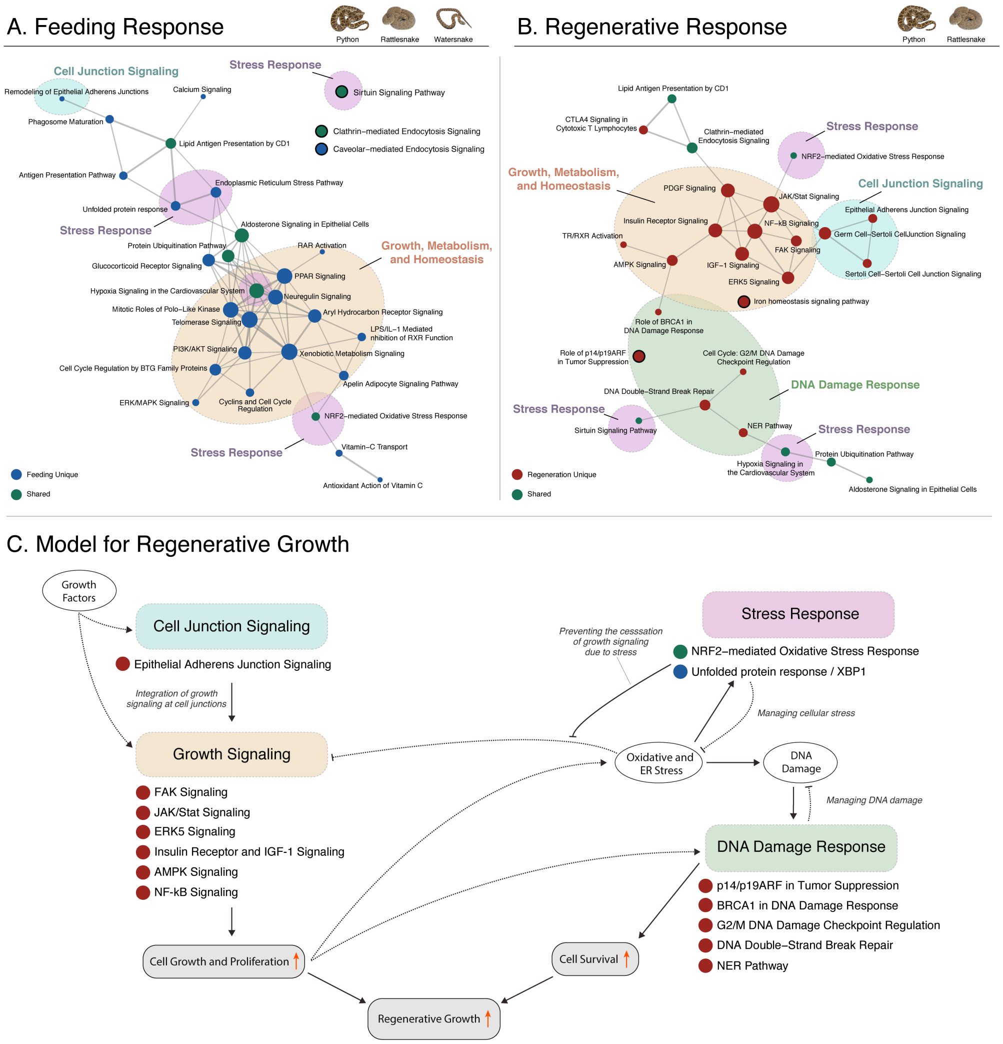

A. Feeding Response
Python
Rattlesnake
Watersnake
Cell Junction Signaling
Stress Response
Remodeling of Epithelial Adherens Junctions
Calcium Signaling
Sirtuin Signaling Pathway
Phagosome Maturation
Clathrin–mediated Endocytosis Signaling
Caveolar–mediated Endocytosis Signaling
Lipid Antigen Presentation by CD1
Antigen Presentation Pathway
Endoplasmic Reticulum Stress Pathway
Unfolded protein response
Stress Response
Aldosterone Signaling in Epithelial Cells
Protein Ubiquitination Pathway
RAR Activation
Growth, Metabolism, and Homeostasis
Glucocorticoid Receptor Signaling
PPAR Signaling
Hypoxia Signaling in the Cardiovascular System
Neuregulin Signaling
Mitotic Roles of Polo–Like Kinase
Aryl Hydrocarbon Receptor Signaling
Telomerase Signaling
LPS/IL–1 Mediated nhibition of RXR Function
PI3K/AKT Signaling
Xenobiotic Metabolism Signaling
Cell Cycle Regulation by BTG Family Proteins
Apelin Adipocyte Signaling Pathway
ERK/MAPK Signaling
Cyclins and Cell Cycle Regulation
NRF2–mediated Oxidative Stress Response
Stress Response
Vitamin–C Transport
Antioxidant Action of Vitamin C
Feeding Unique
Shared
B. Regenerative Response
Python
Rattlesnake
Lipid Antigen Presentation by CD1
CTLA4 Signaling in Cytotoxic T Lymphocytes
Clathrin–mediated Endocytosis Signaling
Stress Response
NRF2–mediated Oxidative Stress Response
Growth, Metabolism, and Homeostasis
Cell Junction Signaling
PDGF Signaling
JAK/Stat Signaling
Epithelial Adherens Junction Signaling
Insulin Receptor Signaling
NF–kB Signaling
Germ Cell–Sertoli CellJunction Signaling
TR/RXR Activation
FAK Signaling
AMPK Signaling
IGF–1 Signaling
Sertoli Cell–Sertoli Cell Junction Signaling
ERK5 Signaling
Iron homeostasis signaling pathway
Role of BRCA1 in DNA Damage Response
Role of p14/p19ARF in Tumor Suppression
Cell Cycle: G2/M DNA Damage Checkpoint Regulation
DNA Double–Strand Break Repair
DNA Damage Response
Stress Response
Sirtuin Signaling Pathway
NER Pathway
Stress Response
Hypoxia Signaling in the Cardiovascular System
Protein Ubiquitination Pathway
Aldosterone Signaling in Epithelial Cells
Regeneration Unique
Shared
C. Model for Regenerative Growth
Growth Factors
Cell Junction Signaling
Epithelial Adherens Junction Signaling
Integration of growth signaling at cell junctions
Growth Signaling
FAK Signaling
JAK/Stat Signaling
ERK5 Signaling
Insulin Receptor and IGF-1 Signaling
AMPK Signaling
NF-kB Signaling
Cell Growth and Proliferation
Regenerative Growth
Cell Survival
Preventing the cesssation of growth signaling due to stress
Stress Response
NRF2–mediated Oxidative Stress Response
Unfolded protein response / XBP1
Managing cellular stress
Oxidative and ER Stress
DNA Damage
Managing DNA damage
DNA Damage Response
p14/p19ARF in Tumor Suppression
BRCA1 in DNA Damage Response
G2/M DNA Damage Checkpoint Regulation
DNA Double–Strand Break Repair
NER Pathway

Figure 3. Overlapping canonical pathway predictions characterizing regenerative and feeding responses and a hypothesis model for regenerative growth in snakes. Networks showing the overlap in genes underlying canonical pathways with predicted activity from analyses of A) DE genes shared between all three species and B) DE genes shared between the python and rattlesnake but not the watersnake. A connection between two pathways indicates that at least 50% of the genes underlying the significant prediction of activity in one of the pathways also underlie the prediction of the other pathway, whereas pathways that are not connected to any others (circles with grey outlines) do not share > 50% of the genes underlying their prediction with any other pathway. Dotted circles represent manual annotation of pathways with similar functions. C) A hypothesis model for how the integration of growth and stress response signaling drive regenerative growth in snakes.

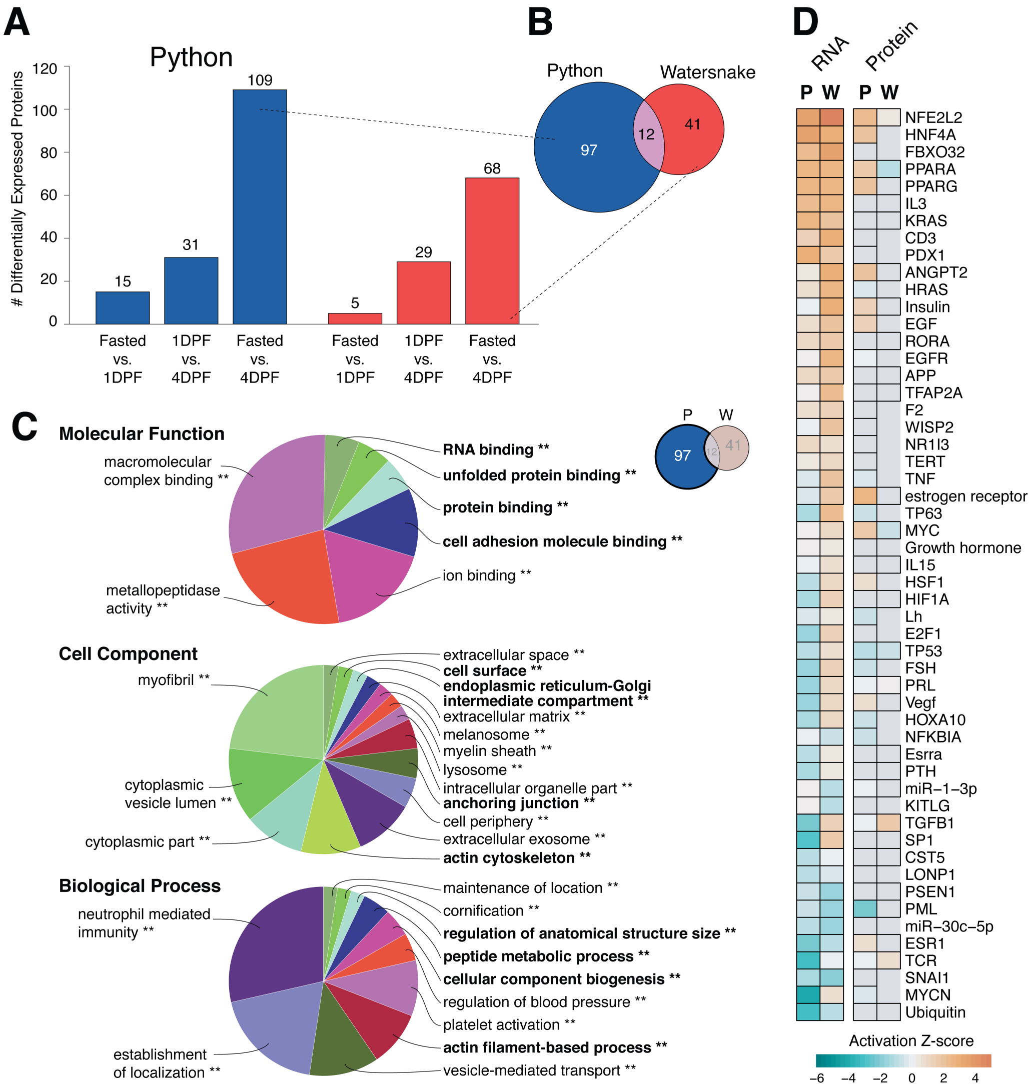
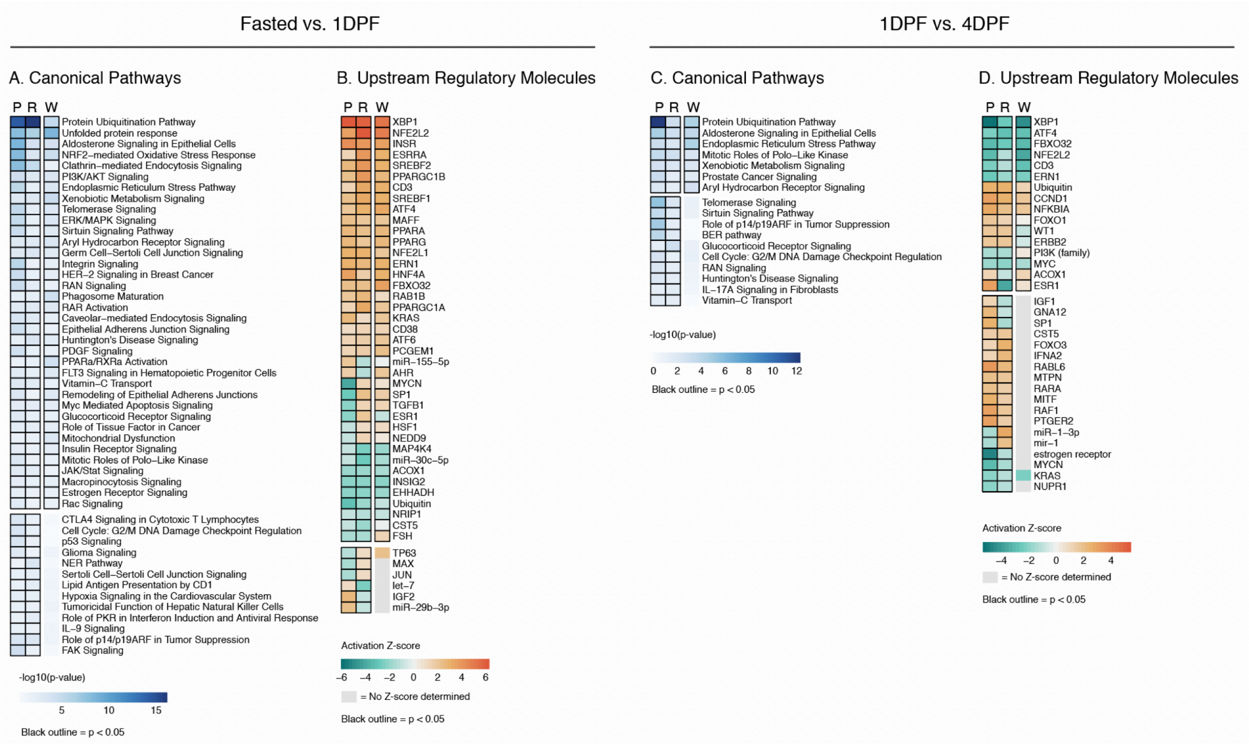
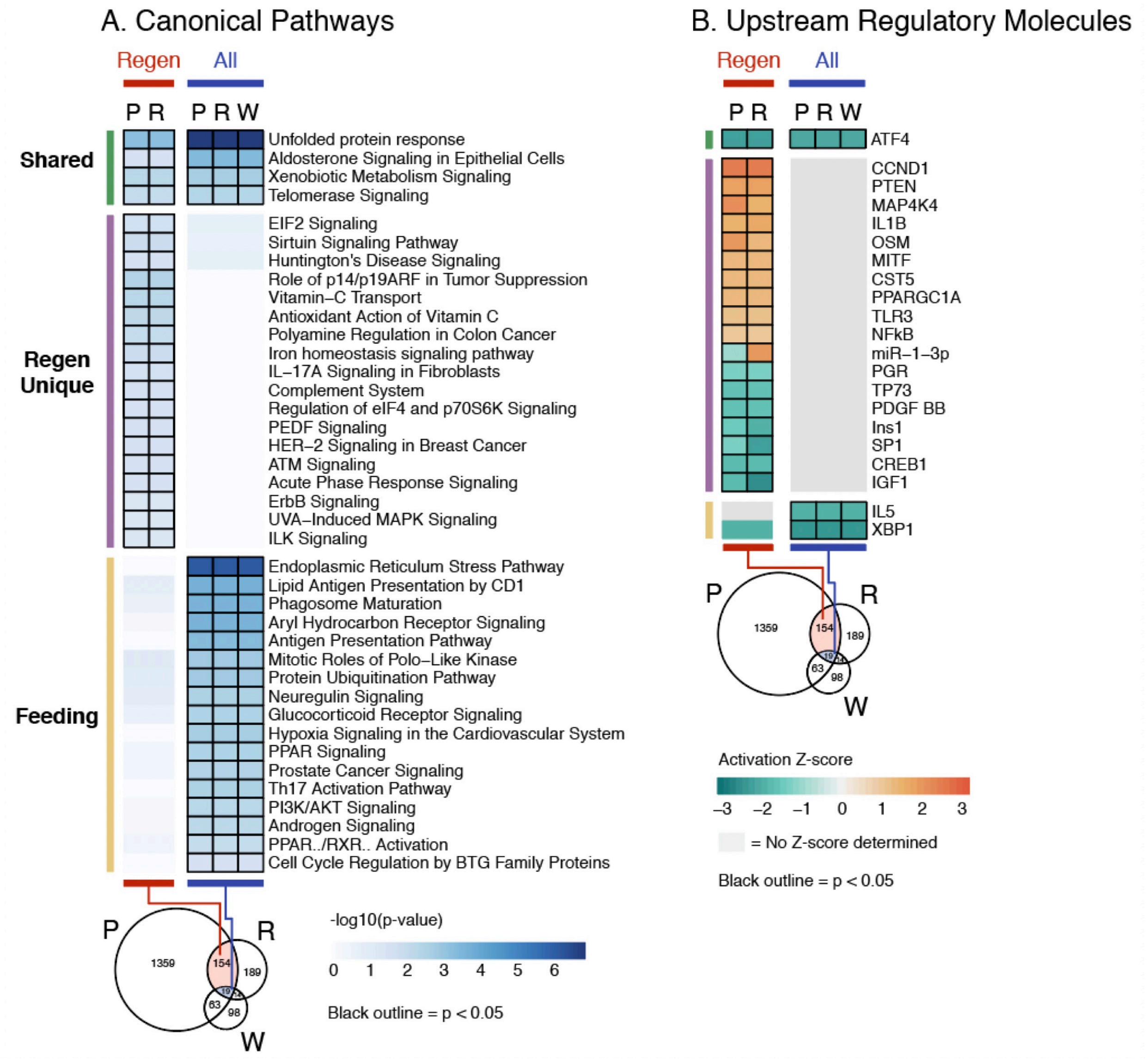


Figure 4. Proteomic comparison of python and watersnake intestine following feeding. A) Numbers of proteins that exhibited significant changes in abundance in pairwise comparisons. B) Venn diagram of proteins showing significant changes in abundance between fasted and 4DPF in the python and watersnake. C) GO term characterization of proteins with significant changes in abundance between fasted and 4DPF in the python only. Asterisks denote significant enrichment of a category ($p < 0.05$), and terms with likely involvement in regeneration phenotypes are bolded. D) Upstream regulatory molecule (URM) activity inferred from significant changes in protein abundance between fasted and 4DPF ($p < 0.1$), and significant DE genes between fasted and 1DPF ($p < 0.05$). Only URMs with significantly inferred activity in the python from both protein and RNA data are shown.

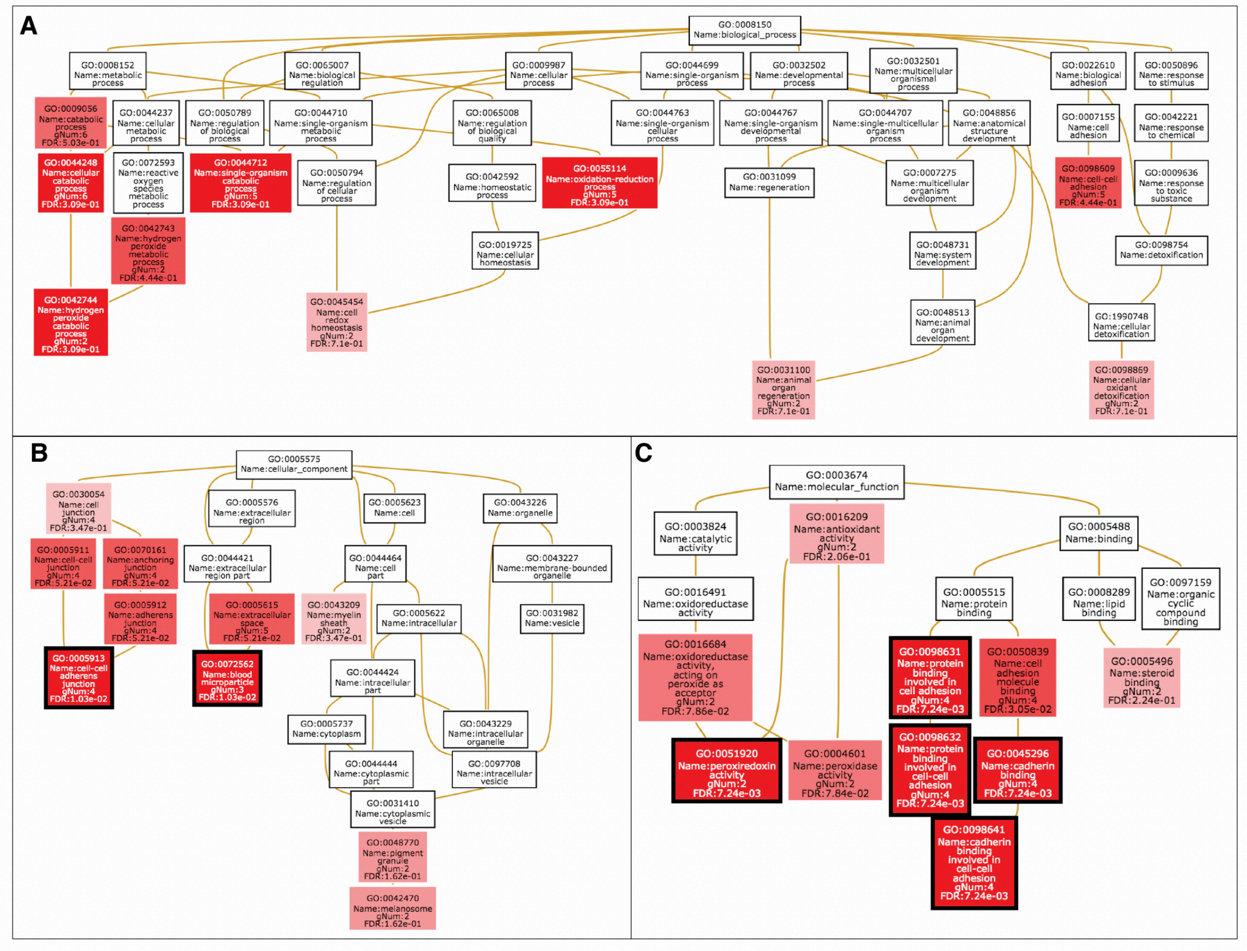
Supplementary Figures



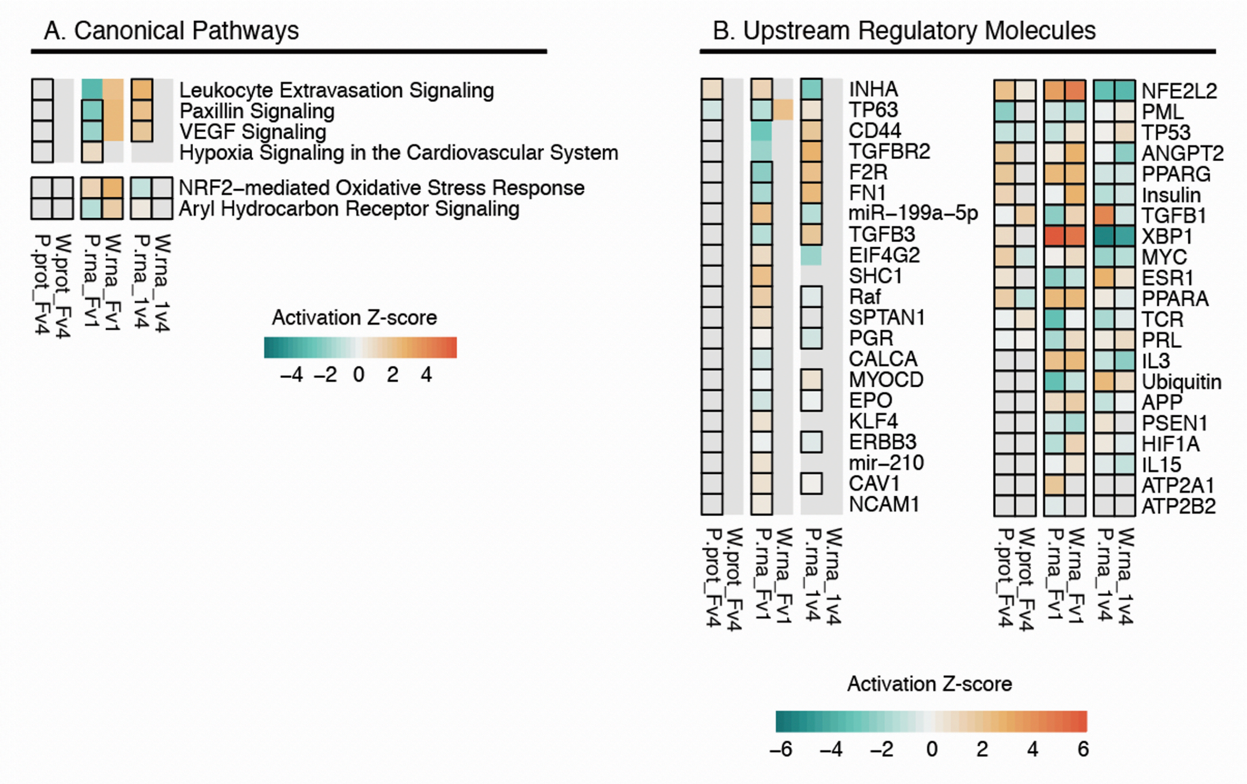
Supplementary Figure 1. Canonical pathway and upstream regulatory molecule activation inferences based on comparisons of all differentially expressed genes. A) Canonical pathways enrichment of differentially expressed genes between fasted and 1DPF. Black outlines denote a p < 0.05. B) Predicted upstream regulatory molecule activity between fasted and 1DPF. Cells with a black outline indicate significant enrichment and predicted activity (p < 0.05 and |z| > 1). C) Canonical pathways enrichment of differentially expressed genes between 1DPFand 4DPF. Black outlines denote a p < 0.05. D) Predicted upstream regulatory molecule activity between 1DPFand 4DPF. Cells with a black outline indicate significant enrichment and predicted activity (p < 0.05 and |z| > 1). For each heatmap, only pathways/regulatory molecules that are either significant in all three species or just the two regenerating species (python and rattlesnake) are shown.



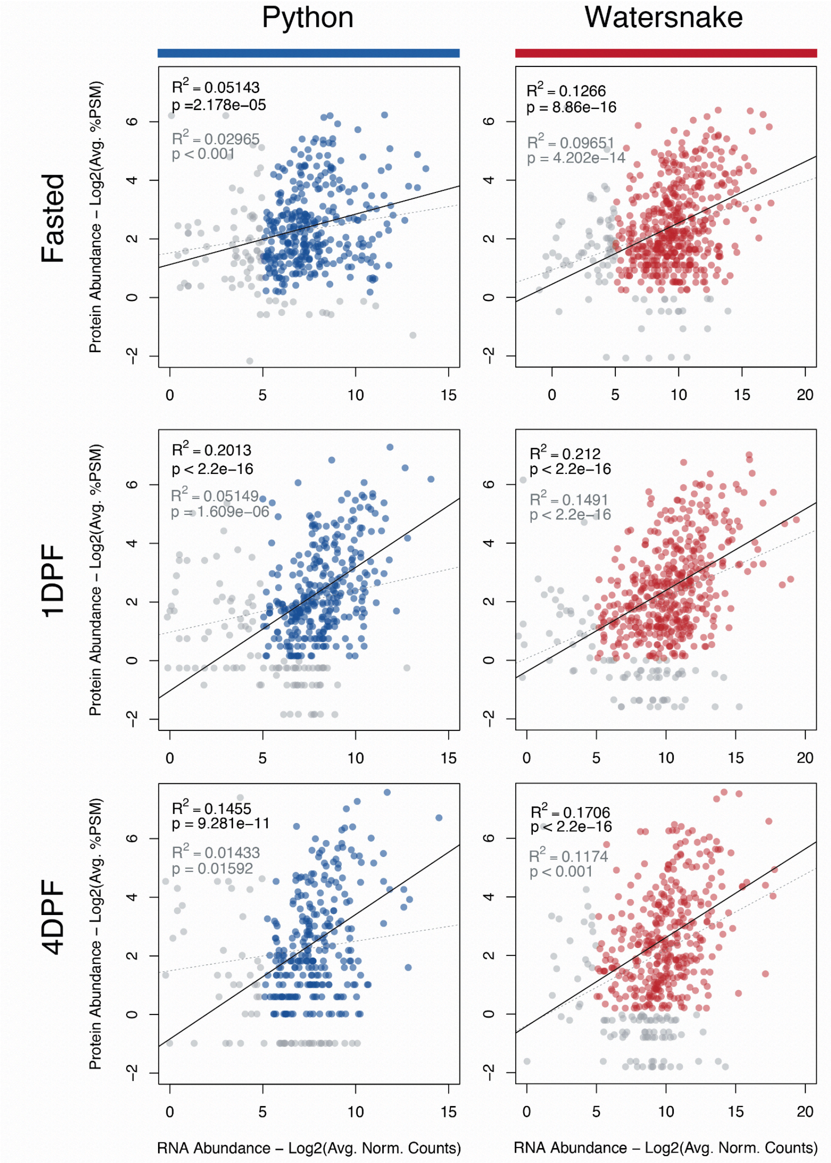
Supplementary Figure 2. Canonical pathway and upstream regulatory molecule activation inferences based on comparisons of 1DPF versus 4DPF RNAseq data. A) Canonical pathways enrichment of differentially expressed genes between fasted and 1DPF based on genes shared uniquely between the the two regenerating species ("Regen," left column) and genes shared between all three species ("All," right column). Black outlines denote a p < 0.05. B) Predicted upstream regulatory molecule activity based on Regen and All gene sets. Cells with a black outline indicate significant enrichment and predicted activity (p < 0.05 and |z| > 1).


A
GO:0008150 Name:biological_process
GO:0008152 Name:metabolic process
GO:0065007 Name:biological regulation
GO:0009987 Name:cellular process
GO:0044699 Name:single-organism process
GO:0032502 Name:developmental process
GO:0032501 Name:multicellular organismal process
GO:0022610 Name:biological adhesion
GO:0050896 Name:response to stimulus
GO:0009056 Name:catabolic process gNum:6 FDR:5.03e-01
GO:0044237 Name:cellular metabolic process
GO:0050789 Name:regulation of biological process
GO:0044710 Name:single-organism metabolic process
GO:0065008 Name:regulation of biological quality
GO:0044763 Name:single-organism cellular process
GO:0044767 Name:single-organism developmental process
GO:0044707 Name:single-multicellular organism process
GO:0048856 Name:anatomical structure development
GO:0007155 Name:cell adhesion
GO:0042221 Name:response to chemical
GO:0044248 Name:cellular catabolic process gNum:6 FDR:3.09e-01
GO:0072593 Name:reactive oxygen species metabolic process
GO:0044712 Name:single-organism catabolic process gNum:5 FDR:3.09e-01
GO:0050794 Name:regulation of cellular process
GO:0042592 Name:homeostatic process
GO:0055114 Name:oxidation-reduction process gNum:5 FDR:3.09e-01
GO:0031099 Name:regeneration
GO:0007275 Name:multicellular organism development
GO:0098609 Name:cell-cell adhesion gNum:5 FDR:4.44e-01
GO:0009636 Name:response to toxic substance
GO:0042743 Name:hydrogen peroxide metabolic process gNum:2 FDR:4.44e-01
GO:0019725 Name:cellular homeostasis
GO:0048731 Name:system development
GO:0098754 Name:detoxification
GO:0042744 Name:hydrogen peroxide catabolic process gNum:2 FDR:3.09e-01
GO:0045454 Name:cell redox homeostasis gNum:2 FDR:7.1e-01
GO:0048513 Name:animal organ development
GO:1990748 Name:cellular detoxification
GO:0031100 Name:animal organ regeneration gNum:2 FDR:7.1e-01
GO:0098869 Name:cellular oxidant detoxification gNum:2 FDR:7.1e-01
B
GO:0005575 Name:cellular_component
GO:0030054 Name:cell junction gNum:4 FDR:3.47e-01
GO:0005576 Name:extracellular region
GO:0005623 Name:cell
GO:0043226 Name:organelle
GO:0005911 Name:cell-cell junction gNum:4 FDR:5.21e-02
GO:0070161 Name:anchoring junction gNum:4 FDR:5.21e-02
GO:0044421 Name:extracellular region part
GO:0044464 Name:cell part
GO:0043227 Name:membrane-bounded organelle
GO:0005912 Name:adherens junction gNum:4 FDR:5.21e-02
GO:0005615 Name:extracellular space gNum:5 FDR:5.21e-02
GO:0043209 Name:myelin sheath gNum:2 FDR:3.47e-01
GO:0005622 Name:intracellular
GO:0031982 Name:vesicle
GO:0005913 Name:cell-cell adherens junction gNum:4 FDR:1.03e-02
GO:0072562 Name:blood microparticle gNum:3 FDR:1.03e-02
GO:0044424 Name:intracellular part
GO:0005737 Name:cytoplasm
GO:0043229 Name:intracellular organelle
GO:0044444 Name:cytoplasmic part
GO:0097708 Name:intracellular vesicle
GO:0031410 Name:cytoplasmic vesicle
GO:0048770 Name:pigment granule gNum:2 FDR:1.62e-01
GO:0042470 Name:melanosome gNum:2 FDR:1.62e-01
C
GO:0003674 Name:molecular_function
GO:0003824 Name:catalytic activity
GO:0016209 Name:antioxidant activity gNum:2 FDR:2.06e-01
GO:0005488 Name:binding
GO:0016491 Name:oxidoreductase activity
GO:0005515 Name:protein binding
GO:0008289 Name:lipid binding
GO:0097159 Name:organic cyclic compound binding
GO:0016684 Name:oxidoreductase activity, acting on peroxide as acceptor gNum:2 FDR:7.86e-02
GO:0098631 Name:protein binding involved in cell adhesion gNum:4 FDR:7.24e-03
GO:0050839 Name:cell adhesion molecule binding gNum:4 FDR:3.05e-02
GO:0005496 Name:steroid binding gNum:2 FDR:2.24e-01
GO:0051920 Name:peroxiredoxin activity gNum:2 FDR:7.24e-03
GO:0004601 Name:peroxidase activity gNum:2 FDR:7.84e-02
GO:0098632 Name:protein binding involved in cell-cell adhesion gNum:4 FDR:7.24e-03
GO:0045296 Name:cadherin binding gNum:4 FDR:7.24e-03
GO:0098641 Name:cadherin binding involved in cell-cell adhesion gNum:4 FDR:7.24e-03

Supplementary Figure 3. GO term overrepresentation for the 12 proteins that exhibited significant changes in abundance between fasting and feeding in both the python and watersnake. A) Biological process. B) Cellular component. C) Molecular function. Boxes with black outline denote significant overrepresentation (FDR < 0.05).



Supplementary Figure 4. Canonical pathway and upstream regulatory molecule activation inferences based on proteomic data. Comparison of predicted activity of A) canonical pathways and B) upstream regulatory molecules based on proteins with significant changes in abundance between fasted and 4DPF and genes significantly differentially expressed between fasted vs. 1DPF and 1DPF vs. 4DPF, filtered to only show pathways and molecules that are significant from both protein and RNA data but only in python, or significant based on both protein and RNA data in both python and watersnake. Cells with a black outline indicate significant predicted activity ($p < 0.05$).



Supplementary Figure 5. Correlation of RNA and protein abundance in the python and watersnake. Grey R^2 values, p-values, and trend lines correspond to all data points (grey and colored), while black values and trend lines correspond to only data points in which average log2 RNA and protein abundance are greater than 5 and 0, respectively (colored points).

Supplementary Tables

Supplementary Table 1. Sample and sequencing information for rattlesnake and watersnake small intestine gene expression data generated for this study.

Species	Time point	Animal ID	Instrument	cDNA Prep Kit	SRA Accession
Prairie Rattlesnake (*Crotalus viridis viridis*)	Fasted	CV1	HiSeq	NEB Next	SAMN12003898
	Fasted	CV2	HiSeq	NEB Next	SAMN12003899
	Fasted	CV4	HiSeq	NEB Next	SAMN12003900
	Fasted	CV7	HiSeq	NEB Next	SAMN12003901
	1DPF	CV3	HiSeq	NEB Next	SAMN12003902
	1DPF	CV6A	HiSeq	NEB Next	SAMN12003903
	1DPF	CV8	HiSeq	NEB Next	SAMN12003904
	4DPF	CV9	HiSeq	NEB Next	SAMN12003905
	4DPF	CV10	HiSeq	NEB Next	SAMN12003906
	4DPF	CV11	HiSeq	NEB Next	SAMN12003907
	4DPF	CV12	HiSeq	NEB Next	SAMN12003908
Diamondback Watersnake (*Nerodia rhombifer*)	Fasted	NR1316	HiSeq	NEB Next	SAMN12003909
	Fasted	NR1317	HiSeq	NEB Next	SAMN12003910
	Fasted	NR1464	HiSeq	NEB Next	SAMN12003911
	Fasted	NR1416	HiSeq	NEB Next	SAMN12003912
	1DPF	NR1357	HiSeq	NEB Next	SAMN12003913
	1DPF	NR1436	HiSeq	NEB Next	SAMN12003914
	1DPF	NR1388	HiSeq	NEB Next	SAMN12003915
	1DPF	NR1331	HiSeq	NEB Next	SAMN12003916
	4DPF	NR1442	HiSeq	NEB Next	SAMN12003917
	4DPF	NR1354	HiSeq	NEB Next	SAMN12003918
	4DPF	NR1327	HiSeq	NEB Next	SAMN12003919
	4DPF	NR1324	HiSeq	NEB Next	SAMN12003920

Supplementary Table 2. Read mapping statistics for all samples used in the study.

Species	Time Point	Sample ID	Number Input Reads	Number Uniquely Mapped Reads	% Uniquely Mapped Reads
Python	Fasted	U25	11,140,509	7,845,947	70.43%
		AI6a	3,409,861	2,898,451	85.00%
		AI6b	992,739	871,455	87.78%
		AJ6a	5,154,838	4,529,326	87.87%
		AJ6b	2,060,920	1,837,885	89.18%
		AJ6c	2,643,193	2,414,192	91.34%
	1DPF	W20	2,043,673	1,960,430	95.93%
		V43	5,656,165	4,193,004	74.13%
		Z14a	4,085,590	3,680,246	90.08%
		Z14b	1,678,087	1,527,872	91.05%
		Z14c	2,401,791	2,183,797	90.92%
	4DPF	Y18a	3,560,391	3,107,561	87.28%
		Y18b	1,364,726	1,203,620	88.19%
		Y18c	2,216,650	1,949,233	87.94%
		Y23a	2,971,652	2,551,637	85.87%
		Y23b	1,145,443	1,001,545	87.44%
		Y5a	3,689,468	3,256,951	88.28%
		Y5b	1,449,679	1,292,496	89.16%
Rattlesnake	Fasted	CV2	12,730,320	9,974,214	78.35%
		CV4	11,271,037	8,782,204	77.92%
		CV7	8,708,873	6,947,966	79.78%
		CV1	1,915,910	1,071,479	55.93%
	1DPF	CV6A	8,556,459	7,335,838	85.73%
		CV8	8,248,596	6,768,378	82.05%
		CV3	3,010,519	1,273,622	42.31%
	4DPF	CV11	9,610,524	8,174,577	85.06%
		CV12	8,926,893	7,465,208	83.63%
		CV9	11,594,658	10,156,595	87.60%
		CV10	2,318,138	1,607,507	69.34%
Watersnake	Fasted	NR1316	9,885,620	6,590,139	66.66%
		NR1317	9,239,753	6,384,714	69.10%
		NR1416	6,598,368	4,319,166	65.46%
		NR1464	11,832,553	7,814,975	66.05%
	1DPF	NR1331	6,486,641	3,561,965	54.91%
		NR1357	10,446,473	7,218,437	69.10%
		NR1388	9,144,085	5,465,528	59.77%
		NR1436	7,363,598	5,135,712	69.74%
	4DPF	NR1324	5,338,747	3,349,379	62.74%
		NR1327	19,008,775	8,448,405	44.44%
		NR1354	5,420,532	3,474,894	64.11%
		NR1442	8,042,516	5,501,847	68.41%

Supplementary Table 3. Relevant statistics pertaining to homolog assignment via OrthoMCL. The values in parentheses represent the number of groups containing a single human ID and multiple human IDs, respectively, in the homolog groups that were not full 1-to-1. See Supplementary Methods for additional details.

Total input genes per species:		
Human	20,182	
Python	17,985	
Rattlesnake	17,480	
Garter snake	17,524	
Total OrthoMCL homolog groups per species:		
Human	17,464	
Python	16,721	
Rattlesnake	14,513	
Garter snake	15,623	
OrthoMCL homolog groups containing all species:		
Total	9,553	
Full 1-to-1	7,151	
Not full 1-to-1	2,402	(832; 1,570)

Supplementary Table 4. Python and watersnake small intestine samples used to generate label-free quantitative proteomics data for this study. See (Andrew et al. 2015; Andrew et al. 2017) for additional information on python tissues.

Species	Time point	Animal ID
Burmese Python (*Python molurus bivittatus*)	Fasted	AI6
	Fasted	AJ6
	Fasted	U25
	1DPF	V43
	1DPF	S6
	1DPF	W20
	4DPF	Y5
	4DPF	Y18
Diamondback Watersnake (*Nerodia rhombifer*)	Fasted	NR1317
	Fasted	NR1464
	Fasted	NR1416
	1DPF	NR1436
	1DPF	NR1388
	1DPF	NR1331
	4DPF	NR1442
	4DPF	NR1354
	4DPF	NR1324

Chapter 4

PHYSIOLOGICAL DEMANDS AND SIGNALING ASSOCIATED WITH SNAKE VENOM PRODUCTION AND STORAGE ILLUSTRATED BY TRANSCRIPTIONAL ANALYSES OF VENOM GLANDS

Blair W. Perry[1], Drew R. Schield[1,2], Aundrea K. Westfall[1], Stephen P. Mackessy[3], & Todd A. Castoe[1]

[1]Department of Biology, 501 S. Nedderman Dr., The University of Texas Arlington, Arlington, TX, 76019, USA

[2]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA

[3]School of Biological Sciences, 501 20th Street, University of Northern Colorado, Greeley, CO 80639, USA

Abstract

Despite the extensive body of research on snake venom, many facets of snake venom systems, such as the physiology and regulation of the venom gland itself, remain virtually unstudied. Here, we use time series gene expression analyses of the rattlesnake venom gland in comparison with several non-venom tissues to characterize physiological and cellular processes associated with venom production and to highlight key distinctions of venom gland cellular and physiological function. We find consistent evidence for activation of stress response pathways in the venom gland, suggesting that mitigation of cellular stress is a crucial component of venom production. Additionally, we demonstrate evidence for an unappreciated degree of cellular and secretory activity in the steady state venom gland relative to other secretory tissues and identify vacuolar ATPases as the likely mechanisms driving acidification of the venom gland lumen during venom production and storage.

Introduction

The snake venom gland is an intriguing yet poorly understood system that holds broad potential as a model for studying the evolution of novel organ function and regulatory architecture, cellular responses to extreme physiological demands, and the production and storage of potent biological toxins. The emerging interest and potential utility of this system is emphasized by the recent development of snake venom gland organoids, an unprecedented resource for the controlled study of snake venom regulation, production, and general snake venom gland physiology and function (Post et al., 2020). However, despite an extensive body of literature focused on the products of snake venom glands (i.e. venoms and their components), and to a lesser degree the genes underlying snake venoms, little is known about the physiological, cellular, and molecular functionality of snake venom glands and how these compare to other secretory systems. The lack of a systems-level understanding of snake venom gland biology presents a major impediment to progress towards a comprehensive understanding of venom system evolution and function.

Our current understanding of venom gland function and physiology indicates that the process of venom production is impressively dynamic. Following the depletion of stored venom (i.e. via a predatory bite or manual venom extraction), the snake venom gland exhibits rapid and high-magnitude upregulation of venom gene transcription, venom protein production and processing, and secretion of venom components into the gland lumen (Carneiro et al., 1991; Currier et al., 2012; Kochva et al., 1980; Mackessy, 1991). Aspects of this process have been described through microscopy, proteomic, and gene expression ((Currier et al., 2012; Kerchove et al., 2004, 2008a; Mackessy and Baxter, 2006; Yamanouye et al., 2000), including some broad characterization of non-venom gene expression (Rokyta et al., 2012; Schield et al., 2019). While these studies have provided insight into venom gland function, their focus has been on processes immediately associated with venom production (i.e. regulation of venom genes and secretion of venom components). Accordingly, a systems-level understanding of the cellular processes that comprise the physiological environment in which venom production occurs remains incompletely characterized. For example, the high demands of gene regulation, protein processing and venom protein production places extreme demands and stress on venom gland epithelial cells. This would necessitate the activation of cellular stress response mechanisms to facilitate successful venom production while preventing damage to protein products, cells or the venom gland tissue. It is therefore expected that the extreme demands placed on venom gland tissue during venom production are associated with similarly extreme cellular physiology to accommodate extreme cellular and physiological performance, yet this has not been examined in previous studies.

In addition to unique physiology and functionality associated with the upregulation of venom production following venom depletion, the steady-state venom gland is tasked with housing an abundance of highly toxic venom components in a manner that protects the venom gland and surrounding tissue from the biological activity of venom components while keeping stored venom stable. Previous studies have shown that acidification of the venom bolus in the gland lumen inhibits venom enzymatic activity during storage and thus plays an important role in self-protection against harmful effects of venom and stabilization of venom proteins (Mackessy and Baxter, 2006). It has been proposed that this acidification is driven by populations of mitochondria-rich cells with morphological and histochemical features similar to parietal cells in mammals, which are responsible for secretion of gastric acid (Mackessy and Baxter, 2006). However, the exact molecular mechanisms underlying venom gland acidification remain unexplored.

In this study, we use gene expression data from multiple sampled timepoints from the venom gland of the Prairie Rattlesnake (*Crotalus viridis*) to facilitate the first detailed analysis of physiological and cellular pathways associated with the rapid and high-magnitude shifts in activity and function of the snake venom gland during venom production. These analyses include both broad characterization of regulatory pathways, molecules, and analysis of differentially expressed genes, as well as targeted dissection of cellular stress response mechanisms that may play an underappreciated role facilitating venom production. We also conduct comparisons of venom gland gene expression to that of multiple non-venom secretory tissues to highlight physiological and functional distinctiveness of the venom gland, including detailed analysis of molecular mechanisms driving venom gland lumen acidification.

Materials and Methods

Generation of mRNA-seq data

Venom gland tissue samples were generated previously (Schield et al., 2019)

. In brief, venom was manually expressed from one of the two venom glands of an adult male Prairie Rattlesnake, and the second venom gland was expressed 2 days later. One day later, the animal was humanely euthanized and venom gland tissues were dissected out and immediately snap frozen in liquid nitrogen. This process resulted in both a one day post-extraction (DPE) and three DPE venom gland tissue sample from the same animal. Unextracted venom gland, skin, pancreas, and stomach tissue was dissected from an additional male individual and frozen in liquid nitrogen. Total RNA was extracted from all tissue samples using Trizol Reagent (Invitrogen) and isopropanol. Four and two technical replicates were extracted for each venom gland and body-tissue treatment, respectively. Illumina mRNAseq libraries generated using poly-A selection and sequenced at Novogene on an Illumina NovaSeq platform using 150 bp paired-end reads.

mRNA-seq processing and pairwise analysis

Raw RNAseq data was quality-trimmed and filtered using Trimmomatic v0.33 (Bolger et al., 2014) and mapped to the Prairie Rattlesnake reference genome using STAR v2.5.2b (Dobin et al., 2013). Raw read counts were generated with featureCounts v1.6.3 (Liao et al., 2013). Count normalization and pairwise comparisons between unextracted and 1DPE venom gland, 1DPE and 3DPE venom gland, and body tissue and unextracted venom gland were conducted in DeSeq2 v1.26.0 (Love et al., 2014), and resulting p-values were corrected using independent hypothesis weighting (IHW) using baseMean from DeSeq2 as the covariate (Ignatiadis et al., 2016). Differentially expressed genes were defined as those with IHW p-value < 0.05.

Inferences of regulatory pathway ad molecular activity and analysis of overrepresented functional groups

Venom genes have been previously curated and annotated in the Prairie Rattlesnake reference genome (Schield et al., 2019), Annotated venom genes were excluded from subsequent analyses that focus on non-venom gene regulation. Differentially expressed non-venom genes were then assigned an orthologous human gene identifier using orthology tables generated previously (Perry et al., 2018). To infer broad patterns of regulatory pathway and molecule activity, differentially expressed genes were then analyzed using Core Analysis in Ingenuity Pathway Analysis (IPA) (Krämer et al., 2013). In the Core Analysis results, the following categories of canonical pathways were excluded: cancer, cardiovascular signaling, cellular immune response, disease-specific pathways, humoral immune response, Ingenuity Toxicity List Pathways, neurotransmitters and other nervous system signaling, pathogen-influenced signaling, and xenobiotic metabolism. Upstream regulatory molecule results were filtered by molecule type to include only genes, RNAs and proteins. Inferences of canonical pathway and upstream regulatory molecule activity with an overlap p-value < 0.05 and absolute activation z-score > 1 were considered significant. To characterize functional groups of differentially expressed genes further, gene ontology (GO) analyses were performed specifically on sets of genes upregulated in the unextracted venom gland relative to non-venom tissues, and for those upregulated in 1DPE relative to unextracted venom gland tissues. GO terms with significant overrepresentation in these gene sets were determined using the ClueGO plugin v2.5.6 (Bindea et al., 2009) for Cytoscape v3.7.2 (Shannon et al., 2003) using a right-sided hypergeometric test of enrichment with default p-value correction, using all genes that met DeSeq2 input cutoffs. Terms with a corrected p-value < 0.05 were considered significantly enriched. Networks of enriched GO terms were further manually characterized and grouped based on similarity of function, tissue, or cellular process.

Mechanisms of venom gland acidification

To investigate potential mechanisms driving venom gland acidification, we first compared gene expression for a set of candidate genes annotated with the "pH reduction" GO term (GO: 0045851) to identify genes with evidence of informative upregulation in the venom gland relative to other secretory tissues and/or during venom production.

To validate inferences related to the roles of H+/K+ versus vacuolar ATPases in driving venom gland acidification, we performed Western immunoblot analyses and immunohistochemical staining of gastric and venom gland membranes. Stomach and venom gland tissues were dissected from an adult Prairie Rattlesnake, and epithelial cells were harvested after removal of connective tissue and fascia. Epithelial tissues were then minced on an ice-chilled glass plate prior to homogenization in 3 ml 10 mM PIPES/tris buffer pH 7.4 with 2mM ethylenediaminetetraacetic acid (EDTA) and 2 mM ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid (EGTA) at 1500 rpm on ice. The remaining muscle and connective tissue was then removed by centrifugation at 3K rpm for 10 min at 4 oC. The resulting supernatant was layered onto 42% sucrose (w/v) in PIPES/tris buffer and overlaid with 5% sucrose. The samples were then centrifuged at 25K rpm for 90 min. at 4 oC in an SW28 swing rotor in a Beckman L8-70 ultracentrifuge. The membrane fraction, located at the interface of the 42 and 5% sucrose layers, was removed with a Pasteur pipette, transferred to a new centrifuge tube and topped with PIPES/tris buffer. Protein-containing membrane fractions were then pelleted by centrifuging at 34K rpm for 45 min at 4 oC. Following aspiration of the supernatant, the pellet was resuspended in tris/pipes buffer. Quantification of protein in membrane fractions was accomplished by a modified Lowry method utilizing the BCA Modified Lowry reagent from Promega (Madison, WI.). This material was then utilized (undiluted) for Western blots.

Thirty µl of each undiluted sample were run on a 7.5% acrylamide SDS-tricine reducing gel and transferred onto nitrocellulose membranes as described previously (Smith and Mackessy, 2016). Non-specific binding was blocked by incubating membranes in 10 ml 5% nonfat milk in PBS-tween20 (20% w/v) for 30 min. The blots were then incubated in 10 ml blocking solution with 1:2000 (v/v) primary antibody for 1 hr at RT; ATPAL1 was at 5.0 µg/µl and αH56 was in 100% mouse serum. The polyclonal ATPAL1 (designed against a C-terminal epitope of the gastric H+/K+-ATPase) and αH56 (designed against the 56kDa subunit of the vacuolar H+-ATPase) antibodies were used to test for the presence of H+/K+ ATPases and vacuolar ATPases, respectively, in gastric and venom gland membrane preparations (Granger et al., 2002; Mercier et al., 1993). Subsequently, the blots were washed for 3 x 15 min in PBS-tween and placed in 10 ml blocking solution with 1:20K secondary antibody. Following incubation at RT for 1 hr, the blots were again washed, incubated in 10 ml Supersignal West Pico chemiluminescent substrate solution (Pierce, Rockville, IL.) for 5 min at RT and exposed to high performance chemiluminescence film (Amersham International, Buckinghamshire, England).

Separately, main venom glands were fixed in 3.7% PBS-buffered formalin prior to being imbedded in paraffin wax and sectioned by microtome. The slides were dewaxed in xylene and rehydrated in an ethanol series, followed by PBS. To reduce background fluorescence during confocal microscopy, the antigen retrieval system (Dako Corp., Carpentaria, Ca.) was employed. The samples were then blocked with Dako Protein Block Serum and incubated overnight at 4 oC with 1:1000 mouse serum containing the αH56 polyclonal antibody. After washing 3x in 100 mM phosphate buffer (PB) pH 7.4, the secondary antibody (anti-mouse IgG conjugated to tetramethylrhodamine isothiocyanate (TRITC) in PB) was added to the samples and incubated for 1 hr in the dark at RT. The slides were then washed 3x in PB and mounted with Dako mounting medium. The labeled venom gland section was then visualized using a Zeiss LSM 550 confocal microscope with an excitation wavelength of 552 nm and long pass emission at 575 nm.

Results

Timing and variation in snake venom gene expression

Because previous gene expression studies on venom glands have primarily focused specifically on expression of venom genes, we first analyzed patterns of venom gene expression to illustrate the degree of upregulation of venom gene production following venom depletion. At 1DPE, average expression of major venom genes is significantly upregulated (Fig. 1b) to the extent that venom gene expression dwarfs that of non-venom genes at a genome-wide scale (Fig. 1c). For nearly all of these venom genes, expression decreases between 1DPE and 3DPE, although expression remains elevated during this interval relative to the expression in unextracted venom gland for most venom genes (Fig 1b). In contrast, venom genes are not expressed at notable levels in the three sampled body tissues (Fig. 1b).

Differential expression of all non-venom genes across time points

To characterize the full cellular and physiological response of the venom gland during venom production, we focused on analyses of differentially expressed genes that are not known venom genes (Fig. 2). Between the unextracted venom gland and 1 DPE venom gland samples, we identified 2,589 differentially expressed non-venom genes, roughly half of which are upregulated (1,322 genes). Comparatively few genes are differentially regulated between 1 and 3 DPE, with only 100 genes differentially expressed, 54 of which were upregulated (Fig. 2a). Given the small number of differentially expressed genes between 1 and 3DPE, we focused on the unextracted versus 1DPE and unextracted versus non-venom tissues gene sets in downstream functional inferences.

Regulatory mechanisms associated with venom gland physiology during venom production

To characterize molecular mechanisms involved in gland physiology during venom production, we performed Core Analysis in Ingenuity Pathway Analysis on the 2,589 genes that showed significant differential expression between unextracted and 1DPE venom gland tissue (Fig. 2a-c). Core Analysis separately infers relative changes in regulatory activity of both canonical pathways and upstream regulatory molecules (URMs) based on observed patterns of differential gene expression. We separately performed GO term overrepresentation analysis to identify functional categories of genes that were enriched in genes that were upregulated during venom production (Fig. 2d). Together, these analyses highlight three distinct categories of molecular regulatory activity in the venom gland during venom regulation (Fig. 2).

All functional analyses of gene expression during venom production provide evidence for activation of stress response mechanisms (Fig. 2b-d). Four of the six canonical pathways inferred to have increased activity at 1DPE are stress response pathways. These include the endoplasmic reticulum stress response and unfolded protein response pathways, which are associated with mitigating cellular stress caused by misfolded proteins and high demands for protein processing, as well as two DNA damage checkpoint regulation pathways that act to prevent replication of cells that have accumulated significant DNA damage (Fig. 2b). Activation of stress response mechanisms is also emphasized in inferences of URM activity, in which endoplasmic reticulum to nucleus signaling 1 (ERN1, a.k.a. IRE1) and activating transcription factor

6 (ATF6), two of the three primary stress sensors that lead to activation of the unfolded protein response, are inferred to be activated (Fig. 2c). Further, x-box binding protein 1 (XBP1), a transcription factor activated by IRE1 in response to ER stress that subsequently upregulates genes associated with protein folding and degradation, is inferred to have the highest increase in activity among URMs between unextracted and 1DPE venom gland samples (Fig. 2c). Two additional stress related URMs are inferred as activated during venom production, including nuclear factor erythroid 2 like 2 (NFE2L2, a.k.a. NRF2), a high-level regulator of the NRF2 oxidative stress response pathway, and STIP1 homology and u-box containing protein 1 (STUB1), which targets misfolded proteins for degradation (Fig. 2c). Similarly, GO term analysis of significantly upregulated genes between unextracted and 1DPE show overrepresentation of several terms related to cellular stress responses, including “response to topologically incorrect protein” and terms related to the proteasome complex (Fig. 2d).

Multiple pathways and URMs related to cellular growth and proliferation, cell cycle regulation and tumor suppression are also inferred to be activated during venom regulation (Fig. 2b,c). The PTEN signaling pathway, which negatively regulates cell growth and proliferation, increases significantly in activity at 1DPE (Fig. 2b), along with URMs involved in regulation of cellular growth and proliferation, including insulin like growth factor 2 (IGF2), insulin receptor (INSR) and phosphoinositide-3-kinase regulatory subunit 1 (PIK3R1; Fig. 2c). Additional URMs with high estimated increases in activity during venom production include sterol regulatory element binding transcription factor 1 (SREBF1) and hydrocarboxylic acid receptor 2 (HCAR2), which are involved in lipid signaling and metabolism; ATPase copper transporting beta (ATP7B), IκB kinase, an upstream regulator within the Nf-κB signaling pathway; and transcription factor EB (TFEB), a regulator of lysosome biogenesis and pro-autophagy signaling, among others (Fig. 2c).

We also find broad evidence of pathways, regulators and functional groups of genes associated with the pronounced secretory function of the venom gland during venom production. These include the RAN signaling pathway, which is inferred to have the highest degree of activation at 1DPE (Fig. 2b) and is involved with nuclear transport of proteins and RNAs, as well as numerous enriched GO term categories associated with the processes of transcription, translation, protein transport, modification, and metabolism (Fig. 2d).

Steady-state regulatory mechanisms that differentiate the venom gland from other tissues

To characterize the physiological and regulatory distinctiveness of the venom gland at a steady state, we identified regulatory mechanisms and enriched categories of differentially expressed genes from comparisons between the unextracted venom gland and a group of other secretory organs (pancreas, skin, and stomach), and compared these to inferences of regulatory activity during venom production (Fig. 3). We identified 8,032 significantly differentially expressed genes, of which 3,134 are upregulated, in the unextracted venom gland tissue relative to body tissues (Fig. 3a). A total of 1,688 differentially expressed genes are identified both in pairwise comparisons of unextracted venom gland to body tissues and unextracted venom gland to 1DPE venom gland (Fig. 3a).

This analysis showed evidence for high activity of cellular stress response mechanisms in the unextracted venom gland relative to non-venom secretory tissues (Fig. 3b-d). The unfolded protein response and endoplasmic reticulum stress pathways, both of which showed evidence of activation during venom production, are here inferred to be activated to a greater degree in the unextracted venom gland relative to body tissues (Fig. 3b), indicating that these pathways exhibit a high baseline activity in the steady-state venom gland and are further upregulated during venom production. Similar to inferences during venom production, XBP1 and ERN1 (IRE1) are inferred to be relatively active in the venom gland compared to non-venom body tissues, as well as eukaryotic translation initiation factor 2 alpha kinase 3 (EIF2AK3, a.k.a. PERK), another high-level regulator within the unfolded protein response (Fig. 3c). Overrepresented GO terms include “response to endoplasmic reticulum stress” and other terms related to responses to misfolded proteins (Fig. 3d), indicating a relatively high degree of endoplasmic reticulum stress in the unextracted venom gland relative to other secretory tissues.

Several pathways and URMs involved in tumor suppression, cell cycle regulation, and regulation of cellular growth and proliferation are relatively highly active in the unextracted venom gland, and these are largely exclusive of regulatory mechanisms implicated in venom production. For example, while PTEN is inferred to be upregulated both in this analysis and during venom production, pathways including protein kinase A signaling, PPAR signaling and Wnt/β-catenin signaling pathways are uniquely activated in the unextracted venom gland (Fig. 3b). Further, while multiple URMs broadly involved in insulin-related growth signaling (i.e. IGF2, INSR, and PI3KR1) show evidence of activation during venom production, these URMs are not observed in the unextracted venom gland and instead two fibroblast growth factor (FGF) transcription factors show evidence of activation. Additional URMs with inferred activation in the unextracted venom gland include Ras-related protein Rab-1B (RAB1B), a regulator of intracellular vesicle transport between the ER and Golgi, which is also activated during venom production (Fig. 3c).

Analyses of GO terms identified a substantially greater number of overrepresented terms in this analysis compared to those during venom production that were directly associated with secretory machinery of the venom gland, including many associated with the activity of the endoplasmic reticulum and Golgi, intracellular transport via vesicles, and protein folding, export, and metabolism (Fig. 2d, 3d). In contrast, fewer terms associated with regulation of transcription and translation are overrepresented in this set of genes upregulated in the unextracted venom gland relative to other secretory tissues (Fig. 2d, 3d).

Detailed analysis of unfolded protein response activity in the snake venom gland

Our analyses of the venom gland both in an unextracted and post-extracted state consistently infer activation of stress response mechanisms, and primarily those of the unfolded protein response (UPR; Figs. 2, 3). To better understand the activation of this pathway, we characterized all genes, URMs, and specific components of the UPR pathway that are activated in the venom gland (Fig. 4). Pairwise comparisons between the unextracted venom gland and body tissues yielded the largest number of differentially expressed genes involved with the unfolded protein response pathway, the majority of which are upregulated in the venom gland relative to other secretory tissues (Fig 4a). A subset of these genes is further differentially expressed at 1DPE with the majority upregulated, and only three are differentially expressed between 1 and 3DPE (Fig 4a). Notably, all three of the primary high-level regulators of the UPR - PERK (EIF2AK3), IRE1 (ERN1), and ATF6 – are upregulated in the unextracted venom gland relative to other secretory tissues (Fig. 4a,b). These analyses also suggest activation of the NRF2 oxidative stress response pathway in unextracted and 1DPE venom gland tissue given the inferred activation of its primary regulator, NFE2L2 (Fig. 4).

Consistent with the prevalence of overrepresented GO terms related to ER stress response and protein degradation, many genes involved in endoplasmic reticulum associated protein degradation (ERAD) are significantly upregulated in the venom gland (Fig. 4b). Additionally, both the inferred activation of SREBP transcription factor as a URM as well as the observed upregulation of its constituent genes (SREBF1 and SREBF2) indicate potential crosstalk between the UPR and lipid signaling mechanisms and pathways (such as PPAR signaling) that are separately inferred to exhibit unique patterns of activity in the venom gland (Fig. 4b).

Candidate gene analysis of venom gland acidification

While previous studies have demonstrated that acidification of the venom gland lumen plays an important role in venom storage, our initial analyses did not identify clear links to mechanisms associated with acidification beyond overrepresentation of the “phagosome acidification” term in analyses of the unextracted venom gland versus other secretory tissues (Fig. 3d). We performed additional post hoc characterization of 57 candidate genes annotated with the gene ontology term “pH reduction” (GO:0045851) to identify genes with potential roles in venom gland acidification. Of these candidate genes, 35 exhibited a detectable degree of expression in our gene expression dataset, 26 of these are significantly upregulated in the unextracted venom gland relative to non-venom body tissues, and six are upregulated between unextracted and 1DPE venom gland samples (Fig. 5a, Supp. Fig. 1). The majority of significantly upregulated genes are vacuolar-ATPases (V-ATPases), and *ATP6V1C2* in particular shows the greatest degree of upregulation in the unextracted venom gland relative to other secretory tissues (Fig. 5a). Notably, all six candidate genes with significant upregulation during venom production are components of V-ATPases (Fig. 5a, Supp. Fig. 1). In contrast, *ATP4B*, a major component of proton pumps driving gastric acid secretion, was not found to be upregulated in the venom gland relative to other secretory tissues, and instead shows high expression in the stomach only (Fig. 5a).

Western blots and immunohistochemical staining on total proteins extracted from Prairie Rattlesnake gastric and venom gland membranes were used to validate inferences of a role of V-ATPases in driving venom gland acidification (Fig. 5 b-d). Western blots using polyclonal ATPAL1 antibody show a high prevalence of gastric H+/K+ ATPases in rattlesnake gastric membrane only, with no detectable presence in the venom gland (Fig. 5b). Conversely, Western blots using αH56 antibody show prevalence of V-ATPases in both gastric and venom gland tissues (Fig. 5c), consistent with gene expression data, and immunohistochemical staining demonstrates that V-ATPases are concentrated in mitochondria-rich cells in the venom gland secretory epithelium (Fig. 5d).

Discussion

This study provides new insight into the dynamic physiology of the snake venom gland and the molecular mechanisms associated with the secretory demands placed on this organ during venom production. We show that the depletion of venom in the rattlesnake venom gland is met with differential regulation of thousands of non-venom genes that orchestrate a complex suite of regulatory responses to support venom production, highlighting the complexity of this process. Our inferences provide the first detailed molecular characterization of this response, including regulatory mechanisms driving secretory function and epithelial maintenance, and highlight a role of cellular stress response activation during venom production. Characterization of the unextracted venom gland gene expression programs compared to other secretory tissues suggests unique activation of additional signaling mechanisms associated with cell growth, survival, and extreme secretory capacity, as well as further evidence for stress response activation, in the venom gland. Our results also provide new evidence for mechanisms of venom lumen acidification that maintains venom toxins in an inactive form during storage prior to envenomation.

Comparisons of gene expression between unextracted and 1 DPE venom gland tissues identified over 2,500 non-venom genes that are differentially expressed within 24 hours of venom depletion (Fig. 2a), emphasizing that venom production entails a complex and highly regulated process necessitating the activation of a large coordinated gene expression response to facilitate production of a comparatively small number of venom components. These genes are inferred to drive peripheral physiological and cellular processes that are likely central to cellular and physiological shifts necessary to facilitate venom production, including the upregulation of regulatory pathways, molecules and genes involved in protein production, transport, and export. This response also includes suites of genes associated with cellular stress response, pro-survival and cell cycle regulation, and cellular growth and proliferation signaling (Fig. 2). In particular, the high number and magnitude of inferences related to stress response activation imply an important balance between upregulated cellular activity to replenish venom stores rapidly and the mitigation of resulting cellular stress to prevent DNA damage, apoptosis, and cellular dysfunction. Activated cell growth and proliferation signaling may also be indicative of epithelial cell turnover and general epithelial maintenance, which would further compound cellular stress during venom production. Collectively, these findings indicate that the extreme performance of the venom gland during venom production necessitates the activation of stress response mechanisms, and that cellular stress may be an important and unappreciated constraint in the evolution and function of venom systems.

Evidence for the extreme physiological regulation of the venom gland during venom production raise questions about what broad physiological and regulatory mechanisms may distinguish the snake venom gland from other body tissues and secretory organs. Similar to analyses of venom production, inferences of pathway and regulatory molecule activation in the unextracted, steady-state venom gland relative to other secretory tissues include mechanisms of cellular stress response, and primarily those related to the UPR (Fig. 3b-d). We also find evidence for the activation of a largely discreet set of pathways and molecules broadly involved in cellular growth and proliferation, lipid metabolism, and cell cycle regulation compared to those activated during venom production (Fig. 3b,c), suggesting that a distinct suites of regulatory pathways drive general epithelial maintenance during this steady state. The PTEN signaling pathway, which typically acts to negatively regulate cell proliferation, was inferred to be activated to a relatively high degree in both analyses of venom production and the comparison between venom and non-venom tissues, suggesting that regardless of the state of venom gland activity, cell proliferation is monitored and controlled more so than in other secretory tissues.

Functional characterization of genes upregulated in the venom gland compared to other secretory tissues identified overrepresentation of terms related to cellular secretory function and machinery (i.e. related to endoplasmic reticulum, Golgi, vesicle transport, protein processing components or function) and responses to endoplasmic reticulum stress and unfolded proteins (Fig. 3d). Together with evidence of activated regulatory mechanisms, these findings appear to indicate a higher degree of cellular and secretory activity occurring in the unextracted venom gland than previously predicted (Mackessy and Baxter, 2006). Alternatively, evidence of heightened regulatory activity and cellular secretory function in the steady state venom gland may represent physiological adaptations of the gland to maintain a primed highly responsive steady state that facilitates venom production to respond rapidly to venom depletion.

All of our analyses emphasize an important role of stress response mechanisms in the snake venom gland. Notably, the venom gland exhibits a high baseline level of activated stress response mechanisms compared to other secretory tissues. This suggests that the storage of venom in the venom gland may induce elevated levels of cellular stress, and may additionally suggest that the venom gland 'steady-state' may actually involve active maintenance of venom stores via low constitutive levels of venom production, further evidenced by elevated venom gene expression in the unextracted gland (Fig. 1). While previous studies have found little evidence for turnover of venom components during storage prior to depletion (Kochva et al., 1982; De Lucca et al., 1974; Mackessy and Baxter, 2006; Rotenberg et al., 1971), and it is unclear to what extent observed mRNA expression corresponds to venom protein production, our results raise the possibility that a degree of venom production is constitutive and ongoing at all times.

Previous studies have shown that acidification of the venom gland lumen inhibits venom activity and thus plays an important role in self-protection against the venom activity during storage (Mackessy and Baxter, 2006). However, the mechanisms of venom gland acidification and how they relate to mechanisms of acidification in other tissues is unknown. In mammalian parietal cells, acidification is driven by hydrogen potassium ATPase proton pumps encoded by the genes *ATP4A* and *ATP4B* (Rabon and Reuben, 1990), which act to secrete hydrochloric acid into the stomach lumen. In the rattlesnake venom gland, *ATP4B* exhibited high expression in the stomach as expected, but low expression in venom gland tissues, pancreas, and skin (Fig. 5), suggesting that hydrogen potassium ATPase pumps are not involved in venom gland acidification. Instead, analysis of candidate genes involved in proton transmembrane transport identified significant upregulation of multiple V-ATPases (Fig. 5) that, while typically associated with acidification of secretory vesicles (Nishi and Forgac, 2002), drive luminal acidification in some tissues (Breton and Brown, 2007; Brown et al., 1988, 1992; Pamarthy et al., 2018). Western blots of rattlesnake stomach and venom gland membranes confirm the lack of gastric ATPases and abundance of V-ATPases in the venom gland, and immunohistochemical staining for V-ATPases indicate that V-ATPases are concentrated in mitochondria-rich cells present in the venom gland epithelium. Collectively, these findings strongly indicate a role of V-ATPases in the direct acidification of the venom gland lumen, and support previous inferences that mitochondria-rich cells in the venom gland epithelium are the primary drivers of gland acidification.

This study provides a valuable perspective on the complex nature of venom gland physiology in the Prairie Rattlesnake, and we expect that many aspects of our findings are likely applicable to snake venom gland physiology in general. However, the high degree of variation in venom phenotypes observed both among conspecific populations and between venomous snake species raises the question of whether there exists corresponding variation in specific aspects of venom gland physiology and function. Future studies that incorporate greater replication both within and across species will be vital to develop a comprehensive understanding of venom gland physiology, regulation, and evolution across the diversity of venomous snakes.

Conclusion

Snake venom systems have emerged as an important model for addressing a broad array of biological, evolutionary and biomedical questions. Our analyses of the underlying physiological regulation of venom gland function provide new insight into the extreme cellular "environment" in which venom production takes place, highlighting that the venom gland itself may be as interesting a model system as the venoms it produces. Evidence for the broad coordination of physiological and cellular programs that accompany venom storage and production illustrate the complexity of regulatory systems that have co-evolved to enable venom gland function. These findings raise intriguing questions about co-evolution of these pathways with venom itself, and whether variation in venom phenotypes across species corresponds with nuanced differences in venom gland physiology and function. Our findings also raise the question of whether particular signaling responses related to venom gland physiological regulation are also coupled to regulation of venom – a process that remains poorly understood (Kerchove et al., 2008a; Post et al., 2020; Schield et al., 2019; Yamanouye et al., 2000).

Our analyses of the physiological functions of snake venom glands consistently emphasize a central role of stress response pathways in mediating cell stress and enabling extreme cellular performance. Beyond the venom gland, snakes are important models for other extreme physiological responses, including exceptional physiological upregulation, metabolic fluctuation, and organ regenerative growth upon feeding (Andrew et al., 2017; Perry et al., 2019). Intriguingly, previous studies of molecular mechanisms underlying this post-feeding regenerative growth in snakes have also suggested an important role of stress responses during extreme bouts of growth and regeneration, including those identified in this study (e.g., UPR, NRF2 (Andrew et al., 2017; Perry et al., 2019)). This raises the question of whether stress response activation may play a broadly important role in the evolution of diverse physiological adaptations in snakes.

Acknowledgments

We thank the Minnesota Supercomputing Institute and the Texas Advanced Computing Center for access to computational resources. Support for this work was provided by National Science Foundation (NSF) grants IOS-655735 to TAC, DEB-1655571 to TAC and SPM.

Data Availability

Gene expression data are available from the NCBI Short Read Archive at (NCBI: PRJNA716163). Relevant scripts and code are available at github.com/blairperry/VenomGlandPhysiol.

Figures

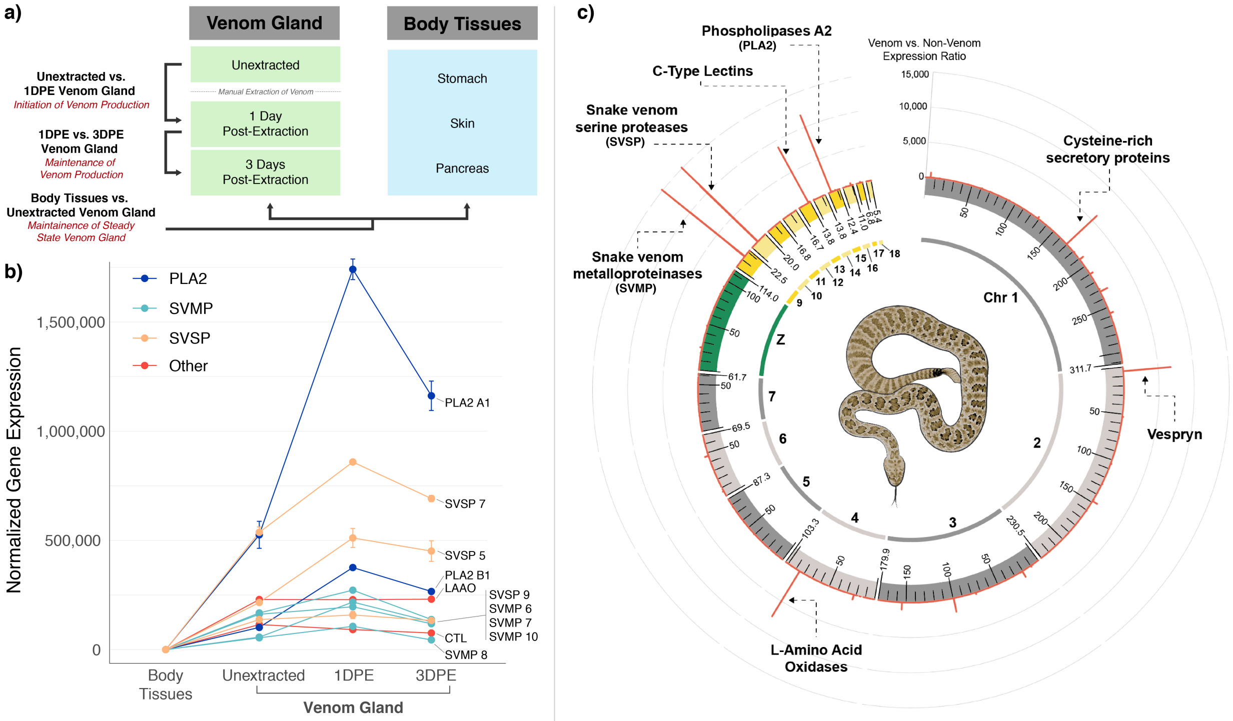


Figure 1. Experimental design and venom gene expression. **A**) Overview of experimental design, highlighting the three pairwise comparisons used to characterize features of venom gland physiology. **B**) Expression of major venom genes across sampled timepoints and in non-venom tissues. **C**) Genome-wide view of gene expression in the rattlesnake venom gland at one day post-extraction (1DPE), with the red line indicating the ratio of expression between 1DPE venom gland and non-venom tissues. Major venom gene clusters (labeled) are easily distinguishable given their high magnitude of expression.

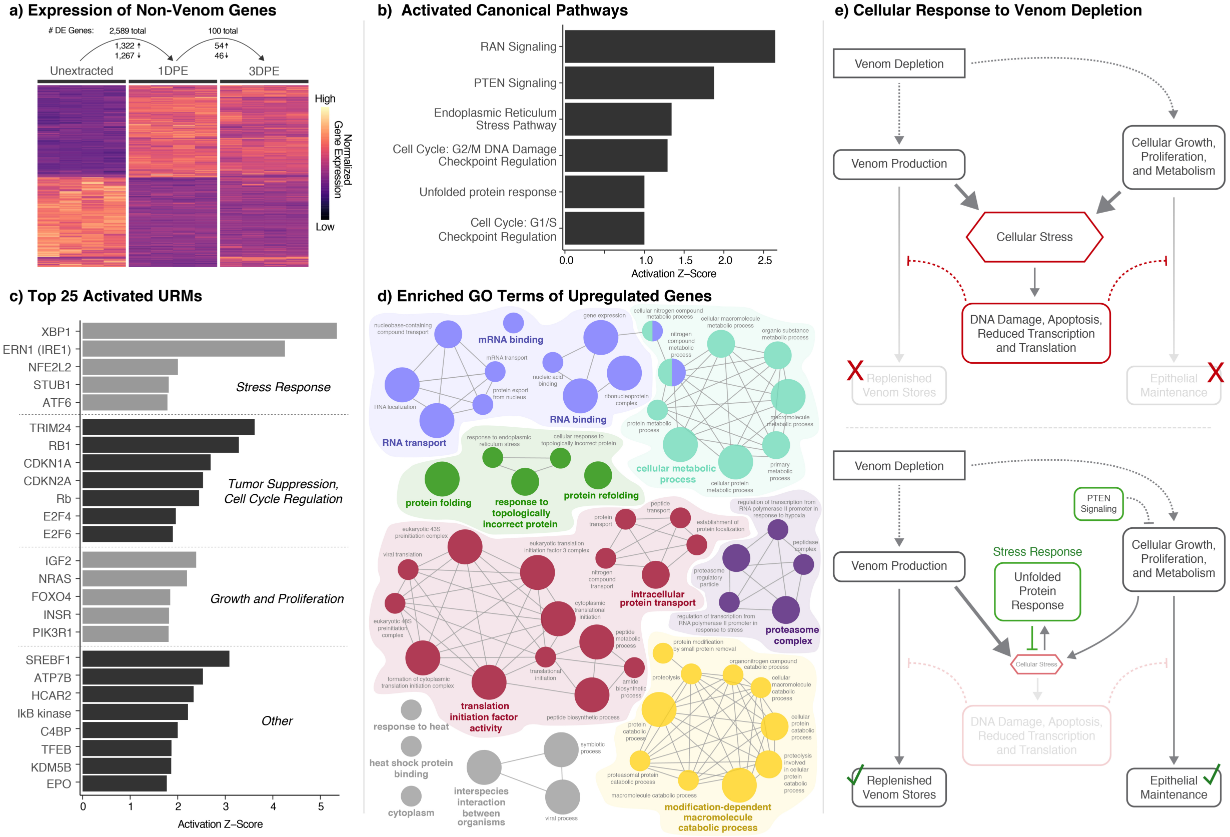


Figure 2. Functional characterization of gene expression in the venom gland during venom production. A) Gene expression heatmap of all genes showing significant differential expression between unextracted vs. 1DPE and 1DPE vs. 3DPE pairwise comparisons. The total number of differentially expressed (DE) genes, as well as the number up- and downregulated, in each comparison is shown above the heatmap. **B**) Inferences of activated canonical pathways based on differentially expressed genes between unextracted and 1DPE venom gland tissues. **C**) Inferences of activated upstream regulatory molecules (URMs) based on differentially expressed genes between unextracted and 1DPE venom gland tissues, with post-hoc categorization based on known activity of URMs. **D**) Significantly enriched GO terms of genes significantly upregulated at 1DPE. All terms shown are significantly enriched ($p < 0.05$), and node size is inversely proportional to enrichment p-value, with larger nodes having a lower p-value. Connected nodes indicate a high proportion of shared genes underlying term enrichment. **E**) Overview of proposed roles of stress response activation in facilitating venom production and epithelial maintenance in the venom gland.

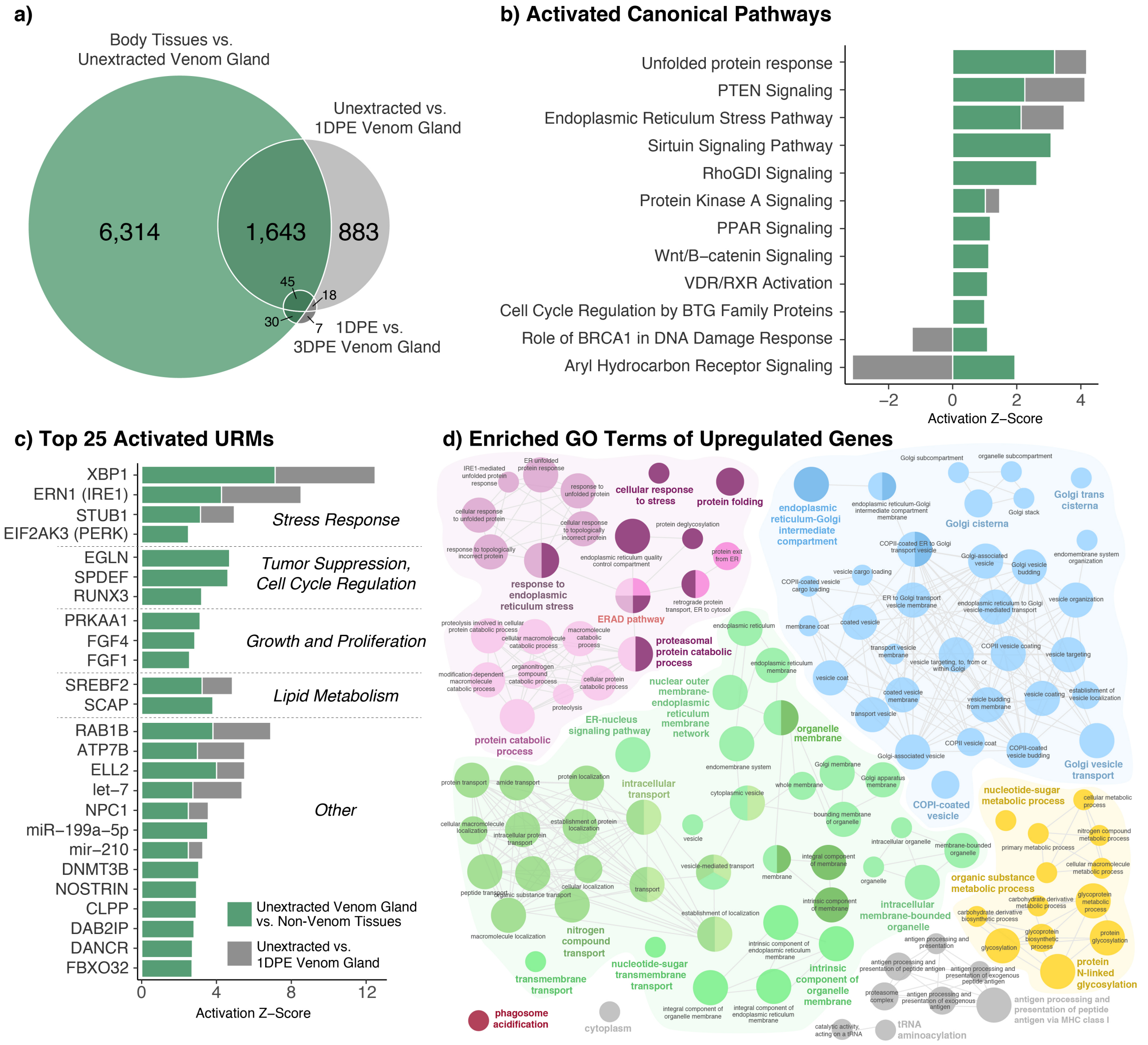


Figure 3. Functional characterization of gene expression in the venom gland relative to other non-venom secretory tissues. A) Venn diagram denoting shared and unique genes between all pairwise comparisons. **B**) Inferences of activated canonical pathways based on differentially expressed genes between non-venom tissues and the unextracted venom gland (green), with inferred activity in analyses of unextracted versus 1DPE venom gland tissues shown in grey if present. **C**) Top 25 inferences of activated URMs based on differentially expressed genes between non-venom tissues and the unextracted venom gland (green), with inferred activity in analyses of unextracted versus 1DPE venom gland tissues shown in grey if present. **D**) Significantly enriched GO terms of genes significantly upregulated in the unextracted venom gland relative to body tissues. All terms shown are significantly enriched ($p < 0.05$), and node size is inversely proportional to enrichment p-value, with larger nodes having a lower p-value. Connected nodes indicate a high proportion of shared genes underlying term enrichment.

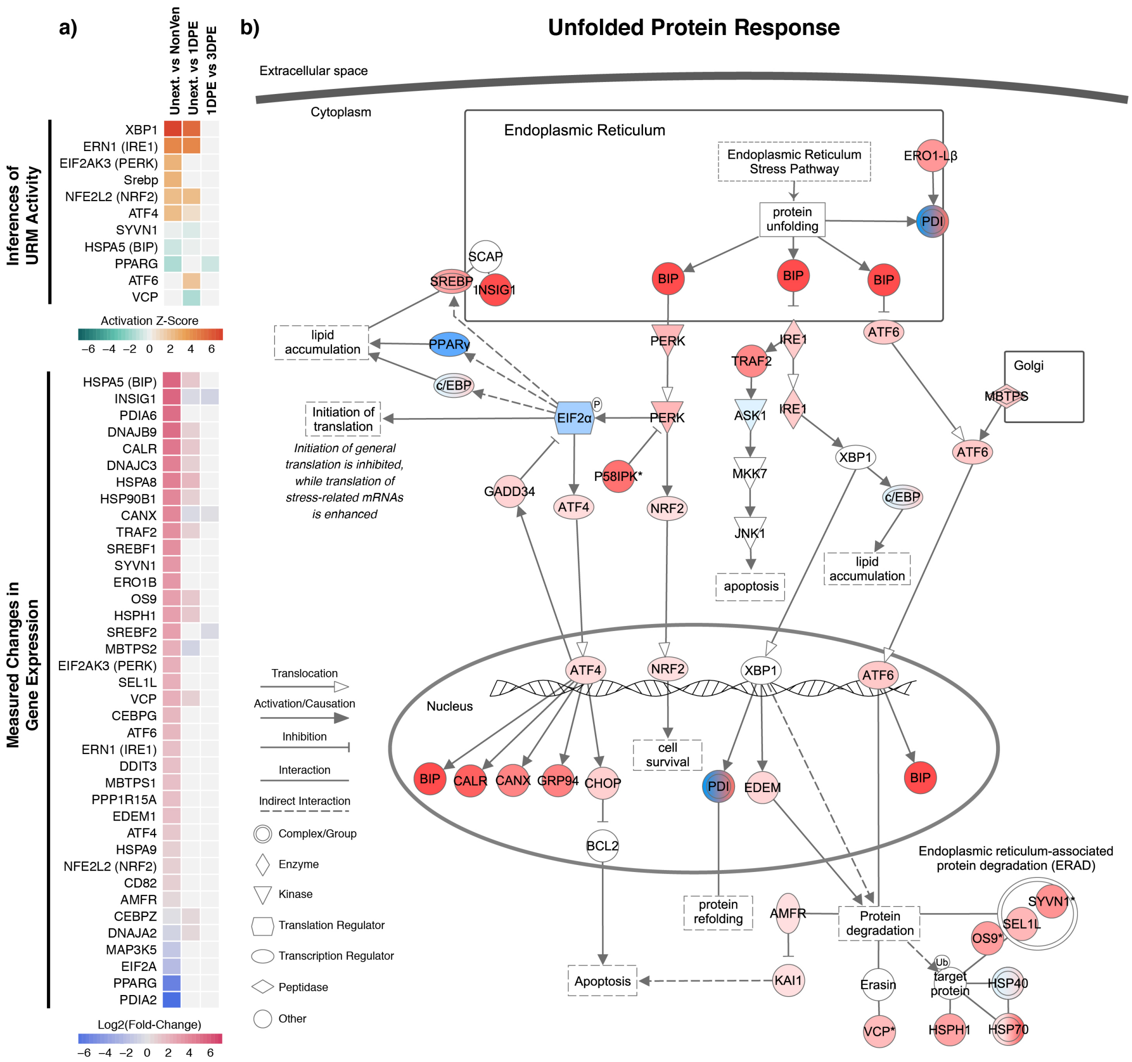


Figure 4. Details of unfolded protein response activation in the venom gland. A) heatmap of inferred upstream regulatory molecule activity (top) and observed differential expression of genes involved in the unfolded protein response (bottom), **B**) Diagram of the unfolded protein response pathway overlaid with observed patterns of differential gene expression in the unextracted venom gland relative to other non-venom tissues, with red indicating upregulation, blue indicating downregulation, and darker colors indicating a higher magnitude change in expression.

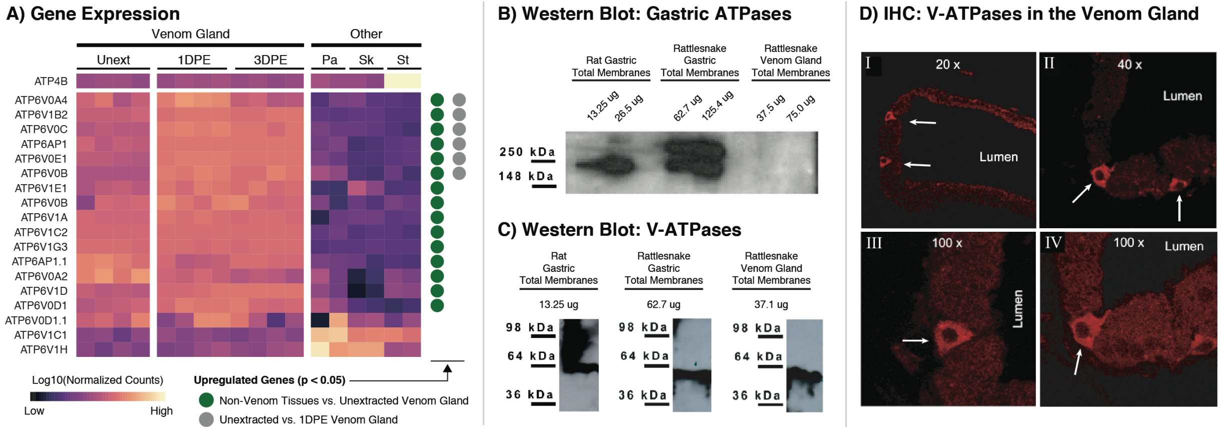
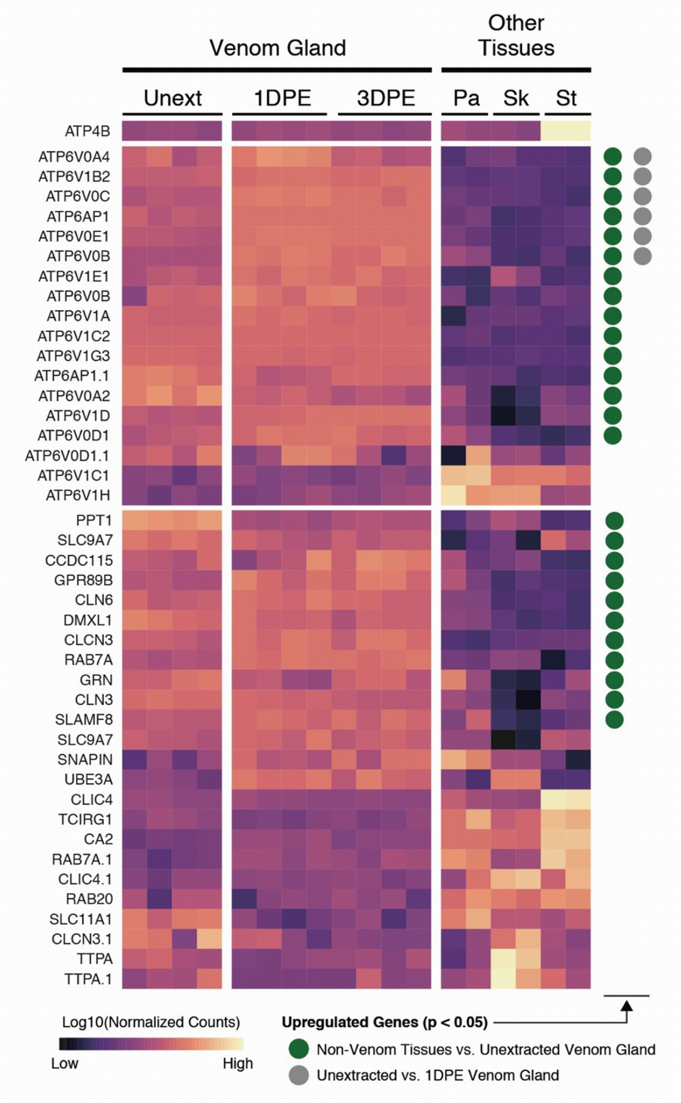


Figure 5. Vacuolar-ATPases are likely mechanisms driving venom gland acidification. A) gene expression heatmap of ATPase genes annotated with the "pH reduction" GO term, with circles on the right side indicating significant differential expression in the two focal pairwise comparisons. **B-C**) Western blots testing for the presence of **B**) gastric H+/K+ ATPases (cropped from a single gel and membrane) and **C**) vacuolar ATPases in rattlesnake stomach and venom gland membranes (cropped from a single gel/membrane), with rat gastric membranes used as a positive control (cropped from a separate gel/membrane). No manipulations beyond cropping were applied to raw images. **D**) Immunohistochemical staining showing an abundance of vacuolar-ATPases in mitochondria-rich cells in the venom gland epithelium (marked by arrows) at I) 20x, II) 40x, and III-IV) 100x resolution.

Supplementary Figures



Supplementary Figure 1. Gene expression heatmap of full set of genes annotated with the "pH reduction" GO term, with circles on the right side indicating significant differential expression in the two focal pairwise comparisons.

Chapter 5

DIVERSE GENOMIC MECHANISMS FACILITATED THE EVOLUTION OF THE SNAKE VENOM GENE REGULATORY NETWORK

Blair W. Perry[1], Siddharth S. Gopalan[1], Giulia I.M. Pasquesi[2], Aundrea K. Westfall[1], Drew. R. Schield[3], Cara F. Smith[4], Ivan Koludarov[5], Paul T. Chippindale[1], Edward B. Chuong[2], Stephen P. Mackessy[4], and Todd A. Castoe[1]

[1]Department of Biology, University of Texas at Arlington, Arlington, TX, 76019, USA.

[2]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO

[3]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO

[4]School of Biological Sciences, University of Northern Colorado, Greeley, CO 80639, USA

[5]Animal Venomics Group, Justus Leibig University, Giessen, Germany

Abstract

Snake venom systems offer a valuable model for understanding the evolution of novel regulatory machinery that drives tissue-specific expression in a highly specialized organ, the venom gland. Here we use multiple functional genomics approaches in the first comprehensive characterization of regulatory mechanisms that coordinate the production of venom in the Prairie Rattlesnake (*Crotalus viridis*). We identify key regulatory sequences for major venom gene families and transcription factors and signaling cascades involved in venom regulation. We also detect topologically associated domains, CTCF loops, and features of chromatin architecture as mechanisms enabling venom gene families to be regulated independently of one another through family-specific transcription factors and regulatory architecture. Our findings further reveal diverse genomic processes that have led to the establishment of novel venom regulatory networks, including tandem duplication of genes and regulatory sequences, cis-regulatory sequence seeding by transposable elements, and diverse transcriptional regulatory proteins controlled by a master regulatory cascade.

Introduction

Understanding processes that underlie the evolution of novel traits has been a long-standing challenge in biology (Wagner and Lynch, 2010). The evolution of novel traits often involves major changes in gene regulatory architecture which may involve the evolution of new regulatory sequences and the co-option or "rewiring" of existing networks and trans-activating factors, or a combination of these (Babu et al., 2004; Teichmann and Babu, 2004; Wagner and Lynch, 2010). Snake venom systems are an ideal model to understand how novel regulatory systems evolve and function (Casewell et al., 2012, 2013; Zancolli and Casewell, 2020) due to the tractable size of venom gene families that comprise venom and the direct relationships to phenotype and fitness (Holding et al., 2016; Rokyta et al., 2015). At the core of snake venom systems is a highly specialized secretory organ – the venom gland (Kochva et al., 1980; Mackessy, 1991; Mackessy and Baxter, 2006; Perry et al., 2020). Within this gland, multiple venom gene families contribute proteins to a venom bolus that is injected during a bite (Mackessy, 2021). Despite an extensive body of literature on snake venoms, the mechanisms and evolution of snake venom gene regulation remain poorly understood. To date, studies have implicated various transcription factors and signaling pathways that may play a role in particular species and venom gene families, but no studies have identified comprehensive mechanisms that explain venom regulation (Hargreaves et al., 2014a, 2014b; Junqueira-de-Azevedo et al., 2015; Kerchove et al., 2004, 2008b; Margres et al., 2021; Schield et al., 2019). Previous studies have identified a small number transcription factors with potential roles in the regulation of particular venom gene families, including AP-1, NF-kB, and Fox-, NF1- and GRHL-family members (Luna et al., 2009; Margres et al., 2021; Nakamura et al., 2014; Schield et al., 2019). Others have implicated the high-level regulatory involvement of alpha- and beta-adrenergic receptors and downstream signaling through the ERK/MAPK pathway in venom production based on stimulation of this pathway after venom depletion (Kerchove et al., 2008b; Yamanouye et al., 2000). While these studies identify some potential mechanisms, an integrated model for the regulation of venom, and the degree to which distinct venom gene families are regulated by common mechanisms, has not been developed. Additionally, although it has been shown that venom gene expression is correlated with open chromatin near gene promoters (Margres et al., 2021), the combined role of chromatin structure and enhancer-promoter architecture in regulating snake venom, as well as the evolutionary origins of these features, have not been investigated.

To address the many outstanding questions about the regulation of snake venom systems, we leveraged a chromosome-level genome assembly for the Prairie Rattlesnake (Schield et al., 2019) together with analyses of RNA-sequencing (RNA-seq), Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq), Chromatin Immunoprecipitation sequencing (ChIP-seq) for insulators (CTCF), open promoters (H3K4me3), and open enhancers (H3K27ac), and Hi-C chromatin contact data to discover regulatory interactions underlying venom gene expression. We use these data to reveal, for the first time, the fundamental architecture and evolution of snake venom regulatory systems. We identify conserved enhancer and promoter sequences within venom gene families, which we demonstrate to interact through the binding of a broad suite of transcription factors (TFs) to coordinate expression. We show that many of these TFs are controlled by ERK (extracellular receptor kinase) signaling, which suggests evolutionary co-option of a conserved vertebrate regulatory network to coordinate venom expression. Our results highlight the roles of diverse evolutionary strategies including TF co-option, regulation of many TFs by a master regulator, tandem duplication, and regulation by sequences seeded by transposable elements that have collectively led to the rewiring of venom gene regulation. Our results further emphasize the role of chromatin state and three-dimensional conformation in directing the precise regulation of venom genes. These discoveries provide the first mechanistic and evolutionary characterization of the complex regulatory architecture underlying snake venom, and provide a valuable example of how multiple genomic processes can synergistically act to generate novel regulatory networks underlying a polygenic trait.

Results

Massive upregulation of venom gene expression to replenish venom

The secreted venom proteome of the Prairie Rattlesnake is dominated by snake venom serine proteases (SVSP), snake venom metalloproteinases (SVMP), phospholipase A2 (PLA_2), and peptide myotoxins (Fig. 1a), which all derive from tandemly-arrayed multi-copy gene families (Fig 1b; Saviola et al., 2015; Schield et al., 2019). In SVMP and PLA_2 clusters, one or more non-venom paralogs (i.e., venom gene paralogs that do not contribute to venom activity; NVPs) are present at one end of the array (ADAM28 for SVMP, PLA_2 gIIe for PLA_2; Fig. 1b). The myotoxin cluster has not been successfully assembled and annotated in the Prairie Rattlesnake, so myotoxins were excluded from this study. A smaller fraction of the Prairie Rattlesnake venom proteome is composed of additional proteins and peptides encoded by genes in small tandem arrays (i.e. 2 genes; "CRISPs" in Fig. 1b) or as individual genes (Saviola et al., 2015). For analyses, we grouped venom genes by SVMP, SVSP, PLA_2, and a category of "other" venom genes that comprise the minor fraction of venom proteins (Fig. 1c). Venom genes were highly expressed in the venom gland, with vastly lower expression in non-venom gland tissues (Fig. 1c). Venom gene expression was massively upregulated in the venom gland following venom depletion, with major venom gene clusters readily distinguished at a genome-wide scale (Fig 1c-d).

Identification of candidate transcription factors associated with venom regulation

We identified candidate TFs likely to be important for venom regulation using three independent approaches (Fig. 2, Supp. Table S1). First, we identified 82 TFs with significantly upregulated gene expression between unextracted and one-day post-extraction (DPE) venom gland tissue (Fig. 2b, c). Of these 82 TFs, 46 were also expressed at higher levels in the venom gland compared to non-venom tissues (Supp. Table S1). We separately identified a second set of TFs that fall within or adjacent to super-enhancer (SE) regions identified using H3K27ac ChIP-seq data. SEs represent "regulatory hotspots" in the genome associated with genes critical to cell- and tissue-specific functions, including genes encoding TFs central to tissue-specific function (Chapuy et al., 2013; Hnisz et al., 2013; Jia et al., 2019; Lovén et al., 2013; Whyte et al., 2013). We identified 504 SEs in the 1DPE venom gland (Supp. Fig. S1a) and 946 SE-associated genes, with 813 directly overlapping an SE and the remaining 133 adjacent to an SE. SE-associated genes were enriched for venom genes, genes involved in adrenergic receptor binding, and transcription factor activity (Supp. Fig. S1b, c). This approach identified 81 SE-associated TFs (Fig. 2b), 14 of which were identified above as upregulated during venom production (Fig. 2a, b). In a third approach, we used differential footprinting analysis of ATAC-seq data to identify 55 TFs with evidence for increased binding activity in the post-extraction versus pre-extraction venom gland (Fig. 2b; Supp. Fig. S1d). Of these 55 total TFs with evidence of increased binding in the 1DPE venom gland, 7 and 3 TFs were also identified in SE-associated or upregulated candidate approaches, respectively, with a single TF (CAMP Responsive Element Modulator; CREM) identified by all three approaches (Fig. 2b).

Candidate venom regulation TFs linked to adrenergic receptor signaling and ERK

Our analysis identified candidate TFs which are consistent with previously hypothesized mechanisms regulating venom gene expression. Multiple SE-associated candidate TFs are involved in adrenergic receptor binding activity, and additional candidate TFs (e.g., FOS and JUN) are components of the AP-1 transcription factor complex previously implicated in venom gene regulation (Fig. 2a; Luna et al., 2009). An additional SE-associated candidate TF (EHF) was previously shown to play a role in regulating venom PLA_2 genes in *Protobothrops* (Nakamura et al., 2014). Other candidate TFs have been identified as potential regulators of venom genes in genomic studies, including FOX-family TFs, GRHL1, NFIA and NFIB, XBP1 (Fig. 2a; Margres et al., 2021; Schield et al., 2019). Several of these TFs (i.e., EHF, GRHL1, JUN, multiple FOX-family TFs) showed increased binding activity during venom production, based on ATAC-seq footprinting (Fig. 2a).

Direct interactions with ERK, the primary regulator of the ERK signaling pathway previously implicated in venom regulation (Kerchove et al., 2008b; Yamanouye et al., 2000), were identified for 27 candidate

TFs (Fig. 2a, c), and 32 additional TFs had known interactions with these 27 TFs (i.e., second-degree interactions with ERK; Fig. 2c). This was corroborated by KEGG pathway overrepresentation analysis of the full candidate TF set indicating overrepresentation of TFs involved in the ERK/MAPK signaling pathway (enrichment ratio = 2.6404, FDR = 0.025745; Supp. Table S2).

Promoter chromatin state and venom expression

We used H3K4me3 ChIP-seq data to investigate the relationships between gene expression and open promoters at 1DPE (Fig. 3; Supp. Fig. S1e-j). Open promoters were inferred if H3K4me3 ChIP-seq peaks were identified within ±1kb of transcription start sites (TSSs) or the bounds of promoter ATAC-seq peak regions for key venom genes (see Methods). Genome-wide, 38% of promoters were inferred to be open at 1DPE, and genes with open promoters were more highly expressed (Supp. Fig. S1e); this result was also observed for venom genes when analyzed together (Supp. Fig. S1f). Within specific venom gene families, SVMPs and the combined group of "other" venom genes with open promoters were more highly expressed than those without (Supp. Fig. S1g,j; Mann-Whitney ∪ test, $p < 0.05$), and all PLA_2 gene promoters overlapped with an H3K4me3 peak (Supp. Fig. S1i). SVSPs with open promoters did not show significantly higher expression than those without open promoters (Supp. Fig. S1h).

Venom gene promoters contain suites of transcription factors linked to ERK

To infer regulatory activity of our candidate TFs, we identified transcription factor binding sites (TFBS) for candidate TFs with enrichment in venom gene promoters compared to the promoters of non-venom genes and used RNA-seq and ATAC-seq footprint scores to weight the likely importance of specific transcription factors and TFBS. A total of 12, 11, and 7 TFBS were significantly enriched in SVMP, SVSP, and PLA_2 promoters, respectively, compared to non-venom gene promoters. Several TFs were enriched in two of the three families (JUN, Creb3l2, TFAP4, Arnt); none were in all three (Supp. Fig. S2a). Promoter regions in each of the three venom gene families were alignable and exhibited substantial sequence similarity (Fig. 3d-f). Sequence similarity was highest for promoters of the most highly-expressed genes of each family, while promoters of lowly-expressed genes tended to be more divergent. Inferences of TFBS positions in promoter alignments are highly consistent among gene family members, suggesting that TFBS are conserved among paralogs within venom gene clusters (Fig. 3d-f).

Binding sites important for the regulation of a particular venom gene family were expected to be present, conserved, and show evidence of being bound by transcription factors in the promoter regions of highly-expressed venom genes. Conversely, TFBS should be absent, altered, and/or inaccessible in the promoters of lowly-expressed genes of the same family. Based on this logic, we developed a "position score" using binding site conservation, ATAC-seq footprint scores, and gene expression to quantify the degree to which a given TFBS adhered to these patterns and may therefore be involved in the regulation of a given venom gene family (see Methods; Supp. Fig. S2b). TFBS with a position score equal to or greater than the mean of all position scores for a gene family were considered "high-scoring" (Fig. 3d-f); this threshold is intentionally inclusive of TFBS with moderate to strong evidence of involvement in venom gene regulation, while excluding TFBS with weak evidence of involvement. A total of 11, 9, and 6 TFs had one or more high-scoring binding sites in SVMP, SVSP, and PLA_2 promoters, respectively. For all TFBS that were enriched and had high position scores in at least one venom gene family ("primary" TFBS hereafter), we then identified and scored any of these TFBS present, but not enriched, in the promoters of other venom gene families (Fig. 3g). Any newly-identified TFBS with a position score equal to or greater than the original mean score threshold for a given family was considered to have evidence of being bound, and we refer to these as "secondary" TFBS. The overlap of primary and secondary TFBS inferences produced four TFs showing evidence of involvement in the regulation of all three major venom gene families: FIGLA, TBX3, EHF, and ELF5 (Fig. 3g).

We applied a similar approach to characterize secondary TFBS in the promoters of the "other" venom gene set. Because these genes do not occupy large multi-gene clusters with multiple highly- and lowly-expressed family members, the position score calculation was not applicable, and we instead weighted TFBS inferences in their promoters by the 1DPE ATAC-seq footprint score (Fig. 3h). These analyses further suggested the importance of a number of candidate TFs, most notably ELF5 and TBX3 (Fig. 3h).

Several high-scoring TFs across multiple venom gene families are involved in, or are regulated downstream of, the ERK signaling pathway. These include TBX3 and EHF, for which high-scoring primary and secondary TFBS were identified in promoters of all three major venom gene families. Others included ETS1, CREB3L2, JUN, and ATF4, all of which were associated with one or more venom gene families.

To investigate the possibility that venom genes are regulated by a novel TF that would be missed using only our candidate approach, we analyzed regions of venom gene promoters with elevated ATAC-seq footprint scores for enrichment of *de novo* sequence motifs. No such sequences were identified (Supp. Table S5).

Candidate TFBS in promoters of non-venom paralogs

For all annotated non-venom paralogs for the three major venom gene families, we scanned promoter sequences for the presence of TFBS implicated in each corresponding venom gene family (Supp. Fig. S2c,d). Across the three families, only three TFBS were enriched in the promoters of both the venom and corresponding non-venom paralogs (ATF4 and ZBTB26 in SVSP and NFE2 in SVMP; Supp. Fig. S2d). EHF, which was inferred to be a "secondary" TFBS in PLA_2 promoters, was enriched in PLA_2 non-venom paralogs (Supp. Fig. S2d). A larger number of TFBS implicated in venom gene regulation were present in one or more non-venom paralogs of a given venom gene family despite not being enriched, and each group of non-venom paralogs were enriched for multiple TFBS not implicated in regulation of related venom genes (Supp. Fig. S2c).

Roles of enhancers in regulating venom expression

We used the Activity-by-Contact (ABC) model approach, which identifies putative enhancer regions and estimates their contribution to gene expression (Fulco et al., 2019) using contact information (Hi-C), chromatin accessibility (ATAC-seq), histone modifications (H3K27ac ChIP-seq) to identify enhancer sequences involved in the regulation of venom genes. We inferred 7,059 putative enhancer regions (PERs) associated with 7,119 genes in the 3DPE venom gland, with an average of 1.75 PERs per gene. PERs were most commonly located in intergenic regions (62.8% of all PERs), and the median distance between PERs and associated genes was 14,693 bp.

Venom genes were associated with a total of 56 PERs ("vPERs" hereafter; Fig. 4a-c), with many venom genes inferred to be regulated by one or more vPERs (Supp. Fig. S3a). A high proportion of vPERs were located in intergenic regions (72.5%), and vPERs tended to be located relatively close to target genes with a median distance of 4,896 bp (Supp. Fig. S3b). In the SVMP cluster, most genes with high expression were associated with a single vPER located ~4kb upstream of the TSS (SVMP 6, 7, 8, and 10), while SVMP 2 and 5 each had two inferred PERs (Fig. 4a). SVMP 1 did not have any gene-specific vPERs and was instead inferred to be regulated by the same vPER as SVMP 2 (Fig. 4a). The two lowly-expressed NVPs in the SVMP region were not associated with any PERs in the post-extraction venom gland (Fig. 4a). vPERs in the SVSP cluster were more variable, with most SVSPs associated with multiple vPERs and a greater number of vPERs shared between multiple genes (Fig. 4b). All PLA_2 vPERs showed evidence of regulating multiple genes, with a single vPER regulating both highly expressed PLA_2 genes (PLA_2 A1 and B1) (Fig. 4c). This is the only enhancer in this region associated with the most highly expressed PLA_2 gene, PLA_2 A1. All three enhancers in the PLA_2 region also regulate the nearby non-venom paralog, PLA_2gIIe (Fig. 4c).

We also identified 11 vPERs associated with 7 "other" venom genes (those not belonging to the three main venom gene families; Supp. Figs. S3c-i). Three of these "other" venom genes (Vespryn, LAAO3, and CTL) were inferred to have a vPER within an intron (Supp. Figs. S3d,e,g). Interestingly, these three genes are also the most highly expressed of these "other" venom genes (Fig. 1c,d).

Patterns of enhancer TFBS and links to ERK signaling

Similar to promoters, vPERs in venom gene families exhibited sequence similarity (Supp. Figs. S4b-d). We therefore followed the approach used for promoter sequences to search vPER sequences for enrichment of candidate TFBS compared to non-venom PER sequences. Both SVMP and SVSP vPERs were enriched for TFBS of 35 candidate TFs, 14 of which are enriched in both families, while PLA_2 vPERs are enriched for 9 candidate TFs (Supp. Fig. S4). Similar to results for promoter regions, no single candidate TF was enriched in all three major venom gene families, although 21 of the 79 enriched TFs have TFBS in two of three families (Supp. Fig. S4). We also used position scores analogous to our inferences for promoter regions to identify primary TFBS with enrichment and evidence of binding in vPERs of specific venom families, then identified secondary TFBS for these TFs in other venom gene families (Fig. 4d,e; Supp. Figs. S5-7). We identified a total of 14 primary and secondary TFBS in enhancers of all three venom gene families, including multiple ETS-family TFs (e.g., EHF and ELF5), three SOX family TFs (SOX9, SOX17 and SOX18), FIGLA, and FOXP1 (Fig. 4d). Secondary TFBS in "other" venom genes included FOXP3 and FOXL1, which were found in all PERs (although with low footprint scores in some cases; Fig. 4e). Many TFs identified as putatively binding to vPERs were also functionally linked to ERK signaling, 16 of which were also present in venom promoter regions, including ELF5, EHF, FIGLA, JUN, CREB3L2, and ATF5 (Supp. Fig. S8a). In summary, we identify a diverse set of TFs associated with vPERs, many of which appear important for regulating multiple venom gene families and are regulated by ERK signaling.

De novo motif analysis identified two unannotated motifs that were enriched in regions with elevated ATAC-seq footprint scores in vPERs compared to non-venom gene PERs (Supp. Fig. S8b), although neither motif was present in all vPERs. Additionally, these motifs both exhibit similarity to other candidate TFBS, including GATA1, JUN, and TBX3 (Supp. Fig. S8c), and therefore do not likely represent distinct binding sites of a novel TF.

Venom gene enhancers are conserved across species

Consistent with evidence that vPERs are relevant regulatory sequences, we found high-similarity hits to Prairie Rattlesnake vPER sequences in other venomous snake species, suggesting that vPER sequences are highly conserved (Fig. 4f-h). Many of these orthologous sequences exhibit substantial sequence similarity within pit vipers, including conservation of predicted TFBS among species spanning at least ~40 million years (MYs) of divergence (Kumar et al., 2017; Fig. 4h). We also identified similar sequences to the SVMP vPER in a cobra (*Naja naja*) and rear-fanged colubrid (*Thamnophis sirtalis*), suggesting conservation of this enhancer across ~70 MYs of divergence (Fig. 4f,h). Evaluating the conservation of vPERs associated with the SVSP cluster was not feasible because of the high-copy number of this vPER sequence in the rattlesnake and in other snake genomes, which we explore in detail below.

vPERs are not associated with non-venom paralogs

We surveyed the Prairie Rattlesnake genome, and non-venom paralogs specifically, for the presence of sequences similar to vPER regions for the three major venom gene families. For PLA_2s and SVMPs, BLAST only returned hits to the immediate venom gene clusters, the majority of which were to other identified vPERs (Supp. Fig. S9a,c). In the SVMP region, one BLAST hit was upstream of SVMP 11 (Supp. Fig. S9a). This hit is located near the annotated non-venom SVMP paralogs adjacent to the SVMP cluster, although there is no evidence from our data that this acts as an enhancer to those genes in the venom gland. For PLA_2 vPERs, BLAST identified a region spanning the third exon of the PLA_2gIIe non-venom paralog with high similarity to the vPER inferred to regulate highly expressed venom PLA_2s (PLA_2A1 and PLA_2B1; Supp. Fig. S9c). In stark contrast, SVSP vPER BLAST searches returned nearly 5,000 genome-wide hits (Supp. Fig. S10a).

SVSP regulatory sequences are associated with transposable elements

Motivated by the high frequency of SVSP vPER BLAST hits throughout the rattlesnake genome, we investigated links between SVSP vPER sequences and transposable elements (TEs) by comparing SVSP vPER sequences to annotated TEs from C. viridis (Schield et al., 2019) and other snakes (Pasquesi et al., 2018). Multiple SVSP vPER sequences shared high homology (>90%) with the consensus sequence of an annotated DNA transposon (DNA-hAT-Tip100, referred to hereafter as "Cv1-hAT-Tip100"). This TE sequence is significantly enriched for overlap with SVSP promoters and vPERs compared to the genomic background (one-tailed Fisher's exact test; $p < 0.05$). Further, chromosome 10, which houses the SVSP venom cluster, exhibits the highest density of Cv1-hAT-Tip100 elements compared to all other chromosomes (Fig. 5a). Cv1-hAT-Tip100 elements are abundant within the SVSP array, with 8 of 11 SVSP promoters and 6 of 18 vPERs overlapping one or more Cv1-hAT-Tip100 elements (Fig. 5b). Additional Cv1-hAT-Tip100 elements occur in intergenic regions of the SVSP cluster. Sequence divergence estimates between genome-wide Cv1-hAT-Tip100 elements and those in the SVSP region indicate that these Cv1-hAT-Tip100s were active/inserted within the last 75 MY (Supp. Fig. S10b). Using the genome-wide set of these TEs, we estimated the ancestral consensus sequence of the Cv1-hAT-Tip100 TE and scanned this consensus and individual Cv1-hAT-Tip100 elements in the SVSP region for the presence of candidate TFBS enriched in SVSP vPERs and promoters. Multiple TFBS that we predict are important for SVSP regulation are also present in the ancestral consensus Cv1-hAT-Tip100 sequence and are largely conserved in the SVSP region copies (Fig. 5c; Supp. Fig. S11). There are also multiple instances where a small number of single base substitutions from the consensus have apparently led to the gain/loss of new TFBS in SVSP Cv1-hAT-Tip100 elements implicated as regulatory sequences (Supp. Fig. S11a). Many of these TFBS within SVSP Cv1-hAT-Tip100 elements have high ATAC-seq footprint scores (i.e. > 3; Supp. Fig. S11b), suggesting that these TFBS are bound by TFs during venom production. Additionally, many TFBS are conserved in the Cv1-hAT-Tip100 copies that are not within SVSP regulatory elements ("Other" in Fig. 5b, c), and 1DPE ATAC-seq footprint scores for these TFBS are in some cases similar to or higher than those in copies within SVSP vPERs and promoters (Fig. 5c; Supp. Fig. S11b).

Chromatin organization contributes to the precision of venom regulation

To investigate the role of the three-dimensional organization of the genome in venom regulation, including topologically associated domain (TAD) structure and insulation by CTCF, we inferred TADs and chromatin loops using Hi-C data from 1DPE venom gland (Schield et al., 2019) and incorporated CTCF ChIP-seq data to investigate the role of CTCF binding and insulation in venom gene regions (Fig. 6). Our results indicate that only the PLA_2 venom gene cluster falls entirely within a single TAD, with SVMP and SVSP clusters each spanning multiple adjacent TAD regions (Fig. 6b). The most highly expressed genes in the SVMP cluster (SVMP 7-10) occupy a single TAD, which also contains nearly all H3K27ac and H3K4me3 ChIP-seq peaks in this region (Fig. 6b), suggesting regulatory isolation of highly-expressed genes (Fig. 6b). Additionally, no ATAC-seq peaks from the unextracted venom gland are identified within this TAD, despite peaks present in adjacent TAD regions (Fig. 6f). We observed a similar lack of unextracted ATAC-seq peaks at the center of the TAD containing the PLA_2 cluster (Fig. 6f). The most highly expressed SVSPs are also within a single TAD, along with several more lowly-expressed genes. Two other lowly-expressed SVSPs (SVSP 1 and 2) are in an adjacent TAD (Fig. 6). Unlike the SVMP and PLA_2 regions, ATAC-seq peaks from the unextracted venom gland occur within this TAD, although at lower density than peaks at 1DPE (Fig. 6b, f).

We inferred chromatin loops in venom gene regions by combining Hi-C and CTCF ChIP-seq data (Fig. 6a-d). Loop boundaries in the three main venom clusters generally did not correspond closely with TAD boundaries (Fig. 6a-d), and only the relatively small PLA_2 cluster was contained within a CTCF-bound chromatin loop (Fig. 6d). There were numerous chromatin loops in the SVMP and SVSP regions, suggesting substructure in these clusters (Fig. 6d). We note, however, that the high TE content of the SVSP region may result in spurious loops due to mapping errors of Hi-C data. Evidence for bound CTCF was also frequent in major venom gene family regions despite not always occurring at either end of inferred chromatin loops, the pattern normally associated with the insulation of loops by CTCF (Fig. 6c, d). These

CTCF sites may indicate the presence of CTCF-bound loops not detected by our Hi-C data, or other regulatory roles of CTCF in these regions.

The two lowly-expressed NVPs adjacent to the SVMP cluster are found to occupy a TAD that is distinct from highly-expressed SVMPs, and inferred chromatin loops suggest that they are excluded from the loop housing active SVMP genes (Fig. 6d, e). In contrast, the NVP in the PLA_2 cluster (PLA_2gIIE) resides within the same TAD, Super-Enhancer, and CTCF-bound loop, and shares enhancers with nearby venom PLA_2s (Fig. 6d, e). Despite being considered a non-venom gene, this PLA_2 NVP exhibits higher expression in the venom gland compared to non-venom tissues, consistent with being “linked” to venom PLA_2 regulation, but is not upregulated during venom production (Supp. Fig. S12a). Hypotheses are illustrated for local chromatin structure of each venom cluster in Fig. 6g based on inferences from TAD, chromatin loop, CTCF binding, and gene expression datasets, and highlight that chromatin loops and in some case insulation by CTCF likely contribute to the isolation of highly-expressed venom genes from more lowly expressed venom genes and nearby non-venom paralogs.

Discussion

Considering the complexities of eukaryotic gene regulation, understanding how evolution can “rewire” novel regulatory networks underlying polygenic traits is challenging. While previous studies implicated several TFs and regulatory pathways that may be involved in the regulation of particular snake venom genes or families (Hargreaves et al., 2014b; Junqueira-de-Azevedo et al., 2015; Kerchove et al., 2004, 2008b; Margres et al., 2021; Schield et al., 2019; Yamanouye et al., 2000; Zancolli and Casewell, 2020), no mechanisms have been proposed to explain how venom systems are globally regulated, and how this global regulatory cascade interacts to precisely control the regulation of multiple venom gene families. Our findings reveal the integrated roles of high-level signaling by ERK, acting through a diverse suite of TFs that bind promoters and newly discovered enhancer sequences to regulate venom expression (Fig. 7). We also provide new evidence for patterns of chromatin accessibility and genomic organization that direct the precise regulation of snake venom production. We further show that multiple distinct processes have contributed to the evolution of venom regulatory mechanisms in different gene families and highlight the regulatory complexity involved in the control of venom expression, including specific TF activity and features of chromatin organization that control precise venom regulation. We also provide new evidence that venom is globally regulated through the co-option of diverse and often gene-family-specific sets of transcription factors that are controlled by higher-level ERK signaling activity. This regulatory network appears to have originated through idiosyncratic evolution of TFBS in enhancers and promoters for TFs regulated by ERK signaling. Tandem duplication, compact regulatory structure, and the involvement of TEs seeding nascent regulatory sequences, also likely facilitated this process.

A plausible *a priori* hypothesis for venom gene regulation is that a single TF, or a small number of shared TFs, regulate all venom genes, and a recent study in the Tiger Rattlesnake provided evidence in support of this hypothesis (Margres et al., 2021). However, our exhaustive candidate transcription factor analyses derived from complementary functional genomics datasets, together with *de novo* motif searches and integration of results from previous studies (Hargreaves et al., 2014b; Junqueira-de-Azevedo et al., 2015; Kerchove et al., 2004, 2008b; Margres et al., 2021; Schield et al., 2019; Yamanouye et al., 2000; Zancolli and Casewell, 2020) do not support this hypothesis. Instead, our findings support a model in which different venom gene families have evolved a combination of shared and unique regulatory connections to a common upstream signaling network that responds to venom depletion. Importantly, this overarching regulatory network encompasses many TFs previously speculated to be involved in venom regulation in different gene families and species, and thus for the first time presents a model that links preliminary findings from multiple studies with newly identified mechanisms to explain the global regulation of venom systems (Fig. 7). Furthermore, many TFs that we identified (e.g., AP-1, GRHL1, NFIA, CREB3, and FOX family TFs) are “pioneer” TFs that regulate local chromatin accessibility and recruit histone modifying proteins, and may therefore directly regulate chromatin structure and accessibility required for the expression of venom genes (Biddie et al., 2011; Fane et al., 2017; Jacobs et al., 2018; Khan and Margulies, 2019; Zaret and Carroll, 2011). Together, the suite of TFs we identify as being involved in venom regulation appear to encompass an important diversity of functional roles during venom gene regulation, including response to higher-level signaling, opening chromatin, recruiting other TFs and driving promoter-enhancer interactions (e.g., Grossman et al., 2018).

Through the first identification of putative enhancer sequences involved in the regulation of venom genes, we show that regulatory sequences of venom genes are relatively compact and that enhancers typically occur close to (or within) the genes they regulate (see also Hargreaves et al., 2014a). Tandem duplication of these genes often appears to have included the duplication of nearby enhancer sequences (this is particularly apparent in the SVMP cluster; Fig. 4a), resulting in the propagation of duplicated genes that are “pre-wired” to contribute to the polygenic basis of venom. Accumulating evidence for high structural diversity in venom regions between populations and species (Dowell et al., 2016; Giorgianni et al., 2020) suggests that ectopic recombination and gene conversion may further re-shuffle regulatory regions within venom gene clusters. Related to this hypothesis, we discovered that enhancer sequences putatively regulating PLA_2s are associated with, and perhaps derived from, exonic debris resulting from incomplete duplication of the non-venom paralog PLA_2gIIe (Supp. Fig. S12b-e). A previous study showed conservation of exonic debris in the PLA_2 region of rattlesnakes (Koludarov et al., 2020), further supporting a role of exon debris in PLA_2 regulation.

Prominent examples of regulatory rewiring (Ellison and Bachtrog, 2013; Lynch et al., 2015) have emphasized the roles of TEs in seeding regulatory elements, with recent studies reinforcing that TEs are often co-opted for the regulation of host genes (Chuong et al., 2016, 2017; Feschotte, 2008). In the case of SVSPs, we show that a class of DNA transposons contributed sequences that appear to regulate an entire cluster of 11 venom genes, including both promoter and putative enhancer sequences. This suggests that while TE seeding of TFBS has not driven the rewiring of venom gene regulation entirely, it has contributed substantially to the novel regulatory network of one of the largest venom gene families in vipers.

Beyond those derived from TEs, our analyses reveal that most regulatory elements of venom genes are not enriched in non-venom paralogs, although presence of TFBS sequences in some non-venom paralog promoters suggests that the "raw material" for certain functional TFBS may have been present in these non-venom paralogs while other TFBS likely evolved *de novo* and were further propagated by tandem duplications retaining compact cis-regulatory sequences. This presents a challenge for understanding how stochastic evolutionary changes could result in the evolution of many new TFBS in different venom gene families (Hargreaves et al., 2014a). However, the diverse set of TFs implicated in venom regulation in this study, and the broad overarching regulation of these and many other vertebrate TFs by ERK signaling, suggest that there is a broad sequence space across which *de novo* mutations have the potential to produce TFBS that can be targeted by ERK-controlled TFs. Together, the complement of gene structure, tandem duplication, TE seeding of regulatory elements, and the co-option of TFs linked by a shared regulatory regime make venom a prime example of how evolution can re-program regulatory networks for new polygenic traits. Overall, our findings suggest that in addition to the propagation of some functional cis-regulatory regions from non-venom paralogs, the evolution of venom genes entailed the evolution of novel enhancer sequences, and new TFBS for both pioneer TFs and TFs regulated by ERK signaling. However, there is some evidence that a subset of non-venom paralogs may have already been partially responsive or 'pre-wired' to ERK signaling (e.g., matrix metalloproteinases related to SVMPs; Arai et al., 2003).

Our findings raise the question of whether venom evolution has favored a venom regulatory architecture capable of being fine-tuned by selection. Venom composition varies widely among snake species (Casewell et al., 2020) and populations (Jorge et al., 2015; Mackessy, 2010), and is hypothesized to be under intense selection pressures (Aird et al., 2017; Casewell et al., 2011; Juárez et al., 2008). Despite variation in venom composition among species, we discovered putative venom gene enhancer sequences that are largely conserved within venom families and also show considerable conservation across distantly related snake species with common ancestry spanning tens of millions of years (Fig 4k; Supp Figs. S23-24). We postulate that venom regulatory systems have evolved as a multi-tiered, and therefore easily tunable, system in which conserved enhancers regulate the tissue specificity of venom gene expression, while promoter activation (mediated by both chromatin and TF binding) modulates the magnitude of expression of specific loci in response to selection without disrupting global venom regulation. As additional genomic resources for snakes become available, this hypothesis should be tested through comparative studies of venom gene regulatory sequences across diverse species.

Regulation of gene expression in eukaryotes is dependent on many factors, including cis-regulatory sequences, the chromatin state of these sequences, and the three-dimensional loop structures that promote or restrict transcription (Cremer and Cremer, 2001; Cremer et al., 1993). We identified multiple TADs in venom gene clusters indicating a degree of isolation between and interaction within subsets of genes in the two large families (SVMP and SVSP), whereas the entire PLA_2 cluster occupies the center of a large TAD and may be isolated from adjacent genes by a CTCF loop. No inferred CTCF-insulated loops spanned the SVMP and SVSP clusters, but we found evidence of chromatin loops and bound CTCF sites in these regions that likely play roles in directing enhancer-promoter activity. Consistent with this hypothesis, lowly-expressed NVPs adjacent to the SVMP cluster appear physically isolated from SVMPs via a TAD boundary and chromatin loops. Meanwhile the PLA_2 NVP is contained within a CTCF-loop along with venom PLA_2s and exhibits higher expression in the venom gland than in non-venom tissues despite having no known function in secreted venom. Further, many other venom genes are isolated physically and, from a regulatory standpoint, isolated via chromatin loops during venom production.

Key discoveries in this study frame the first comprehensive model for global and gene family-specific regulatory architecture underlying snake venom systems, and a compelling example for how multiple genomic processes may coalesce to establish a novel regulatory system for a polygenic trait. This new resolution of an overarching regulatory network of venom regulation, together with recent advances in snake organoid systems (Post et al., 2020; Puschhof et al., 2021) together represent an important step towards establishing venom systems as models for investigating the origins and regulatory networks (Casewell et al., 2012, 2013; Post et al., 2020; Zancolli and Casewell, 2020). Understanding how genomic variation may percolate through regulatory networks to drive intraspecific variation in venom is highly relevant to snakebite treatment and use of venoms as therapeutics (Casewell et al., 2014). In light of the recent declaration by the World Health Organization of snakebite as a neglected tropical disease, estimated to be responsible for over 100,000 deaths worldwide each year (Gutiérrez et al., 2017; World Health Organization, 2019), our results provide a model for understanding regional and species-level variation in venoms.

Methods

Tissue sampling

Venom was manually extracted from one of the paired venom glands of an adult male Prairie Rattlesnake (*Crotalus viridis viridis*) three days prior to sacrifice, and the other gland was subsequently extracted one day prior to sacrifice. This allowed for acquisition of venom gland and accessory gland tissue at two stages of post-extraction venom production (1 day post-extraction (DPE) and 3DPE) from the same individual. Venom gland, accessory gland, pancreas, skin, stomach, and small intestine tissue samples were dissected out and snap frozen in liquid nitrogen following humane sacrifice of the individual via deep anesthesia with Isoflurane followed by decapitation. To sample venom gland tissue at an "unextracted" steady state, one additional Prairie Rattlesnake was sacrificed with no prior extraction of venom, and venom gland tissues were collected and snap frozen. Both individual animals were collected from the same population in order to control for genetic background in subsequent analyses. All animals were housed and sampled at the University of Northern Colorado under approved and registered IACUC protocols.

Refining venom gene annotations

During early exploratory analysis of ChIP-seq and ATAC-seq data described below, we noticed patterns of elevated read mapping density (i.e., ChIP-seq and ATAC-seq peaks) in several intergenic regions of the SVSP gene array that closely resembled patterns associated with annotated SVSP genes. Following the approach described in (Schield et al., 2019), we used FGENESH+ (Solovyev, 2004) and known guide protein sequences to identify two previously unannotated SVSP genes, which we include in analyses below. Following the naming convention of the nine originally annotated SVSPs, these new SVSP genes were named *SVSP 10* and *SVSP 11*.

RNA isolation, sequencing and analyses

Total RNA was extracted from snap frozen tissues with Trizol reagent (Invitrogen). All tissues were subsampled to produce three technical replicates. Poly-A selected mRNA libraries were sequenced on an Illumina NovaSeq using 150bp paired-end reads.

Raw RNA-seq reads were quality trimmed using default settings with Trimmomatic v0.39 (Bolger et al., 2014). RNAseq reads were mapped to the annotated *Crotalus viridis* genome (NCBI: GCA_003400415.2) using STAR v2.7.3a (Dobin et al., 2013) and raw gene expression counts were estimated using featureCounts v1.6.3 (Liao et al., 2013). Count normalization and pairwise comparisons between time-series venom gland tissues and between venom and non-venom tissues were conducted using DEseq2 v1.30.1 (Love et al., 2014) in R (R Core Team, 2013), with independent hypothesis-weighting p-value correction via the IHW package v1.18.0 (Ignatiadis et al., 2016) using baseMean expression from DEseq2 as the covariate. Annotated venom genes from the *Crotalus viridis* genome publication (Schield et al., 2019)

were considered to be relevant venom genes if they were found to be significantly upregulated in the venom gland (IHW p-value < 0.05) in pairwise comparisons of all post-extraction venom gland samples (1DPE and 3DPE) versus non-venom tissues. Venom genes were considered to be "highly-expressed" if their average normalized expression at 1DPE exceeded 100,000 normalized counts. Gene expression heatmaps were generated with the pheatmap package v1.0.12 (github.com/raivokolde/pheatmap) in R, and genome-wide visualization of venom to non-venom expression ratios was generated using circos v0.69-9 (Krzywinski et al., 2009). Visualization of gene expression in the context of venom gene family arrays conducted using the gggenes package v0.4.1 (github.com/wilkox/gggenes) in R.

Hi-C Sequencing and Analysis

Hi-C data for a *Crotalus viridis* venom gland at one day post-extraction was generated previously (see Schield et al., 2019 for details; NCBI BioProject: PRJNA413201). Raw Illumina paired-end Hi-C reads were mapped to the rattlesnake reference genome using the Juicer pipeline (Durand et al., 2016a), and Hi-C contact maps were generated using KR normalization at 10kb and 5kb resolution. Topologically-associated chromatin domains (TADs) and sub-TADs were determined at 10kb resolution using rGMAP v1.3.1 (Yu et al., 2017) in R. Chromatin loops were identified using the HICCUPS algorithm in Juicer v1.9.9 (Durand et al., 2016a) with default settings. Hi-C contact heatmaps were generated using the Sushi package v1.28.0 (Phanstiel et al., 2014) in R.

Chromatin Immunoprecipitation (ChIP) data generation and analysis

ChIP-seq libraries were generated for post-extraction (1DPE) venom gland tissue by Active Motif (Carlsbad, CA) for bound CTCF and histone modifications H3K4me3 and H3K27ac. Basic ChIP-seq data processing was performed by Active Motif using their standard analysis pipeline. In brief, libraries were sequenced on an Illumina NextSeq 500 using 75-nt reads and mapped to the UTA_CroVir_3.0 genome assembly (GCA_003400415.2) using BWA v0.6.1 (Li and Durbin, 2009) with default settings. Reads that failed to pass Illumina's purity filter, aligned with greater than 2 mismatches, were not uniquely mapped, or were identified as PCR duplicates were removed for all subsequent analyses. Aligned reads were extended in silico at the 3' end using Active Motif's in-house software, and fragment densities were determined for 32-nt bins across the genome. Intervals of enriched ChIP-seq fragment density were determined using MACS2 v2.1.0 (Zhang et al., 2008). Super enhancers were determined by merging enriched H3K27ac intervals if their inner distance was equal to or less than 12,500 bp, and classifying merged regions with the top 5% strongest enrichment as super enhancers.

To infer sites of bound CTCF and putative CTCF-bound chromatin loops, we used MEME v5.1.1 (Bailey et al., 2009) to reconstruct the Prairie Rattlesnake CTCF binding motif (Supp. Fig. S8d) from CTCF ChIP-seq peak sequences. We then scanned all CTCF ChIP-seq peaks for this binding motif, and peaks with a binding motif were considered "verified" ChIP-seq peaks. We then used the pairtobed tool in bedtools v2.29.2 (Quinlan and Hall, 2010) to intersect chromatin loops with verified CTCF ChIP-seq peaks, and considered a chromatin loop to be CTCF-bound if a verified CTCF ChIP-seq peak was identified within 10kb of both ends of the loop.

ATAC-seq data generation and analysis

ATAC-seq libraries were generated for unextracted venom gland, and post-extraction (3DPE) venom gland, and skin tissue by Active Motif (Carlsbad, CA). ATAC-seq library preparation, sequencing, and initial data processing were performed by Active Motif using their standard analysis pipeline. In brief, ATAC-seq libraries were sequenced on an Illumina NextSeq 500 using 42-bp paired-end reads. Reads were then mapped to the UTA_CroVir_3.0 genome assembly using BWA with default settings and the same quality filtering processed described above for ChIP-seq data. Intervals of enriched transposition events (i.e., ATAC-seq peaks) were determined using MACS2, during which paired reads were treated as separate, independent reads, and fragment densities were determined genome-wide using 32 bp bins. ATAC-seq data was normalized between samples by randomly down-sampling the number of tags to that of the sample with the fewest tags, which in this case was the unextracted sample with 55,337,325 tags.

ATAC-seq footprinting analysis was conducted using TOBIAS v.0.12.4 (Bentsen et al., 2020). TOBIAS ATACorrect was run on both the unextracted and post-extraction venom gland samples using default parameters and a file of blacklisted regions defined as any region in which enriched peaks of ATAC-seq read density were identified in both venom samples and skin samples (i.e., pile-ups of ATAC-seq reads likely to be spurious, or associated with constitutively and non-tissue-specific open chromatin). To calculate footprint scores, TOBIAS ScoreBigwig was run for unextracted and post-extraction venom gland samples using the bias-corrected output from ATACorrect and a shared peak intervals file comprised of all peaks present in both venom gland ATAC-seq samples, with overlapping peak regions first merged using bedtools merge. Differential footprinting analysis was then performed using TOBIAS BINDetect using the JASPAR 2020 Core Vertebrates Non-Redundant TFBS motif database (Sandelin et al., 2004) and default parameters.

ATAC-seq data, as well as ChIP-seq, RNA-seq, and Hi-C based inferences (chromatin loops, TADs), were visualized using Integrated Genomics Viewer v2.8.7 (Robinson et al., 2011).

Identifying candidate transcription factors

We constructed a candidate set of TFs with evidence for activity in the venom gland during venom production through independent analysis of gene expression, ChIP-seq, and ATAC-seq datasets. To first identify annotated TFs in the Prairie Rattlesnake genome, we downloaded Uniprot (UniProt Consortium, 2019) gene lists that were annotated with one or more of the following gene ontology terms: DNA binding transcription factor activity, protein binding transcription factor activity, and transcription factor co-regulatory activity. These gene lists were used to parse the Prairie Rattlesnake gene annotation for known TFs using previously published rattlesnake-to-human orthology tables (Perry et al., 2020). Resulting rattlesnake TFs were cross-referenced with the results of differential gene expression analyses described above to identify TFs with evidence of upregulation in the venom gland during venom production (i.e. higher expression at 1DPE compared to unextracted, IHW p-value < 0.05). Separately, we identified TF genes that are associated with super-enhancers (SEs), regions of elevated H3K27ac ChIP-seq read density (described above). A TF gene was considered to be SE-associated if it overlapped with an annotated SE region, or was the nearest gene to a SE that does not overlap with any annotated genes. Third, as described above, differential binding analysis based on TFBS footprints in ATAC-seq data was used to identify TFs with evidence for increased binding in the post-extraction venom gland compared to unextracted. As this TFBS-based approach does not take into consideration the annotated location or expression of each TF gene, the full JASPAR 2020 Core Vertebrates Non-Redundant TFBS motif database was used for this analysis.

Overlap between these three independently-derived candidate TF sets was assessed and visualized using the UpSetR v1.4.0 (Conway et al., 2017) package in R, and these lists were merged to form one master set of candidate TFs for subsequent analyses. To characterize candidate TFs, we used WebGestalt 2019 (Liao et al., 2019) to identify GO Terms and KEGG Pathways with overrepresentation in our candidate set compared to a background of all TFs annotated in the rattlesnake genome and default parameters. To assess known involvement of our candidate TFs with ERK, a central regulatory molecule within the ERK/MAPK signaling pathway previously implicated in venom gene regulation, we used StringDB v11.0 (Szklarczyk et al., 2019) to identify interactions between candidate TFs and ERK/MAPK1, filtering to include only interactions from curated databases or that were experimentally determined. Resulting interactions were visualized using Cytoscape v3.8.2 (Shannon et al., 2003). A custom binding site motif database for candidate TFs was then created by filtering the JASPAR 2020 Core Vertebrates Non-Redundant TFBS motif database to only include motifs corresponding to our candidate TFs.

Identifying promoters and relevant promoter regions for manually-annotated venom genes.

For all genes except those belonging to the venom families discussed below, the promoter region of a given gene was defined as 1kb upstream of the transcription start site (TSS) through the TSS. Genes within the SVMP, SVSP, and PLA_2 venom gene clusters were originally annotated manually using FGENESH+ (Solovyev, 2004) and known guide protein sequences (see Schield et al., 2019). Because FGENESH+ does not take into account gene expression data, rather than identifying a TSS it instead attempts to identify a likely TATA box based on nucleotide sequence. Thus the “TSS” position labeled by FGENESH+ may not actually represent the true TSS. In order to focus on a region most likely to represent the transcription start site and adjacent sequence for these genes, we defined promoter ATAC-seq peak (PAP) regions by taking 1kb in either direction from the FGENESH+ TSS location (2kb region total) and identifying ATAC-seq peak regions that overlap with this window using bedtools intersect. In the event that more than one ATAC-seq peak was found within a region, the most downstream peak (relative to the associated gene) was taken to be the PAP. Nucleotide sequences for promoters and PAPs were then extracted from the Prairie Rattlesnake genome using bedtools getfasta.

Identification of putative enhancer regions (PERs) and PER-gene interactions.

We used the Activity-by-Contact (ABC) model v0.2 (Fulco et al., 2019) to identify putative enhancer regions (PERs) and infer PER-gene regulatory interactions in the post-extraction venom gland. Candidate enhancer regions were first determined by filtering post-extraction ATAC-seq peaks to exclude regions identified as peaks in the two skin ATAC-seq samples and taking 250bp on either side of the peak summit (position with highest ATAC-seq read density) of these peaks unique to the post-extraction venom gland. Any overlapping peaks were merged into a single region. ABC was then run using an ABC score threshold of 0.05 and otherwise default parameters on these candidate enhancer regions using H3K27ac ChIP-seq density, ATAC-seq density, KR normalized Hi-C data at 5kb resolution, and average gene expression at 1DPE. Venom PERs (vPERs) were defined as resulting PER regions inferred to interact with one or more annotated venom genes. Nucleotide sequences for PERs were then extracted from the Prairie Rattlesnake genome using bedtools getfasta. Enhancer-gene interactions were plotted using ggplot v3.3.3 (Wickham, 2011) and ggforce v0.3.2 (https://github.com/thomasp85/ggforce/) packages in R.

Transcription factor binding site (TFBS) prediction, enrichment analyses, and TFBS alignment

Transcription factor binding site prediction and enrichment analyses were conducted using CIIIDER v0.9 (Gearing et al., 2019) with the default deficit threshold of 0.15 and a gene coverage p-value cutoff of 0.05 and using the custom motif database generated above for candidate TFBS. For TFBS enrichment analyses in venom promoters/PAPs, sequences for a given family were used as the target sequences and compared to promoter sequences for all non-venom genes as a background. Similarly, for vPER TFBS enrichment analyses, all non-venom PERs were used as the background. For use in position score calculations (described below), Tobias ScoreBed was used to annotate TFBS with the mean post-extraction footprint score determined above.

Venom promoter and vPER sequences were aligned using MAFFT v7.475 (Katoh and Standley, 2013) with flags –reorder, --adjustdirectionaccurately, --allowshift, --unalignlevel 0.8, --maxiterate 0, and –globalpair and visualized using the msa package v1.22.0 (Bodenhofer et al., 2015) in R. Locations of enriched TFBS identified in unaligned sequences via CIIIDER were then converted to their corresponding positions in the MAFFT-aligned sequences using a custom python script (See Data Availability). This custom python script also calculates a simple "consensus score" for each alignment, defined as the maximum percent of sequences with an identical nucleotide at a given position in the alignment not including alignment-introduced gaps. TFBS alignments were visualized in R using ggplot2 v3.3.3 (Wickham, 2011). Overlap of enriched TFBS between venom gene families and promoters and enhancers was plotted using the ggVennDiagram v0.5.0 package (https://github.com/gaospecial/ggVennDiagram) in R.

Position score calculations and primary/secondary TFBS analyses

TFBS important for the regulation of a particular venom gene family are expected to be present, conserved, and have evidence of being bound by transcription factors in the promoter regions of highly-expressed venom genes, and absent, altered, and/or inaccessible in the promoters of lowly-expressed genes of the same family. Based this logic, we developed a "position score" that uses binding site conservation, ATAC-seq footprint scores, and gene expression to quantify the degree that a given TFBS adheres to these patterns and may is likely to be important in the regulation of a given family. For a given set of sequences, for example SVMP vPER sequences, the position score of a particular TFBS at a particular position is calculated by taking the sum of the post-extraction footprint score x normalized gene expression at 1DPE for all sequences containing that TFBS at that site, dividing that value by the sum of normalized gene expression at 1DPE for all sequences that do not possess that TFBS at that site, adding 1, and log transforming the resulting value (Supp. Fig. S2b). This results in a high score in instances where a TFBS is conserved at a given position with high footprint scores in highly expressed genes (i.e. important venom genes) and absent in lowly expressed genes (i.e. minor venom genes), low scores when a TFBS is present only in lowly expressed genes, and intermediate scores for situations in which TFBS are present and have a high footprint score in subsets of the highly-expressed genes or in both highly and lowly-expressed genes (Supp. Fig. S2b). For vPERs that are inferred to interact with multiple genes, the mean expression of these genes was used as the associated expression value for that sequence. In a given set of sequences, 'high-scoring' TFBS were defined as those with a position score exceeding the mean of all TFBS position scores in that set of sequences.

TFBS that were enriched in a given set of sequences and had one or more TFBS sites with a high scoring position score were considered "primary" TFBS as they have the most evidence for being important to binding and regulation of that set of sequences. We then identified "secondary" TFBS in other families – TFBS that are present and high scoring in a particular sequence that were not inferred to be enriched in that set of sequences, but were identified as primary in one or more other sets of sequences. For example, if TFBSA is identified as primary in SVMP promoters (enriched in and has a position score > the mean of all enriched TFBS in SVMP promoters) but is not enriched in SVSP promoters, there may still be binding sites for TFBSA in SVSP promoters that could be opportunistically bound assuming that TFBSA is active during venom production. If TFBSA sites are identified in SVSP promoters and have position scores higher than the mean of primary TFBS in the SVSP promoters, it is considered a "secondary" TFBS.

Because the position score method is not applicable to one-off venom genes (constituting the "other" category of venom genes reference throughout), we instead searched promoters and enhancers of these genes for TFBS identified as primary in the promoters or enhancers of one or more of the three major venom gene families, and weighted identified TFBS by post-extraction footprint scores alone.

Novel TFBS motif searches in venom regulatory sequences

We used de novo motif identification analyses in elevated ATAC-seq footprint regions to identify any novel TFBS motifs that would not be otherwise detected by our candidate approach described above. We confined these searches to regions with evidence of being bound by a transcription factor by only searching regions with an average ATAC-seq footprint score greater than or equal to half of the "bound threshold" determined during the BINDetect step of the ATAC-seq footprint analysis described above (Bound threshold = 4.97183; Half threshold = 2.485915). This cutoff was chosen in order to filter out regions of promoter and enhancer sequences with little to no evidence of being bound by a transcription factor, while being permissive to regions with intermediate evidence. These regions were generated by converting the bigwig file of post-extraction ATAC-seq footprint scores to a bedgraph file using bigWigToBedGraph (Kent et al., 2010), filtering out regions with a footprint score less than the half threshold, and merging any remaining regions within 5bp of one another using bedtools merge. Bedtools intersect and getfasta were then used to select elevated footprint regions overlapping with putative enhancers or promoter regions and extract fasta sequences for each region.

Novel motifs were identified and annotated using MEME v5.3.3 (Bailey and Elkan, 1994) and TomTom v5.3.3 (Gupta et al., 2007) within the online MEME-ChIP tool v5.3.3 (Machanick and Bailey, 2011). MEME was run in Differential Enrichment mode using a background of all elevated footprint regions in enhancers or promoters not associated with venom genes, a minimum and maximum motif width of 6 and 15, respectively, the expectation of Zero or One Occurrence Per Sequence (zoops mode), and was set to identify at most 25 motifs. MEME motifs with an e-value < 0.05 were considered significant, and these motifs were compared to motifs in the JASPAR 2020 non-redundant vertebrate motif database using TomTom with default settings. MEME motifs not successfully annotated with TomTom (i.e., E-value > 1;

Gupta et al., 2007) were considered novel motifs. No novel motifs were enriched in venom promoters relative to non-venom gene promoters (Supp. Table S5).

Position-weight matrices of novel motifs were converted from MEME to JASPAR motif format using the UniversalMotif v1.8.3 (Tremblay, 2020) package in R, and CIIIDER was used to scan all input sequences (i.e., elevated footprint regions within vPER regions) with default parameters. MEME motif sites were tallied for each regulatory sequence in R and plotted using ggplot2. To assess the degree to which MEME motifs were similar to candidate TFs specifically, we again used TomTom to compare MEME motifs with annotated TFBS, this time using our custom JASPAR database containing only candidate TFs identified above.

Comparisons of venom regulatory sequences with those of non-venom paralogs.

Non-venom paralog genes were identified previously (Schield et al., 2019). To assess whether any vPER sequences identified for the three major venom gene families were also present near a family's non-venom paralogs, we used BLASTn to identify similar sequences genome-wide using as a query the sequence of the vPER(s) associated with the most highly expressed venom gene in each family. We then surveyed non-venom paralogs and adjacent regions for significant vPER BLAST hits ($e < 0.000001$).

To compare venom gene and non-venom paralog promoters, we scanned non-venom paralog promoters for candidate TFBS using CIIIDER with default settings and our custom candidate TF motif database, and filtered the results to include TFBS implicated in the regulation of the corresponding venom gene family (i.e. "primary" or "secondary" TFBS in the promoters a given venom gene family; see above). We also tested for enrichment of any candidate TFBS in non-venom paralog promoters for each family compared to a background of all promoters (excluding all venom gene and non-venom paralog promoters).

Identifying potential conserved vPER sequences in other venomous snake species.

To investigate whether vPER sequences are conserved in other venomous snakes, we used BLASTn (Altschul et al., 1990) to search all snake nucleotide sequences on NCBI (via the online BLAST platform) and BLASTn in BLAST+ v2.6.0 to search a set of existing snake genome assemblies - *Naja naja* (NCBI: GCA_009733165.1, Suryamohan et al., 2020), *Deinagkistrodon acutus* (Yin et al., 2016), *Thamnophis sirtalis* (NCBI: GCA_001077635.2, Perry et al., 2018), *Protobothrops flavoviridis* (NCBI: GCA_003402635.1, Shibata et al., 2018), and *Python bivittatus* (NCBI: GCA_000186305.2, Castoe et al., 2013). We used as the query sequence the vPER inferred to interact with the highest expressed gene(s) in the SVMP and PLA_2 regions (SVSP was excluded from these analyses for reasons discussed below). The at most five best hits from each species with were selected based on e-value scores. Alignments were generated using MAFFT with parameters described above. For PLA_2 vPER BLAST searches against all snake nucleotide sequences on NCBI, a subset of returned hits were small (i.e., covered less than 25% of the query sequence) and only hit to regions on the extremities of the query vPER sequence with no similarity to the center of the vPER sequence where functionally-relevant TFBS are inferred to be located; these sequences were manually removed from alignments. “Primary” TFBS for SVMP and PLA_2 enhancers were scanned using CIIIDER with default parameters and visualized in R using ggplot2. An approximated phylogeny for lineages represented in these analyses was downloaded from TimeTree (Kumar et al., 2017).

Analyses of transposable elements (TEs) associated with SVSP regulatory sequences.

Given the fact that BLASTs of SVSP vPER sequences yielded a large number of hits throughout the Prairie Rattlesnake genome, we investigated this could be explained by an association between these sequences and transposable elements (TEs). Using TE annotations from (Schield et al. 2019), we used Giggle v0.6.3 (Layer et al., 2018) to test whether SVSP regulatory regions (promoters and vPERs) were significantly enriched for overlap with any particular TE (one-tailed Fisher’s exact test; $p < 0.05$). This analysis identified a DNA/hAT-Tip100 element (Cv1-hAT-Tip100) that was enriched in the SVSP regulatory regions and generally common on chromosome 10. A genome-wide consensus sequence for this element was generated using mafft by providing the DNA hAT-Tip100 consensus from the repeat element library as reference. This preliminary alignment was then manually curated by removing the DNA hAT-Tip100 reference, re-aligning the genomic copies, and removing major regions with limited coverage by using the Gblocks server v0.91b (Talavera and Castresana, 2007). The final consensus sequence was derived by using the Unipro UGENE software (Rose et al., 2019). This consensus sequence was then used to calculate sequence divergence (pairwise-pi) for all Cv1-hAT-Tip100. For this calculation, we excluded alignment positions where the highest nucleotide frequency exceeded 0.7. Using these pairwise-pi values, we estimated TE age as pi ÷ 2 x 2.4×10^9 following (Pasquesi et al., 2018). For Cv1-hAT-Tip100 copies within the SVSP region, including those in regulatory and "other" intergenic sequences, we used CIIIDER to identify TFBS identified above as "primary" in SVSP promoters and/or enhancers. ATAC-seq footprint scores were calculated for TFBS sites using TOBIAS ScoreBed. These sequences and TFBS positions were then aligned using mafft and the custom python script described above, and plotted in R using ggplot2. Aligned TFBS positions were visualized using ggplot2 in R and inset detailed alignments were generated using Jalview v2.10.1 (Waterhouse et al., 2009).

Identification of exonic debris in the PLA_2 gene cluster.

We used BLAST feature of ncbi-blast v.2.7.1+ suite (tblastx, e-value of 0.01, default restrictions on word count and gaps) to perform initial search of the exons against a pre-compiled database of exons previously successfully used to annotate PLA_2GIIe family of genes in vertebrate animals (Koludarov et al., 2020). We then manually assessed each result and established exon boundaries using Geneious v11 (https://www.geneious.com).

Data Availability

Transcriptomic data for the Prairie Rattlesnake venom gland and body tissues are accessioned at the NCBI Sequence Read Archive (NCBI BioProject: PRJNA716163). Raw and processed ChIP-seq and ATAC-seq data are available at the NCBI Gene Expression Omnibus (NCBI: GSE169217). Previously-generated Hi-C data are available on NCBI SRA (NCBI BioProject: PRJNA413201).

Acknowledgments

Support was provided by a National Science Foundation grants DEB-1655571 (to TAC and SPM), IOS-655735 to TAC, and DBI-1906188 to DRS, NIH 1R35GM128822 to EBC, and a Research Dissemination and Faculty Development grant from the University of Northern Colorado to SPM.

Figures

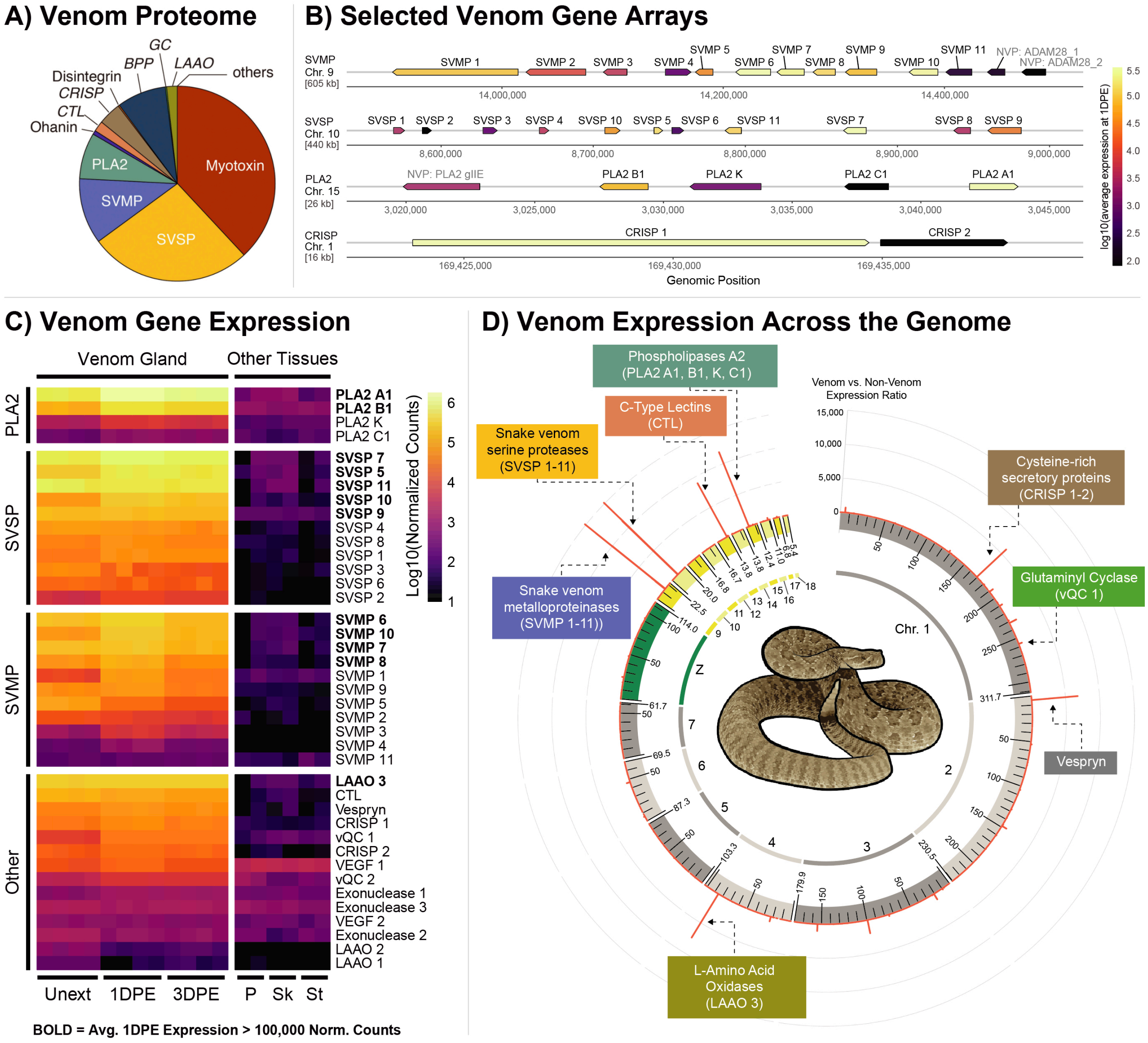
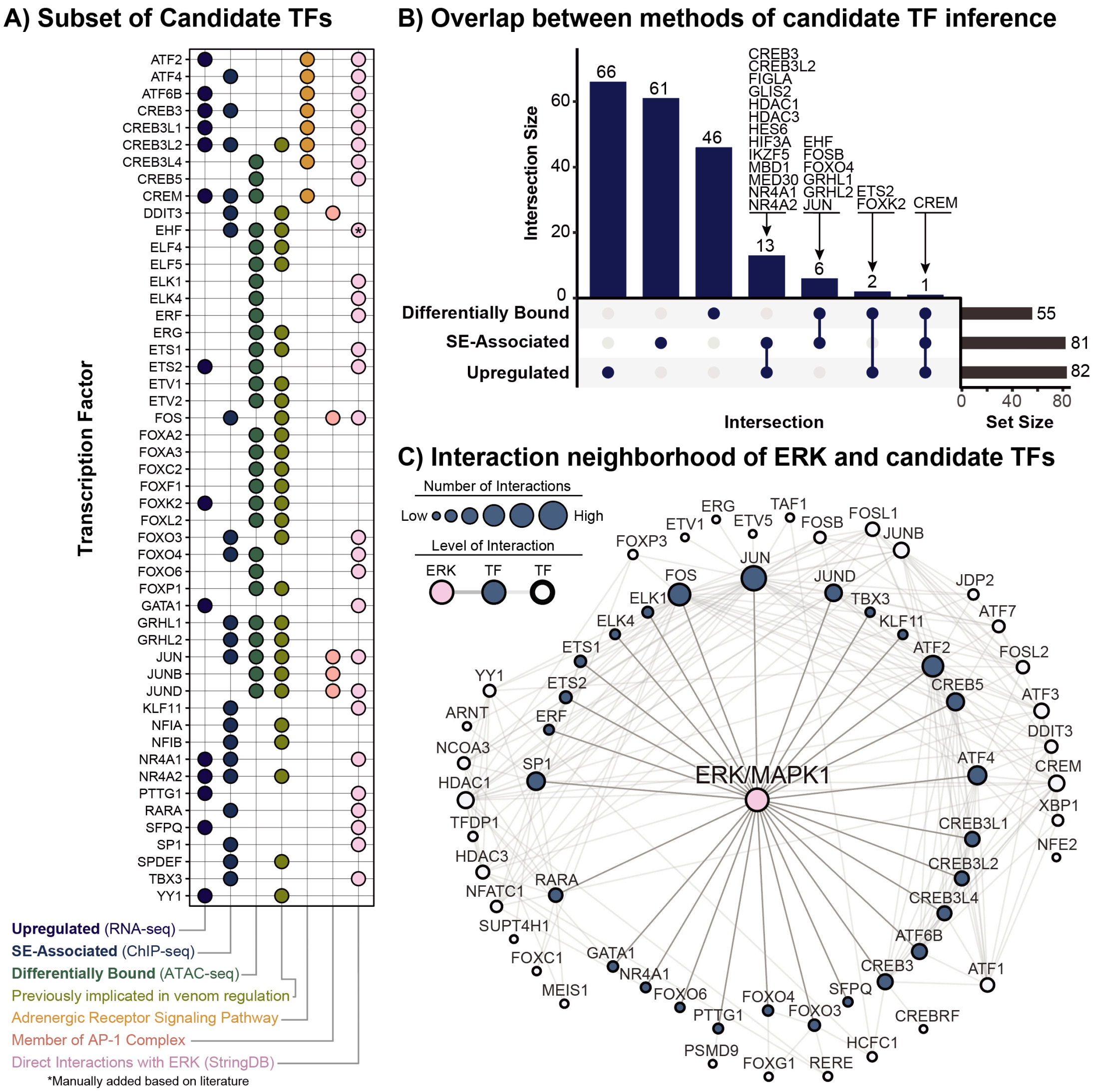


Figure 1. Venom gene expression and structure in the Prairie Rattlesnake. A) Proportions of venom protein components in Prairie Rattlesnake venom (adapted from Schield et al., 2019). **B**) Structure of tandemly-duplicated venom gene families, with genes colored by 1DPE expression. **C**) Venom gene expression in unextracted (Unext), 1DPE, and 3DPE venom gland, and in body tissues (P: pancreas, Sk: skin, St: stomach). **D**) Venom gene expression visualized on the Prairie Rattlesnake genome, with red lines indicating the ratio of average venom gland expression compared to average body tissue expression.



A) Subset of Candidate TFs

B) Overlap between methods of candidate TF inference

C) Interaction neighborhood of ERK and candidate TFs

Figure 2. Candidate transcription factors (TFs) underlying venom gene regulation. A) Candidate TFs involved in venom regulation, with the first three columns representing three approaches for TF identification, and the last three columns indicating TF membership in functional categories. **B**) Numbers of candidate TFs identified by analyses of gene expression (Upregulated), association with super-enhancers (SE-Associated), or differential ATAC-seq footprinting (Differentially Bound). Vertical bars showing the numbers of TFs identified uniquely or in a combination of analyses, as indicated by the dots below. Horizontal bar plots indicate the total number of TFs identified. **C**) Interactions between candidate TFs and ERK (MAPK1) based on StringDB. TFs with direct interactions with ERK are shown in blue, and TFs that interact with these are shown in white. Node sizes are scaled by the number of interactions with other TFs in the network.

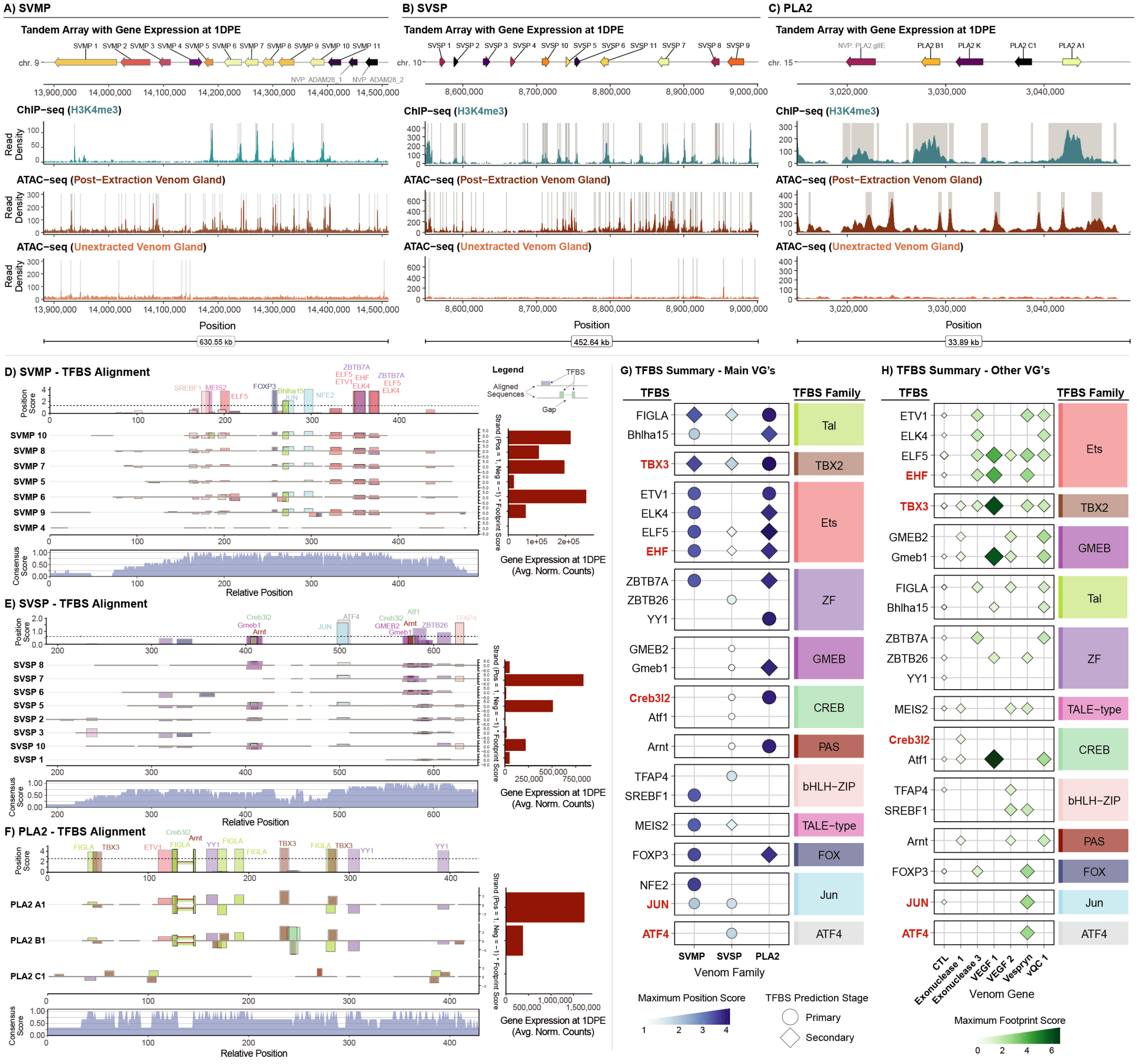

A) SVMP
Tandem Array with Gene Expression at 1DPE
SVMP 1 SVMP 2 SVMP 3 SVMP 4 SVMP 5 SVMP 6 SVMP 7 SVMP 8 SVMP 9 SVMP 10 SVMP 11
chr. 9
NVP: ADAM28_1 NVP: ADAM28_2
13,900,000 14,000,000 14,100,000 14,200,000 14,300,000 14,400,000 14,500,000
B) SVSP
Tandem Array with Gene Expression at 1DPE
SVSP 1 SVSP 2 SVSP 3 SVSP 4 SVSP 10 SVSP 5 SVSP 6 SVSP 11 SVSP 7 SVSP 8 SVSP 9
chr. 10
8,600,000 8,700,000 8,800,000 8,900,000 9,000,000
C) PLA2
Tandem Array with Gene Expression at 1DPE
NVP: PLA2 gIIE PLA2 B1 PLA2 K PLA2 C1 PLA2 A1
chr. 15
3,020,000 3,030,000 3,040,000
ChIP−seq (H3K4me3)
Read Density
ATAC−seq (Post−Extraction Venom Gland)
ATAC−seq (Unextracted Venom Gland)
Position
630.55 kb
452.64 kb
33.89 kb
D) SVMP - TFBS Alignment
Position Score
SREBF1 MEIS2 ELF5 FOXP3 Bhlha15 JUN NFE2 ZBTB7A ELF5 ETV1 EHF ELK4 ZBTB7A ELF5 ELK4
SVMP 10
SVMP 8
SVMP 7
SVMP 5
SVMP 6
SVMP 9
SVMP 4
Consensus Score
Relative Position
Strand (Pos = 1, Neg = −1) * Footprint Score
Gene Expression at 1DPE (Avg. Norm. Counts)
0 1e+05 2e+05
Legend
TFBS
Aligned Sequences
Gap
E) SVSP - TFBS Alignment
Creb3l2 Gmeb1 Arnt JUN ATF4 Creb3l2 GMEB2 Gmeb1 Arnt Atf1 ZBTB26 TFAP4
SVSP 8
SVSP 7
SVSP 6
SVSP 5
SVSP 2
SVSP 3
SVSP 10
SVSP 1
0 250,000 500,000 750,000
F) PLA2 - TFBS Alignment
FIGLA TBX3 ETV1 Creb3l2 Arnt FIGLA YY1 FIGLA FIGLA TBX3 FIGLA TBX3 YY1 YY1
PLA2 A1
PLA2 B1
PLA2 C1
0 500,000 1,000,000 1,500,000
G) TFBS Summary - Main VG's
TFBS
TFBS Family
FIGLA Bhlha15 — Tal
TBX3 — TBX2
ETV1 ELK4 ELF5 EHF — Ets
ZBTB7A ZBTB26 YY1 — ZF
GMEB2 Gmeb1 — GMEB
Creb3l2 Atf1 — CREB
Arnt — PAS
TFAP4 SREBF1 — bHLH−ZIP
MEIS2 — TALE−type
FOXP3 — FOX
NFE2 JUN — Jun
ATF4 — ATF4
SVMP SVSP PLA2
Venom Family
Maximum Position Score
1 2 3 4
TFBS Prediction Stage
Primary
Secondary
H) TFBS Summary - Other VG's
TFBS
TFBS Family
ETV1 ELK4 ELF5 EHF — Ets
TBX3 — TBX2
GMEB2 Gmeb1 — GMEB
FIGLA Bhlha15 — Tal
ZBTB7A ZBTB26 YY1 — ZF
MEIS2 — TALE−type
Creb3l2 Atf1 — CREB
TFAP4 SREBF1 — bHLH−ZIP
Arnt — PAS
FOXP3 — FOX
JUN — Jun
ATF4 — ATF4
CTL Exonuclease 1 Exonuclease 3 VEGF 1 VEGF 2 Vespryn vQC 1
Venom Gene
Maximum Footprint Score
0 2 4 6

Figure 3 (previous page). Chromatin and transcription factor binding site characteristics of venom gene promoter regions. A-C) Venom gene clusters with gene expression indicated by color (brighter colors are more highly expressed), with ChIP-seq and ATAC-seq data with peaks indicated in gray. **D-F**) Promoter alignments of major venom gene clusters with TFBS inferences in ATAC-seq peaks. Colored vertical bars on each promoter indicate presence of a TFBS for an enriched TF (bar size scaled by ATAC-seq footprint score, and the TFBS orientation indicated by its position above or below the line). Alignment gaps shown by faded regions of center lines, and colored lines connecting TFBS bars indicate TFBS that span gaps. Vertical bar plots on each panel show TFBS position score, with mean gene family score indicated with dotted line. **G**) TFs associated with TFBS enriched in a venom gene family ("primary") with a position score above the family median. "Secondary" TFBS are enriched in at least one other family, with binding site position scores above the median threshold. If multiple TFBS were identified per family, the maximum score is shown. **H**) Post-extraction ATAC-seq footprint scores for TFBS in the promoters of "other" venom genes. TFBS shown in red in **G** and **H** have known interactions with ERK signaling.

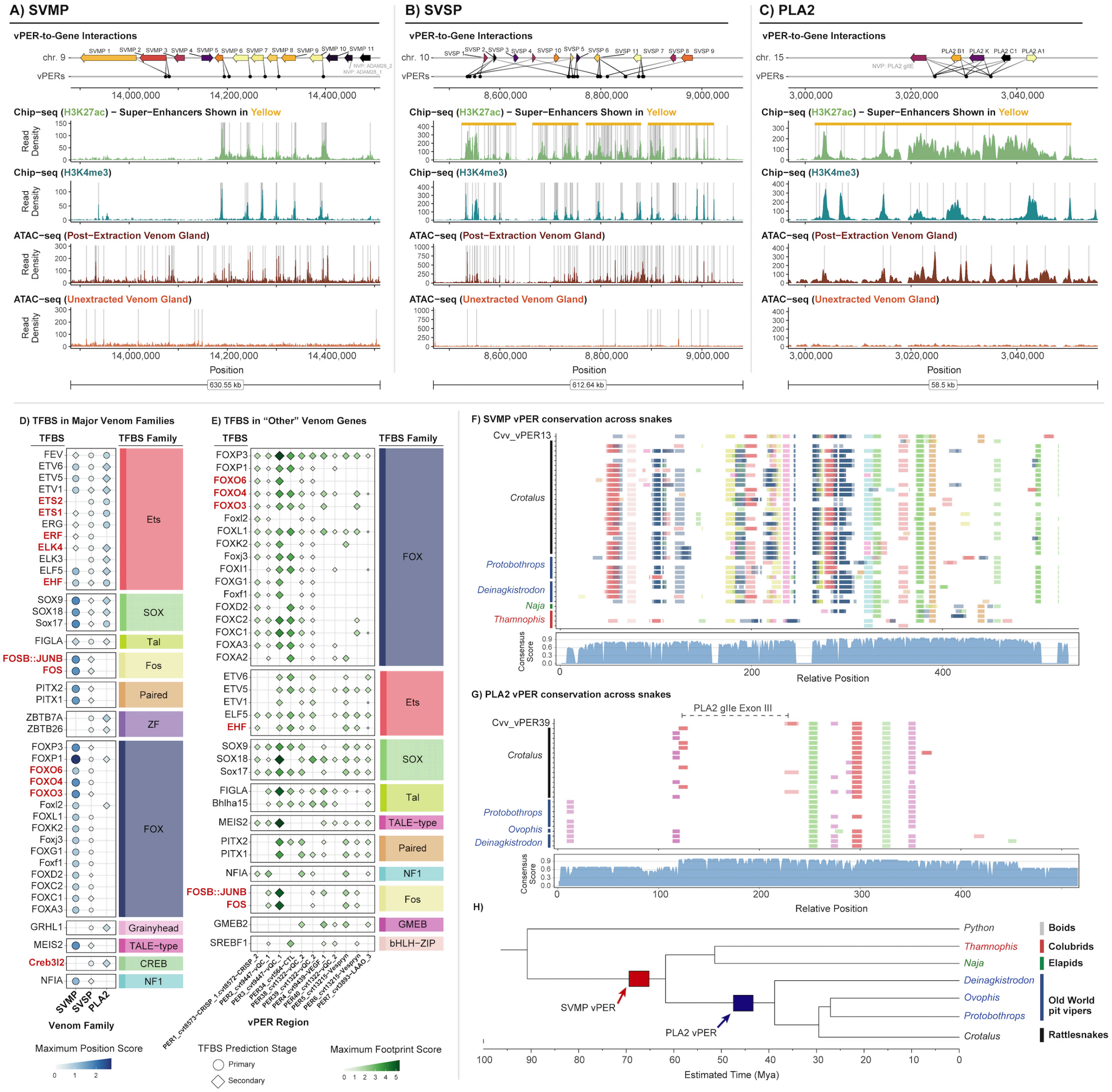

A) SVMP
B) SVSP
C) PLA2
vPER-to-Gene Interactions
chr. 9
vPERs
SVMP 1
SVMP 2
SVMP 3
SVMP 4
SVMP 5
SVMP 6
SVMP 7
SVMP 8
SVMP 9
SVMP 10
SVMP 11
NVP: ADAM28_2
NVP: ADAM28_1
14,000,000
14,200,000
14,400,000
chr. 10
SVSP 1
SVSP 2
SVSP 3
SVSP 4
SVSP 10
SVSP 5
SVSP 6
SVSP 11
SVSP 7
SVSP 8
SVSP 9
8,600,000
8,800,000
9,000,000
chr. 15
NVP: PLA2 gIIE
PLA2 B1
PLA2 K
PLA2 C1
PLA2 A1
3,000,000
3,020,000
3,040,000
Chip−seq (H3K27ac) − Super−Enhancers Shown in Yellow
Chip−seq (H3K4me3)
ATAC−seq (Post−Extraction Venom Gland)
ATAC−seq (Unextracted Venom Gland)
Read Density
Position
630.55 kb
612.64 kb
58.5 kb
D) TFBS in Major Venom Families
TFBS
TFBS Family
FEV
ETV6
ETV5
ETV1
ETS2
ETS1
ERG
ERF
ELK4
ELK3
ELF5
EHF
SOX9
SOX18
Sox17
FIGLA
FOSB::JUNB
FOS
PITX2
PITX1
ZBTB7A
ZBTB26
FOXP3
FOXP1
FOXO6
FOXO4
FOXO3
Foxl2
FOXL1
FOXK2
Foxj3
FOXG1
Foxf1
FOXD2
FOXC2
FOXC1
FOXA3
GRHL1
MEIS2
Creb3l2
NFIA
Ets
SOX
Tal
Fos
Paired
ZF
FOX
Grainyhead
TALE−type
CREB
NF1
SVMP
SVSP
PLA2
Venom Family
Maximum Position Score
0
1
2
E) TFBS in "Other" Venom Genes
FOXP3
FOXP1
FOXO6
FOXO4
FOXO3
Foxl2
FOXL1
FOXK2
Foxj3
FOXI1
FOXG1
Foxf1
FOXD2
FOXC2
FOXC1
FOXA3
FOXA2
ETV6
ETV5
ETV1
ELF5
EHF
SOX9
SOX18
Sox17
FIGLA
Bhlha15
MEIS2
PITX2
PITX1
NFIA
FOSB::JUNB
FOS
GMEB2
SREBF1
GMEB
bHLH−ZIP
PER1_cvt8573−CRISP_1
cvt8572−CRISP_2
PER2_cvt9447−vQC_1
PER3_cvt9447−vQC_1
PER34_cvt564−CTL
PER38_cvt1322−vQC_2
PER39_cvt1322−vQC_2
PER4_cvt9439−VEGF_1
PER5_cvt1322−vQC_2
PER6_cvt1321−Vespryn
PER7_cvt3893−LAAO_3
vPER Region
TFBS Prediction Stage
Primary
Secondary
Maximum Footprint Score
0 1 2 3 4 5
F) SVMP vPER conservation across snakes
Cvv_vPER13
Crotalus
Protobothrops
Deinagkistrodon
Naja
Thamnophis
Consensus Score
0.9
0.6
0.3
0.0
0
200
400
Relative Position
G) PLA2 vPER conservation across snakes
PLA2 gIIe Exon III
Cvv_vPER39
Crotalus
Protobothrops
Ovophis
Deinagkistrodon
0
100
200
300
400
Relative Position
H)
Python
Thamnophis
Naja
Deinagkistrodon
Ovophis
Protobothrops
Crotalus
Boids
Colubrids
Elapids
Old World pit vipers
Rattlesnakes
SVMP vPER
PLA2 vPER
100
90
80
70
60
50
40
30
20
10
0
Estimated Time (Mya)

Figure 4 (previous page). Functional genomics data identify conserved venom gene enhancer sequences and associated transcription factor binding sites. A-C) Venom gene clusters with individual gene expression levels indicated by color per gene (brighter colors are more highly expressed), with ChIP-seq (H3K27ac and H3K4me3) and ATAC-seq data with peaks indicated in gray. **D**) Primary and secondary TFBS in vPERs for the three major venom gene clusters (shown if present in >1 family). **E**) Secondary TFBS in vPERs of "other" venom genes (shown if present in >3 vPERs). TFBS shown in red in **D** and **E** have known interactions with ERK signaling. **F-H**) Alignment and conservation of PERs among species of venomous snakes for SVMP (**F**) and PLA_2 (**G**) PERs. **H**) Tree indicating divergence of venomous snake lineages and conservation of SVMP and PLA_2 PERs.

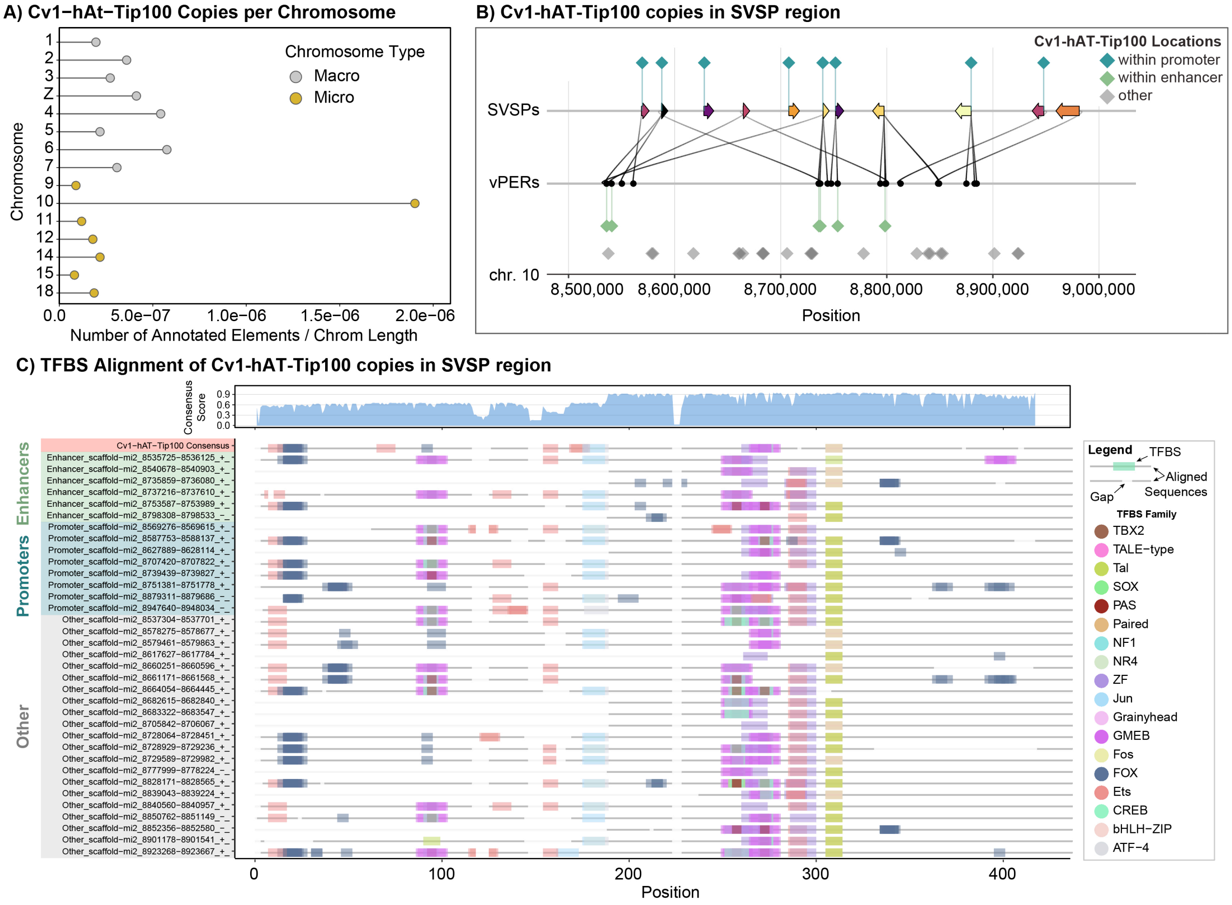


Figure 5. DNA transposons have re-wired SVMP venom cluster regulatory networks. **A**) Cv1-hAT-Tip100 copies per chromosome in the Prairie Rattlesnake, normalized by chromosome length. **B**) SVSP gene array and vPER inferences, with Cv1-hAT-Tip100 copies shown as colored diamonds. **C**) Alignment of SVSP-local Cv1-hAT-Tip100 copies with the genome-wide consensus. TFBS enriched in SVSP promoters or enhancers are colored based on TF families. Faded regions represent alignment gaps.

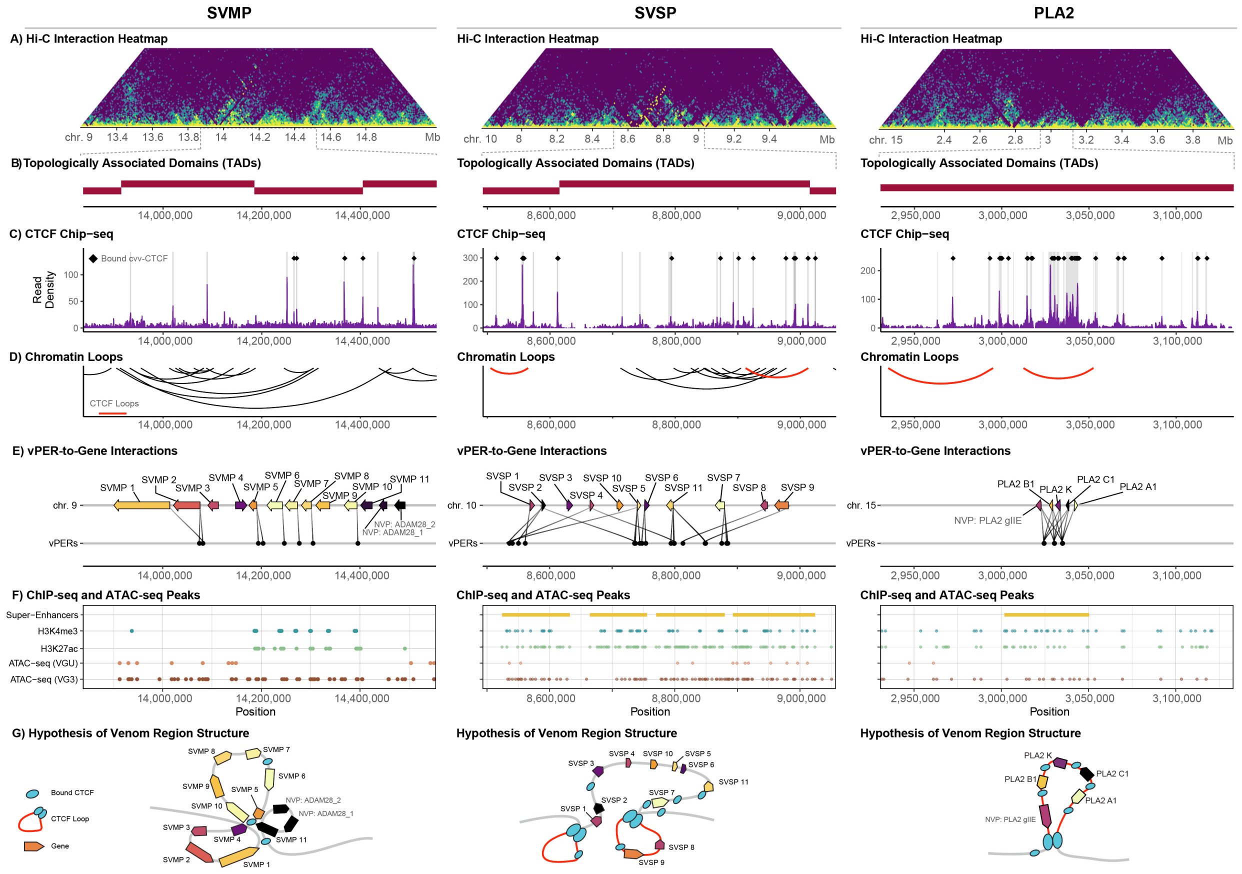


Figure 6. Chromatin structure and organization associated with venom gene arrays. A) Hi-C interaction heatmap (10kb resolution) of 2Mb regions centered on venom gene arrays. Brighter colors indicate higher contact frequency. **B**) Topologically associated domains (TADs) across venom gene arrays. **C**) CTCF ChIP-seq density, with grey vertical lines indicating ChIP-seq peaks and diamonds indicating peaks centered on a verified CTCF binding sites. **D**) Chromatin loops inferred from Hi-C data that span venom gene arrays. Red loops indicate likely CTCF-CTCF bound loops, defined by the presence of a CTCF ChIP-seq peak centered around a CTCF motif within 10kb of chromatin loop ends. **E**) Venom array genes and inferred PER-promoter interactions. **F**) Simplified ChIP-seq and ATAC-seq data, with points indicating the presence of ChIP/ATAC-seq peaks and yellow bars indicating super-enhancers. **G**) Hypotheses for three-dimensional loop structures of venom gene regions.

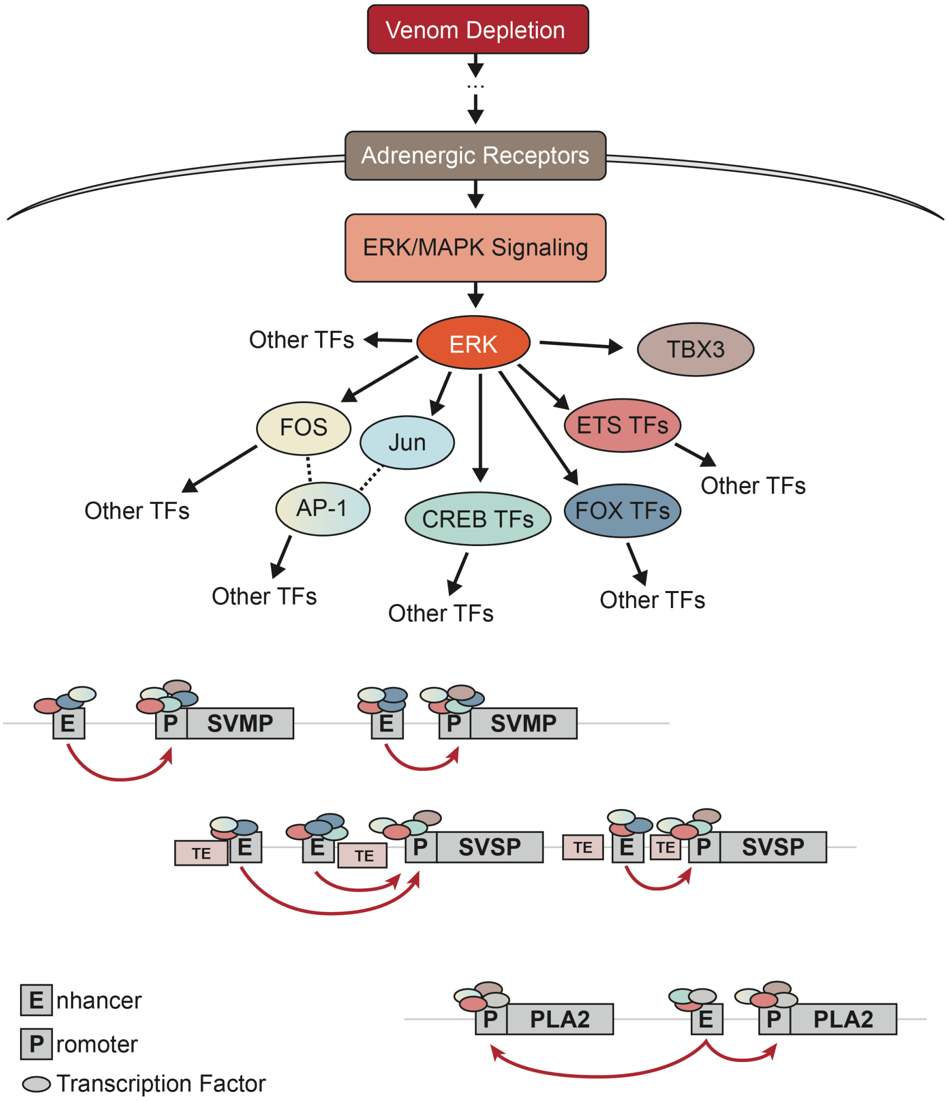
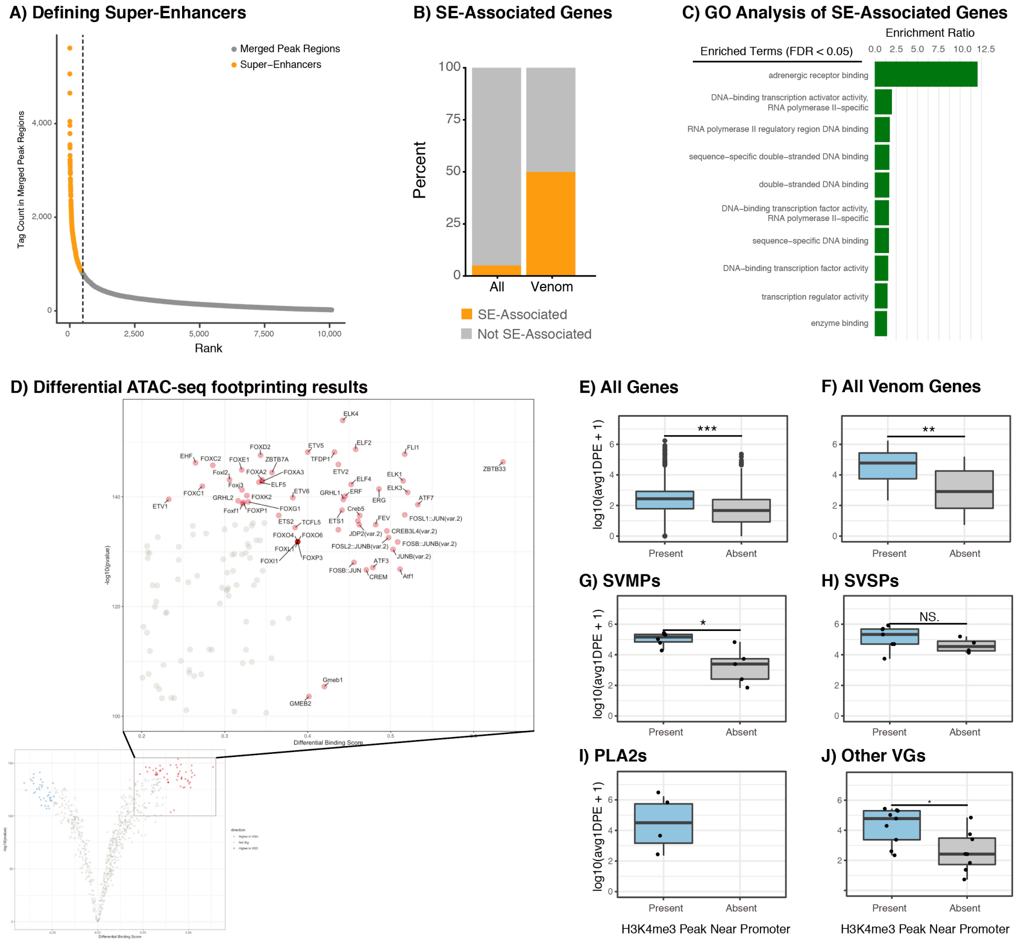
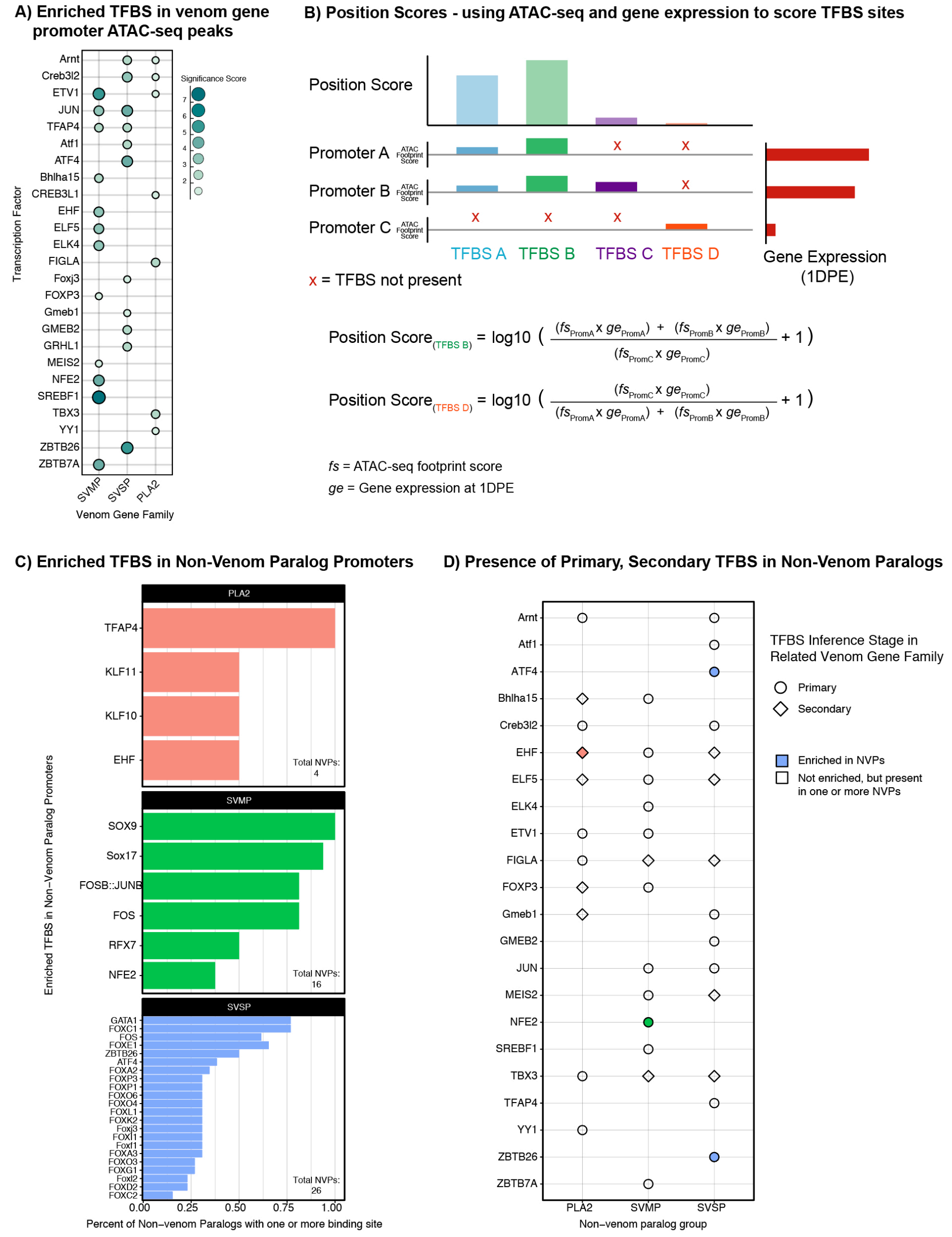


Figure 7. Model of venom signaling and regulatory network. Hypothesis regulatory network that controls venom regulation based on findings presented herein. Red arrows indicate enhancer-promoter interactions.

Supplementary Figures



Supplementary Figure S1. A) Rank-intensity plot of merged H3K27ac ChIP-seq peak regions used to define super-enhancers (regions with top 5% highest ChIP-seq intensity). B) Proportion of genes associated with (within or nearest-to) super-enhancers compared between the all annotated genes ('All') and venom genes ('Venom'). C) Gene ontology overrepresentation analysis results of SE-associated genes compared to a background of all annotated genes in the Prairie Rattlesnake genome. All terms shown are significantly overrepresented in SE-associated genes (FDR < 0.05). D) Transcription factors inferred to have higher (red) and lower (blue) binding activity in the post-extraction venom gland based on ATAC-seq footprint quantity and strength (bottom left). Inset plot shows detail of transcription factors with higher activity during venom production. E-J) Comparisons of normalized gene expression at 1 day post-extraction (1DPE) between genes with and without an H3K4me4 ChIP-seq peak within 1kb of the promoter for E) all genes, F) all venom genes lumped together, G) SVMP, H) SVSP, I) PLA_2, and J) other venom genes (*: p-value < 0.05, ***: p-value < 0.001, NS: not significance; Student's t-test).



A) Enriched TFBS in venom gene promoter ATAC-seq peaks

B) Position Scores - using ATAC-seq and gene expression to score TFBS sites

$$\text{Position Score}_{(\text{TFBS B})} = \log10\left(\frac{(fs_{\text{PromA}} \times ge_{\text{PromA}}) + (fs_{\text{PromB}} \times ge_{\text{PromB}})}{(fs_{\text{PromC}} \times ge_{\text{PromC}})} + 1\right)$$

$$\text{Position Score}_{(\text{TFBS D})} = \log10\left(\frac{(fs_{\text{PromC}} \times ge_{\text{PromC}})}{(fs_{\text{PromA}} \times ge_{\text{PromA}}) + (fs_{\text{PromB}} \times ge_{\text{PromB}})} + 1\right)$$

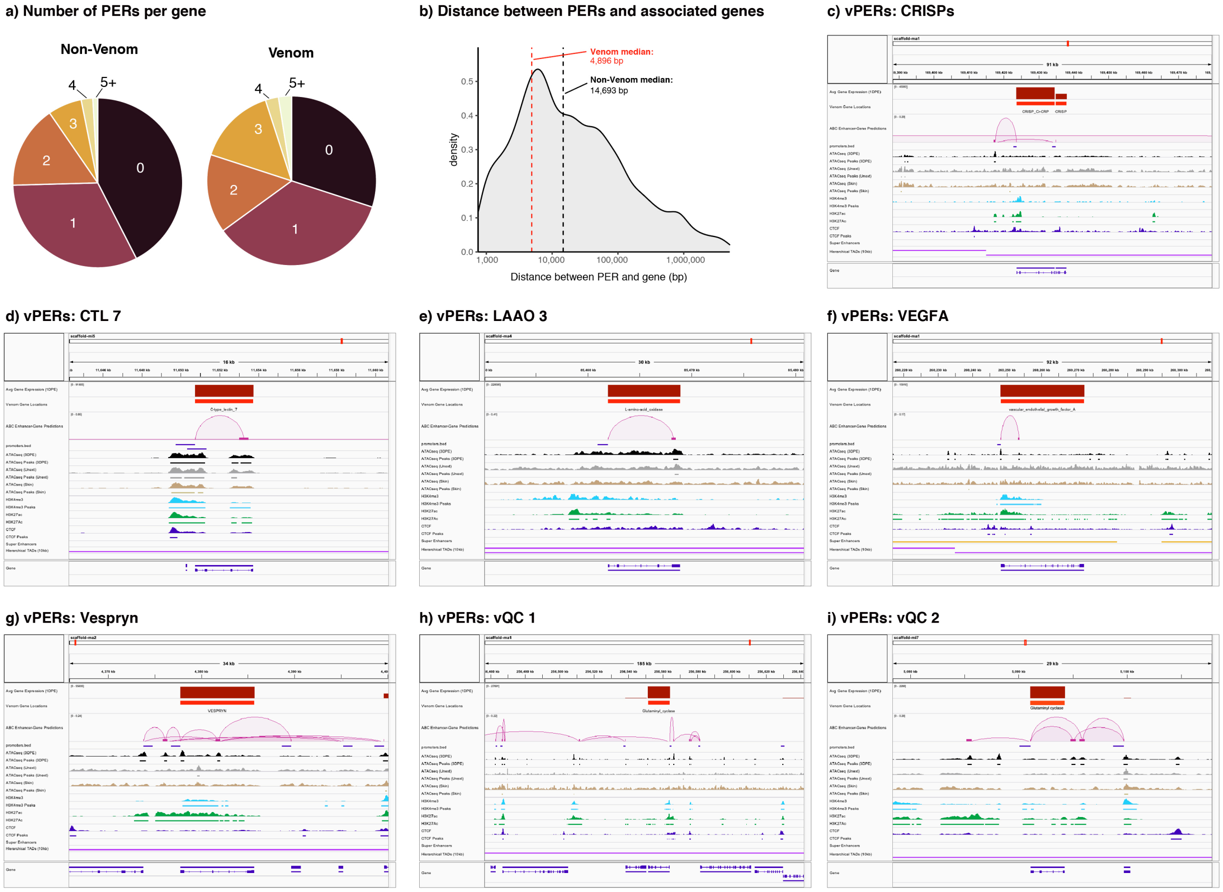
fs = ATAC-seq footprint score

ge = Gene expression at 1DPE

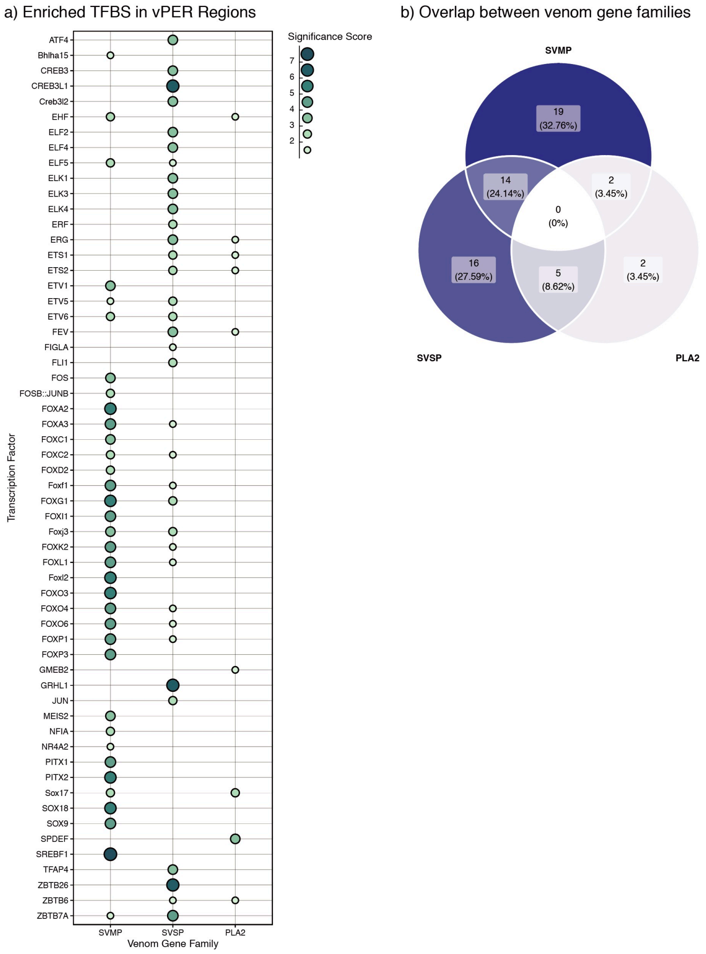
C) Enriched TFBS in Non-Venom Paralog Promoters

D) Presence of Primary, Secondary TFBS in Non-Venom Paralogs

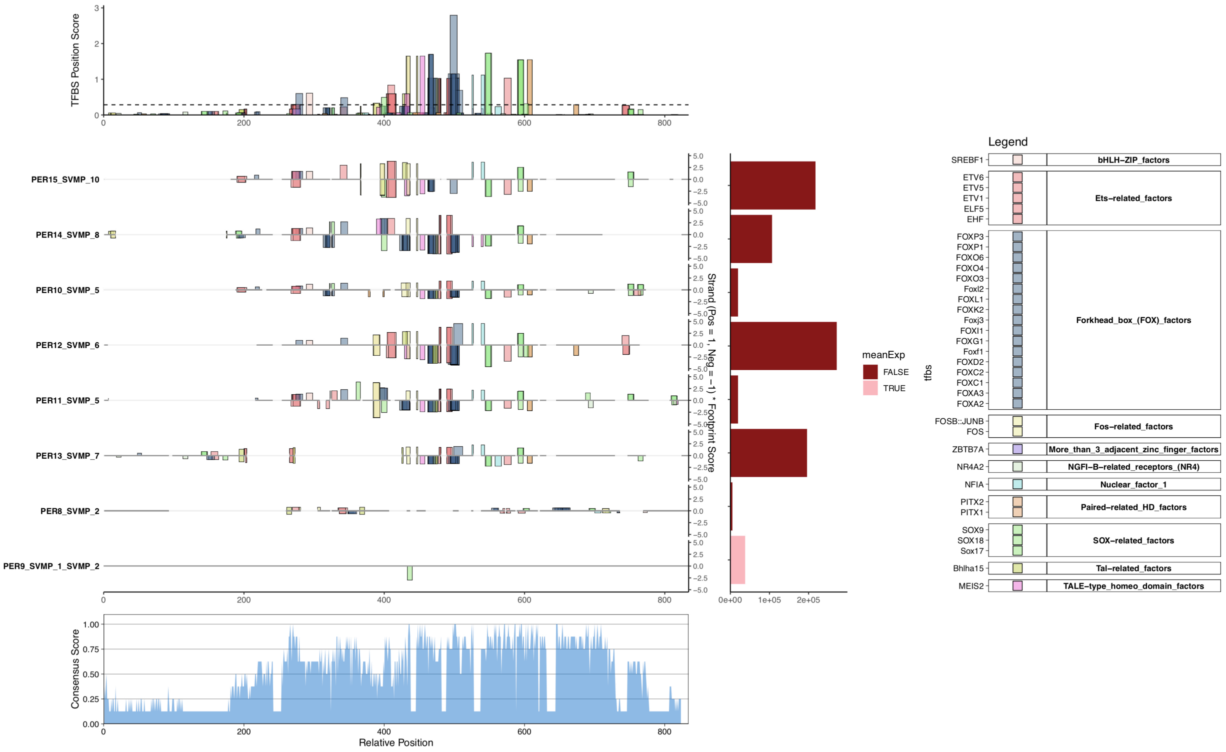
Supplementary Figure S2 (previous page). A) Dots indicate enrichment in the promoter ATAC-seq peaks of a given venom gene family ($p < 0.05$), with color and size showing the significance score (-Log10(p-value) multiplied by sign of log2 enrichment). B) Overview of Position Score basis and calculation. For TFBS that are conserved between two or more sequences in a multiple sequence alignment, the position score first weights the expression of associated genes by the ATAC-seq footprint score of that TFBS (i.e., strength of evidence that that site is bound by a TF), sums these weighted expression values for all genes with that TFBS at that site, and divides that value by the sum for all genes without that TFBS predicted at that site. This score is designed such that TFBS that are shared between highly expressed genes will be scored highly (TFBS A and B), those shared between only lowly expressed genes will be scored lowly (TFBS D), and those with intermediate patterns will have intermediate scores (TFBS C). C) TFBS with significant enrichment ($p < 0.05$) in the promoter regions of NVPs related to PLA_2, SVMP, and SVSP venom genes. All TFBS are enriched, and the x-axis shows the percent of NVPs in each group with one or more of each TFBS. The total number of NVPs in each group is inset in the bottom right. D) For each NVP group, presence (TFBS in one or more NVP in group; empty points) and enrichment (filled points) is shown for primary and secondary TFBS identified in the corresponding venom gene group (see main text Fig. 4g).



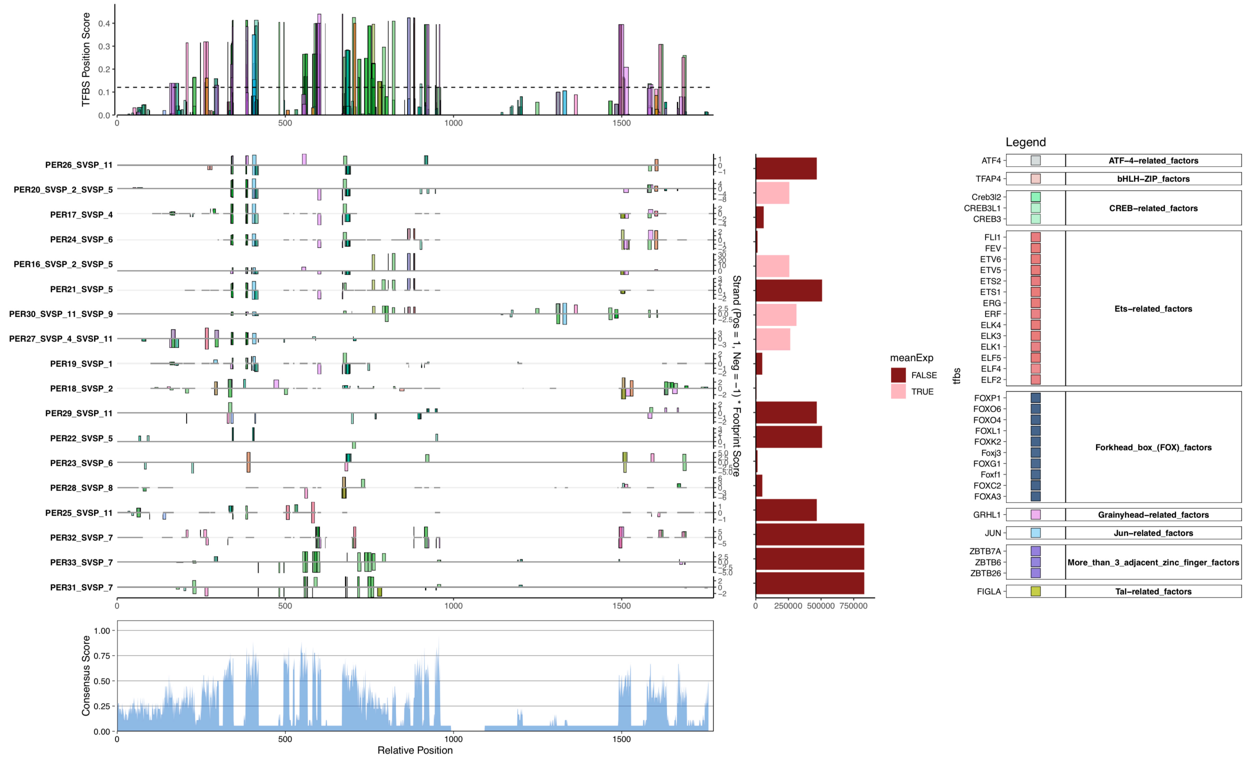
Supplementary Figure S3: A) Comparison of the number of predicted enhancer regions (PERs) per gene for venom and non-venom genes. B) Density plot showing distances between PERs and associated genes for venom and non-venom genes, with median distance shown for each. C-I) vPER predictions for "other" venom genes in the Prairie Rattlesnake. vPER inferences are shown in the "ABC Enhancer-Gene Predictions" track; arcs begin at the promoter and end at the inferred vPER region (marked with a thicker purple bar). For ATAC-seq and ChIP-seq data, peak regions are marked with a bar underneath the read density plots.



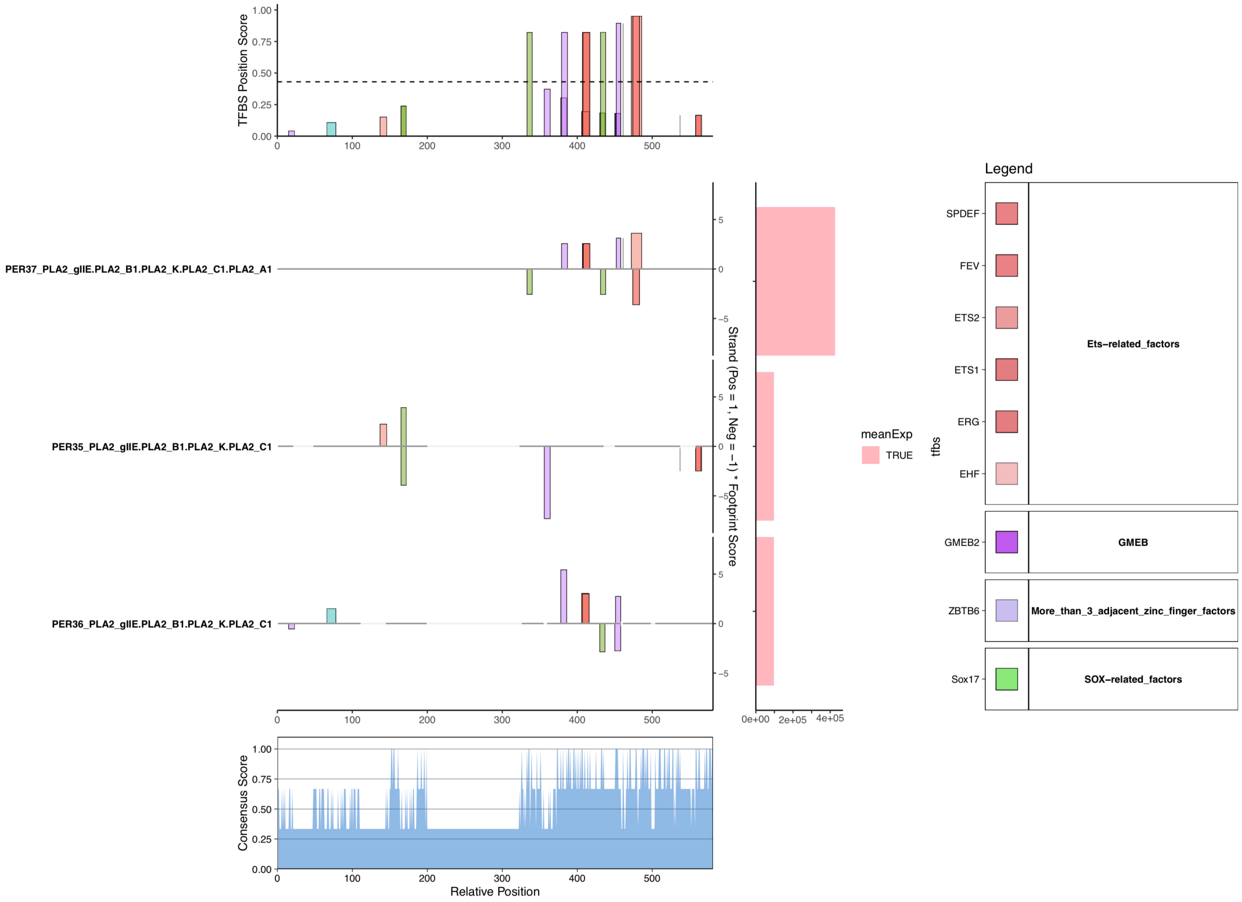
Supplementary Figure S4: Significantly enriched TFBS in putative enhancer regions (vPERs) of the three major venom gene families. A) Dots indicate enrichment in the promoter ATAC-seq peaks of a given venom gene family ($p < 0.05$), with color and size showing the significance score (-Log10(p-value) multiplied by sign of log2 enrichment). B) Venn diagram of shared enriched TFBS between SVMP, SVSP, and PLA_2 venom gene families.



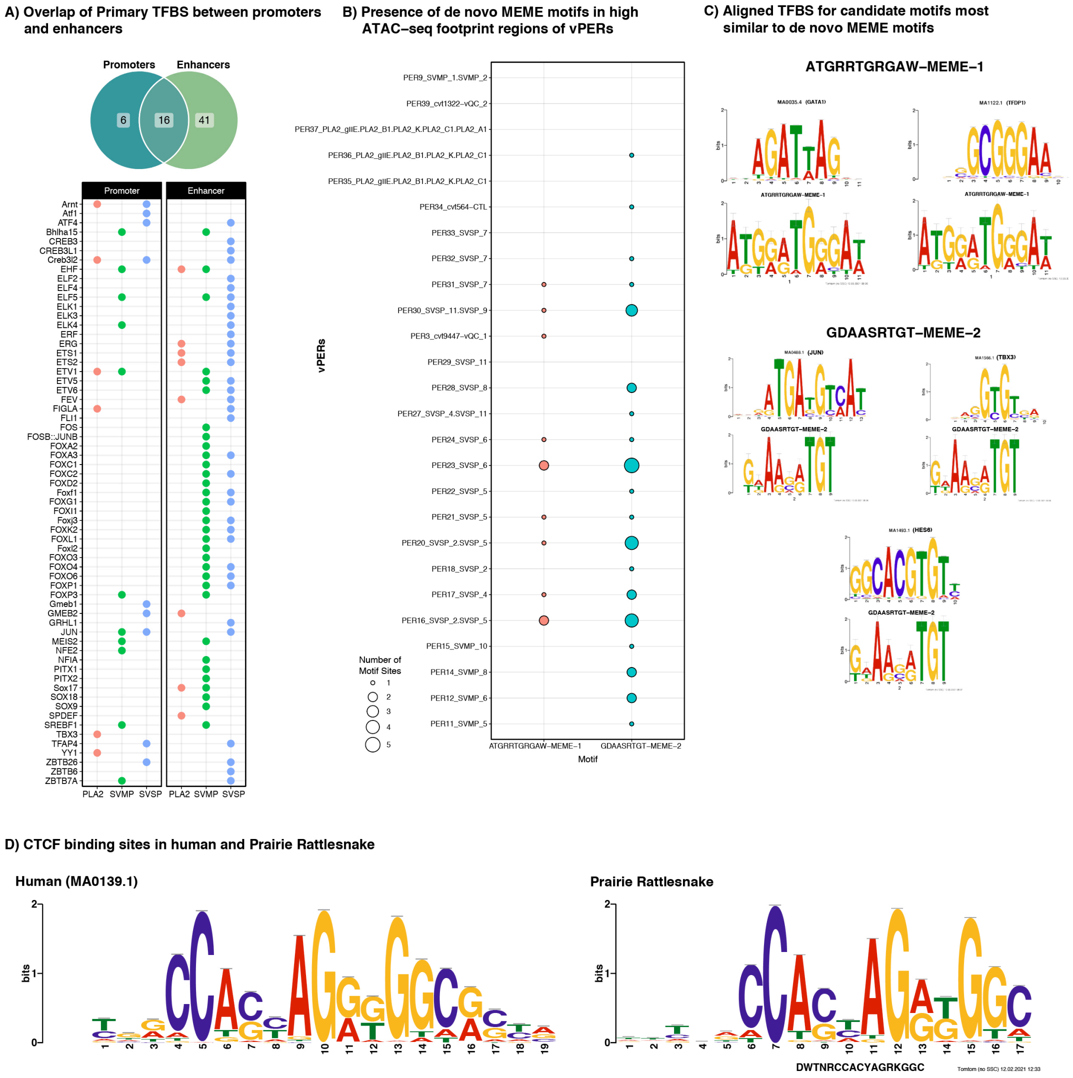
Supplementary Figure S5: Aligned TFBS in SVMP putative enhancer regions (vPERS). Alignments of SVMP vPERs, with colored vertical bars on each sequence indicating presence of a TFBS for an enriched TF (bar size is scaled by the ATAC-seq footprint score for each TFBS, and the orientation of TFBS indicated by its position above (forward) or below (reverse) the center line). Faded regions of the center lines indicate gaps introduced during alignment of the underlying sequences. Bar plots at the top of each panel are position score for each TFBS, which incorporates the footprint score and expression of genes with and without a TFBS inferred at a given site, with the dotted line indicating the mean of all position scores per gene family. A score of alignment consensus is shown beneath vPER alignments. Average gene expression at 1DPE are shown to the right, with pink bars indicating averaged expression of multiple genes associated with a given vPER. TFBS are colored based on motif groups assigned by clustering TFBS with similar binding sites.



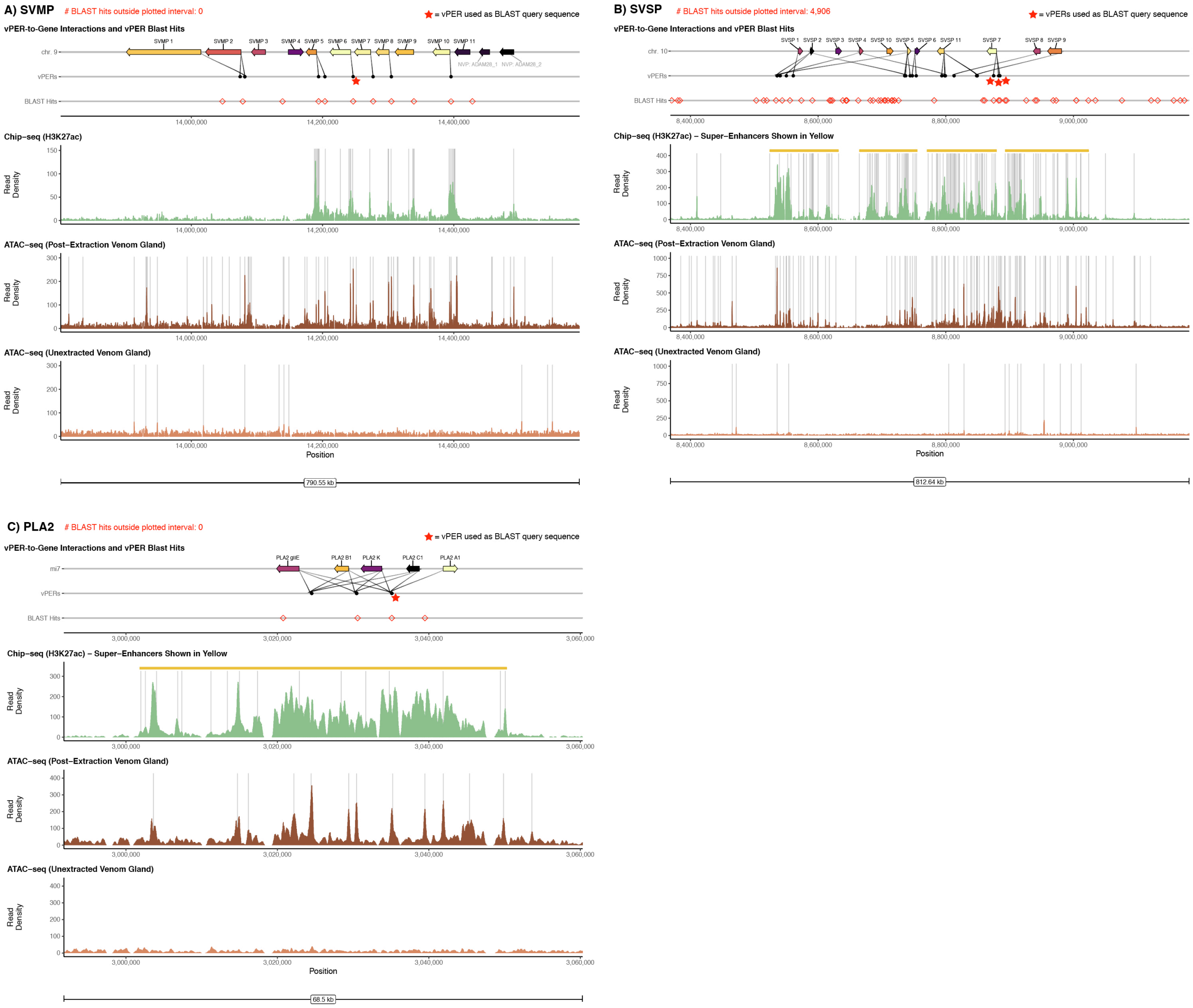
Supplementary Figure S6: Aligned TFBS in SVSP putative enhancer regions (vPERS). Alignments of SVSP vPERs, with colored vertical bars on each sequence indicating presence of a TFBS for an enriched TF (bar size is scaled by the ATAC-seq footprint score for each TFBS, and the orientation of TFBS indicated by its position above (forward) or below (reverse) the center line). Faded regions of the center lines indicate gaps introduced during alignment of the underlying sequences. Bar plots at the top of each panel are position score for each TFBS, which incorporates the footprint score and expression of genes with and without a TFBS inferred at a given site, with the dotted line indicating the mean of all position scores per gene family. A score of alignment consensus is shown beneath vPER alignments. Average gene expression at 1DPE are shown to the right, with pink bars indicating averaged expression of multiple genes associated with a given vPER. TFBS are colored based on motif groups assigned by clustering TFBS with similar binding sites.



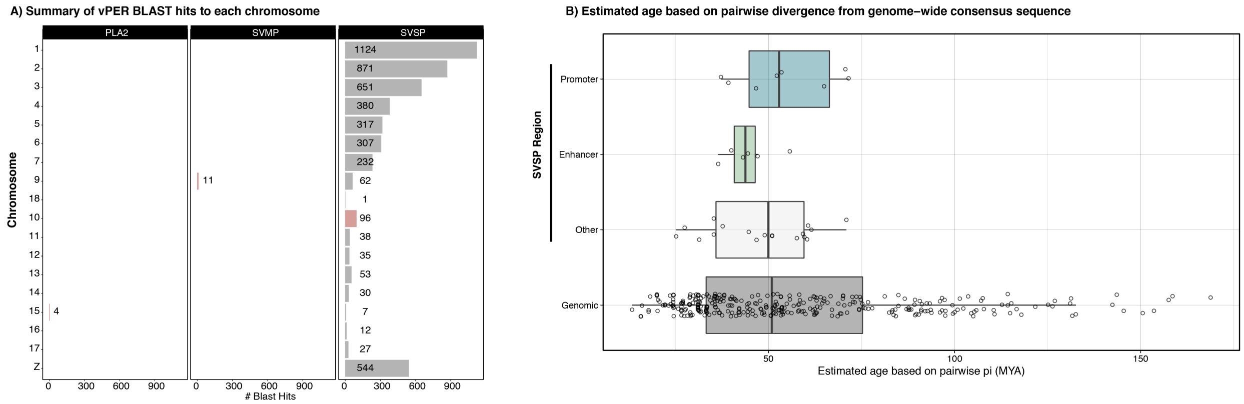
Supplementary Figure S7: Aligned TFBS in PLA$_2$ putative enhancer regions (vPERS). Alignments of PLA$_2$ vPERs, with colored vertical bars on each sequence indicating presence of a TFBS for an enriched TF (bar size is scaled by the ATAC-seq footprint score for each TFBS, and the orientation of TFBS indicated by its position above (forward) or below (reverse) the center line). Faded regions of the center lines indicate gaps introduced during alignment of the underlying sequences. Bar plots at the top of each panel are position score for each TFBS, which incorporates the footprint score and expression of genes with and without a TFBS inferred at a given site, with the dotted line indicating the mean of all position scores per gene family. A score of alignment consensus is shown beneath vPER alignments. Average gene expression at 1DPE are shown to the right, with pink bars indicating averaged expression of multiple genes associated with a given vPER. TFBS are colored based on motif groups assigned by clustering TFBS with similar binding sites.



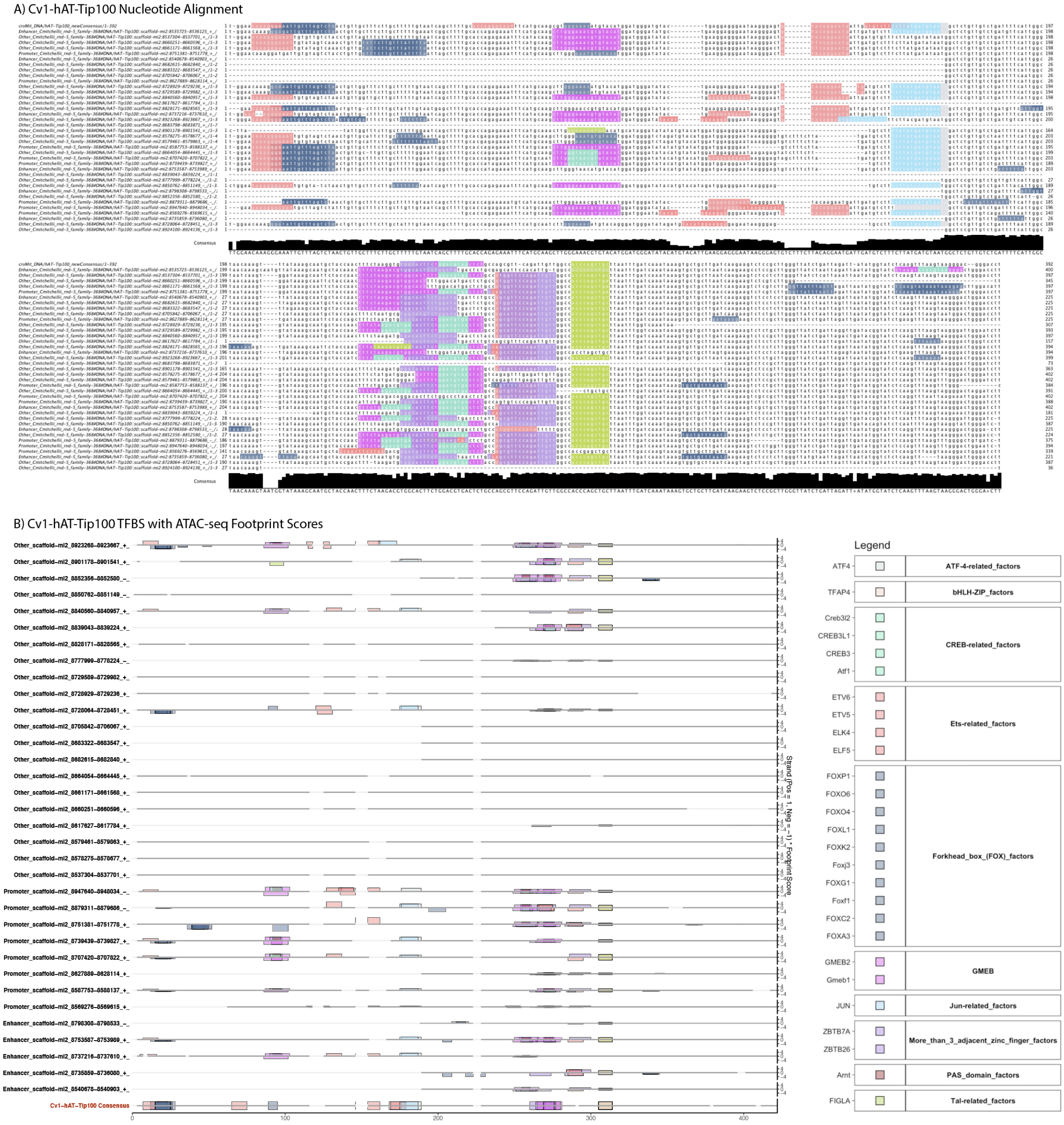
Supplementary Figure S8: A) Overlap of Primary TFBS (enriched and with a position score > mean for a given family) between promoters and enhancers. The Venn diagram at top shows the overall total number of shared and unique primary TFBS across the three main venom gene families. The dot plot below shows details of which Primary TFBS in enhancers and promoters in the three main venom gene families. B-C) Overview of de novo motif analysis in putative venom enhancer regions (vPERs). B) Presence of the two potentially novel motifs identified as enriched in vPERs compared to non-venom PERs. Point size is scaled to reflect the number of motif sites identified within elevated ATAC-seq footprint regions in a given PER. C) Motif comparisons between the two potentially novel motifs and known TFBS from our candidate TF set. D) Comparison of the CTCF binding motif in humans (MA0139.1) and the CTCF binding motif for Prairie Rattlesnake inferred using CTCF ChIP-seq.



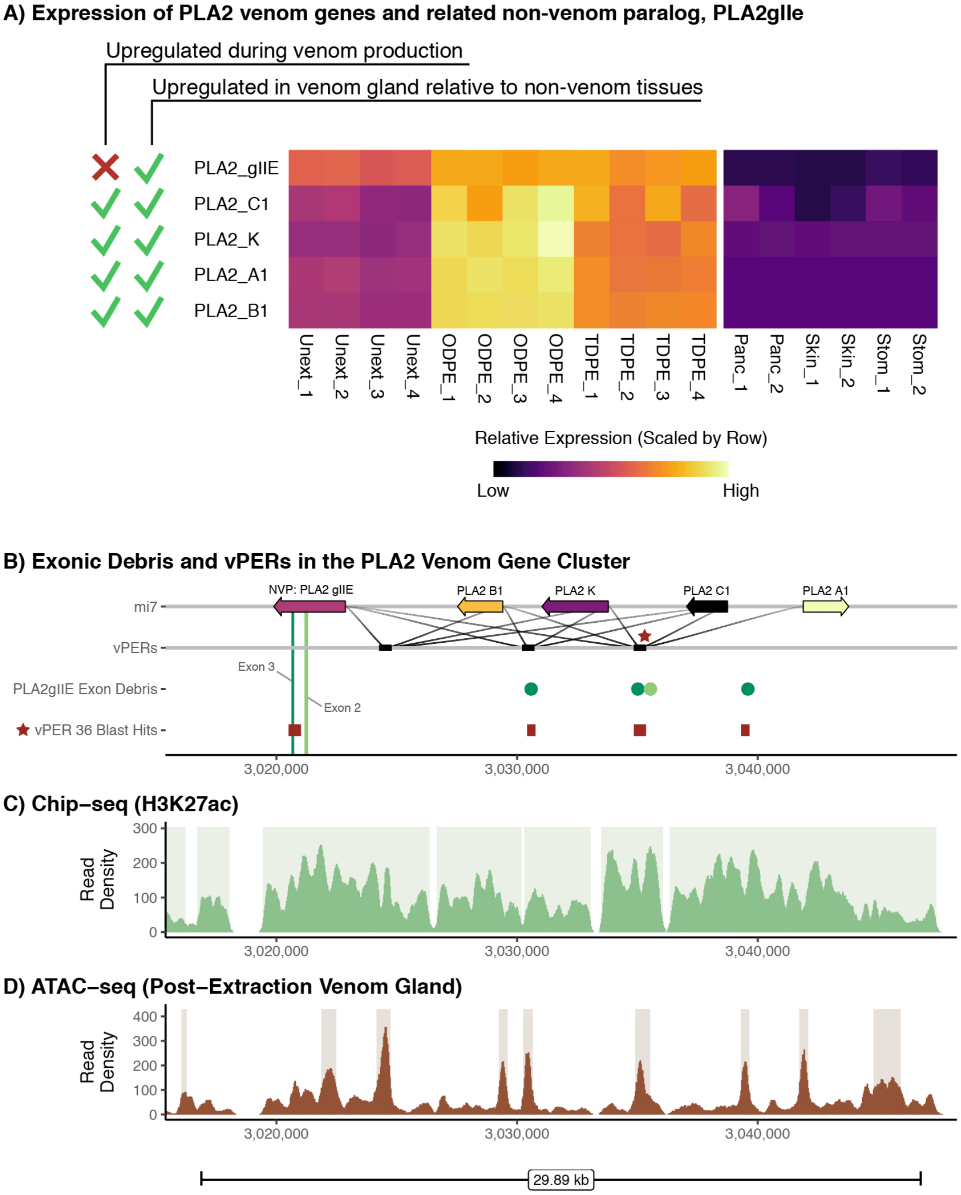
Supplementary Figure S9: Results of BLASTn searches of focal vPERs against the Prairie Rattlesnake genome. For A) SVMPs, B) SVSPs, and C) PLA$_2$s, the vPER(s) associated with the most highly expressed gene (marked with red star) was searched back against the genome using BLASTn. BLAST hits with an e-value < 0.000001 (red diamonds) are shown in the major venom array regions, and the number of hits outside of the focal region is shown next at the top of each panel. Local H3K27ac ChIP-seq (green), post-extraction (dark brown) and unextracted (orange) ATAC-seq read density is shown below, with peak regions shown with vertical grey bars.



Supplementary Figure S10: A) Summary of significant vPER BLAST hits found on each chromosome for a representative vPER from major venom gene families. Bar plot showing the number of significant BLAST hits (e < 0.000001) to the Prairie Rattlesnake genome found when using the vPER associated with the highest expressed venom gene(s) in each family as the query sequence. Bars in red indicate the chromosome on which the query vPER sequence and venom cluster reside. B) Estimated age of Cv1-hAT-Tip100 elements in the Prairie Rattlesnake genome. Estimated age of individual element copies based on pairwise divergence between each copy and the genome-wide consensus sequence, using the mutation rate of 2.24 x 10^9 following (Pasquesi et al., 2018).


A) Cv1-hAT-Tip100 Nucleotide Alignment
Consensus
B) Cv1-hAT-Tip100 TFBS with ATAC-seq Footprint Scores
Other_scaffold–ml2_8923268–8923667_+_
Other_scaffold–ml2_8901178–8901541_+_
Other_scaffold–ml2_8852356–8852580_–_
Other_scaffold–ml2_8850762–8851149_–_
Other_scaffold–ml2_8840560–8840957_+_
Other_scaffold–ml2_8839043–8839224_+_
Other_scaffold–ml2_8828171–8828565_+_
Other_scaffold–ml2_8777999–8778224_–_
Other_scaffold–ml2_8729589–8729982_+_
Other_scaffold–ml2_8728929–8729236_+_
Other_scaffold–ml2_8728064–8728451_+_
Other_scaffold–ml2_8705842–8706067_+_
Other_scaffold–ml2_8683322–8683547_+_
Other_scaffold–ml2_8682615–8682840_+_
Other_scaffold–ml2_8664054–8664445_+_
Other_scaffold–ml2_8661171–8661568_+_
Other_scaffold–ml2_8660251–8660596_+_
Other_scaffold–ml2_8617627–8617784_+_
Other_scaffold–ml2_8579461–8579863_+_
Other_scaffold–ml2_8578275–8578677_+_
Other_scaffold–ml2_8537304–8537701_+_
Promoter_scaffold–ml2_8947640–8948034_–_
Promoter_scaffold–ml2_8879311–8879686_–_
Promoter_scaffold–ml2_8751381–8751778_+_
Promoter_scaffold–ml2_8739439–8739827_+_
Promoter_scaffold–ml2_8707420–8707822_+_
Promoter_scaffold–ml2_8627889–8628114_+_
Promoter_scaffold–ml2_8587753–8588137_+_
Promoter_scaffold–ml2_8569276–8569615_+_
Enhancer_scaffold–ml2_8798308–8798533_–_
Enhancer_scaffold–ml2_8753587–8753989_+_
Enhancer_scaffold–ml2_8737216–8737610_+_
Enhancer_scaffold–ml2_8735859–8736080_+_
Enhancer_scaffold–ml2_8540678–8540903_+_
Cv1–hAT–Tip100 Consensus
0
100
200
300
400
Strand (Pos = 1, Neg = −1) * Footprint Score
Legend
ATF4 ATF-4-related_factors
TFAP4 bHLH-ZIP_factors
Creb3l2 CREB3L1 CREB3 Atf1 CREB-related_factors
ETV6 ETV5 ELK4 ELF5 Ets-related_factors
FOXP1 FOXO6 FOXO4 FOXL1 FOXK2 Foxj3 FOXG1 Foxf1 FOXC2 FOXA3 Forkhead_box_(FOX)_factors
GMEB2 Gmeb1 GMEB
JUN Jun-related_factors
ZBTB7A ZBTB26 More_than_3_adjacent_zinc_finger_factors
Arnt PAS_domain_factors
FIGLA Tal-related_factors

Supplementary Figure S11 (previous page): A) Alignment of CV1-hAT-Tip100 elements in the SVSP venom gene region. Nucleotide alignment of CV1-hAT-Tip100 elements within the SVSP region and the genome-wide consensus (at top). Sequences with "Promoter" or "Enhancer" at the beginning of the identifier overlap with inferred SVSP promoter and enhancer regions, respectively, and those with "Other" do not overlap with any annotated regulatory sequences. Colored regions indicate the presence of candidate TFBS motifs that were inferred to be important for SVSP venom gene regulation. B) TFBS Alignment of Cv1-hAT-Tip100 elements in the SVSP region weighted by ATAC-seq footprint score. Inferred TFBS positions in Cv1-hAT-Tip100 elements that fall within the SVSP region and overlap with Promoters, Enhancers, or Other sequence in the region. The genome-wide consensus sequence is shown at the bottom and does not have a footprint score.



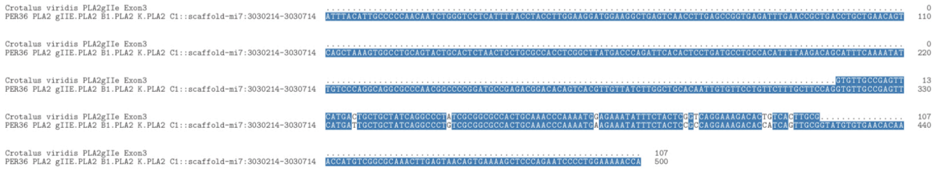
A) Expression of PLA2 venom genes and related non-venom paralog, PLA2gIIe

B) Exonic Debris and vPERs in the PLA2 Venom Gene Cluster

C) Chip–seq (H3K27ac)

D) ATAC–seq (Post–Extraction Venom Gland)

E) Nucleotide alignment of vPER36 and PLA2gIIe Exon 3



```
Crotalus viridis PLA2gIIe Exon3                                  ..............................................................................................................   0
PER36 PLA2 gIIE.PLA2 B1.PLA2 K.PLA2 C1::scaffold-mi7:3030214-3030714 ATTTACATTGCCCCCAACAATCTGGGTCCTCATTTTACCTACCTTGGAAGGATGGAAGGCTGAGTCAACCTTGAGCCGGTGAGATTTGAACCGCTGACCTGCTGAACAGT 110

Crotalus viridis PLA2gIIe Exon3                                  ..............................................................................................................   0
PER36 PLA2 gIIE.PLA2 B1.PLA2 K.PLA2 C1::scaffold-mi7:3030214-3030714 CAGCTAAAGTGGCCTGCAGTACTGCACTCTAACTGCTGCGCCACCTCGGCTTATGACCCAGATTCACACTCCTGATGCCTGCCACATTTTAAGACAGCATTTCAAAATAT 220

Crotalus viridis PLA2gIIe Exon3                                  .................................................................................................GTGTTGCCGAGTT  13
PER36 PLA2 gIIE.PLA2 B1.PLA2 K.PLA2 C1::scaffold-mi7:3030214-3030714 TGTCCCAGGCAGGCGCCCAACGGCCCCGGATGCCGAGACGGACACAGTCACGTTGTTATCTTGGCTGCACAATTGTGTTCCTGTTCTTTGCTTCCAGGTGTTGCCGAGTT 330

Crotalus viridis PLA2gIIe Exon3                                  CATGACTGCTGCTATCAGGCCCTATCGCGGCGCCACTGCAAACCCAAAATGGAGAAATATTTCTACTCGGTCAGGAAAGACACTGTCACTTGCG................ 107
PER36 PLA2 gIIE.PLA2 B1.PLA2 K.PLA2 C1::scaffold-mi7:3030214-3030714 CATGATTGCTGCTATCAGGCCCTGTCGCGGCGCCACTGCAAACCCAAAATGAAGAAATATTTCTACTCCGCCAGGAAAGACACCATCAGTTGCGGTATGTGTGAACACAA 440

Crotalus viridis PLA2gIIe Exon3                                  ............................................................ 107
PER36 PLA2 gIIE.PLA2 B1.PLA2 K.PLA2 C1::scaffold-mi7:3030214-3030714 ACCATGTCGGCGCAAACTTGAGTAACAGTGAAAAGCTCCCAGAATCCCCTGGAAAAACCA 500
```

Supplementary Figure S12 (previous page): A) Gene expression heatmap of the PLA_2 venom genes and related non-venom paralog PLA_2gIIe. Relative expression is show (scaled by row) for each gene in Unextracted (Unext), 1 day post-extraction (ODPE), and 3 day post-extraction (TDPE) venom gland compared to three non-venom tissues: pancreas (Panc), Skin, and Stomach (Stom). On the left, check marks indicate genes that are significantly upregulated ($p < 0.05$) in pairwise comparisons of venom production (between unextracted and 1 day post-extraction venom gland) and between venom tissues (all) versus non-venom tissues (all). B-E) PLA_2 enhancers may be derived from incomplete duplication of their non-venom paralog, PLA_2 gIIE. B) The PLA_2 venom gene cluster and their non-venom paralog PLA_2gIIE with interactions shown to putative venom enhancer regions (vPERs). The second and third exon of PLA_2gIIE are marked with green bars, and the green dots in the PLA_2gIIE Exon Debris row indicate exonic debris corresponding to these exon regions resulting from partial duplication of this gene. Below, significant Blast results to vPER 36 (marked with a star; the only vPER inferred to regulate both highly expressed PLA_2s) correspond to the third exon of PLA_2gIIE and multiple vPER regions. C,D) Active enhancer (H3K27ac) ChIP-seq and ATAC- seq from the post-extraction venom gland, with peaks shown as shaded rectangles. E) Nucleotide alignment of vPER36 and the third exon of PLA_2gIIE show substantial sequence identity (positions highlighted in blue).

Supplementary Tables

Supplementary Table S1. Candidate transcription factors (TFs) with potential involvement in venom gene regulation. TFs identified as upregulated during venom production (RNA-seq), associated with super-enhancer regions (ChIP-seq), and/or positively differentially bound during venom production (ATAC-seq). Functional characterization of candidate TFs highlight those previously implicated in venom gene regulation, those involved in adrenoceptor signaling, and those with known interactions with ERK. Normalized gene expression counts across tissues, differential expression, and differential ATAC-seq footprinting analysis results are included for each candidate TF.

Transcription Factor ID	Prairie Rattlesnake Transcript ID	Upregulated (RNA-seq)	SE-Associated (ChIP-seq)	Differentially Bound (ATAC-seq)	Previously implicated in venom regulation	Interacts with ERK	Adrenoceptor Signaling Pathway	AP-1 Complex Member
AEBP1	crovir-transcript-12782	√						
AEBP2	crovir-transcript-11366		√					
ARID4A	crovir-transcript-6996	√						
ARNT	crovir-transcript-12636		√					
ATF1	crovir-transcript-14479			√				
ATF2	crovir-transcript-8315	√				√	√	
ATF3	crovir-transcript-9649			√				
ATF4	crovir-transcript-12088		√			√	√	
ATF6B	crovir-transcript-13131	√				√	√	
ATF7	crovir-transcript-14662			√				
ATF7IP	crovir-transcript-12040	√						
ATRX	crovir-transcript-10	√						
BHLHA15	crovir-transcript-320		√					
BOLA1	crovir-transcript-1197	√						
BOLA3	crovir-transcript-13007	√						
BUD31	crovir-transcript-144		√					
CAMTA1	crovir-transcript-1130	√						
CARHSP1	crovir-transcript-509	√						
CDC73	crovir-transcript-5698		√					
CEBPZ	crovir-transcript-9446	√						
CIC	crovir-transcript-724		√					
CREB3	crovir-transcript-12236	√	√			√	√	
CREB3L1	crovir-transcript-7444	√				√	√	
CREB3L2	crovir-transcript-12236	√	√		√	√	√	
CREB3L4	crovir-transcript-12594			√		√	√	
CREB5	crovir-transcript-2064			√		√		
CREBRF	crovir-transcript-13972		√					

CREM	crovir-transcript-1813	√	√	√			√	
CSRNP2	crovir-transcript-14714		√					
DACH1	crovir-transcript-3243	√						
DDIT3	crovir-transcript-14617		√		√			√
EHF	crovir-transcript-7549		√	√	√			
ELF2	crovir-transcript-4489			√				
ELF4	NA			√	√			
ELF5	crovir-transcript-7566			√	√			
ELK1	crovir-transcript-14740			√		√		
ELK3	crovir-transcript-11970			√				
ELK4	crovir-transcript-6293			√		√		
ERF	crovir-transcript-709			√		√		
ERG	crovir-transcript-3668			√	√			
ETS1	crovir-transcript-12725			√	√	√		
ETS2	crovir-transcript-3667	√		√		√		
ETV1	crovir-transcript-1984			√	√			
ETV2	crovir-transcript-811			√	√			
ETV5	crovir-transcript-10662			√				
ETV6	crovir-transcript-12831			√				
FEV	crovir-transcript-7975			√				
FIGLA	crovir-transcript-12994	√	√					
FLI1	crovir-transcript-12724			√				
FOS	crovir-transcript-7192		√		√	√		√
FOSB	crovir-transcript-1248		√	√				
FOSL1	NA			√				
JUND	NA			√	√	√		√
FOSL2	crovir-transcript-9524			√				
FOXA2	crovir-transcript-5494			√	√			
FOXA3	crovir-transcript-1318			√	√			
FOXC1	NA			√				
FOXC2	crovir-transcript-11432			√	√			
FOXD2	crovir-transcript-6046			√				
FOXE1	crovir-transcript-15186			√				
FOXF1	crovir-transcript-11433			√	√			
FOXG1	crovir-transcript-7241			√				
FOXI1	crovir-transcript-14004			√				
FOXJ3	NA			√				
FOXK2	crovir-transcript-14093	√		√	√			
FOXL1	NA			√				
FOXL2	crovir-transcript-10736			√	√			
FOXO3	crovir-transcript-16138		√		√	√		
FOXO4	crovir-transcript-16138		√	√		√		
FOXO6	NA			√		√		
FOXP1	crovir-transcript-13360			√	√			
FOXP3	crovir-transcript-14882			√				
GATA1	crovir-transcript-14789	√				√		
GLIS2	crovir-transcript-1669	√	√					
GMEB1	crovir-transcript-919			√				
GMEB2	NA			√				
GPBP1	crovir-transcript-15800	√						
GRHL1	crovir-transcript-8758		√	√	√			
GRHL2	crovir-transcript-4926		√	√	√			
HCFC1	crovir-transcript-14873		√					
HDAC1	crovir-transcript-5542	√	√					
HDAC3	crovir-transcript-5542	√	√					
HES6	crovir-transcript-10728	√	√					
HIF3A	crovir-transcript-16294	√	√					

HIVEP1	crovir-transcript-5324	√						
HYAL2	crovir-transcript-13660	√						
IKZF5	crovir-transcript-10426	√	√					
IRX1	crovir-transcript-5471		√					
IRX2	crovir-transcript-5470		√					
JDP2	crovir-transcript-7235			√				
JUN	crovir-transcript-6797		√	√	√	√		√
JUNB	NA			√	√			√
KAT6A	crovir-transcript-12928		√					
KDM5B	crovir-transcript-6418		√					
KLF10	crovir-transcript-4917	√						
KLF11	crovir-transcript-6772		√			√		
KLF13	crovir-transcript-11070		√					
KLF16	crovir-transcript-6772		√					
KLF8	crovir-transcript-16156	√						
LITAF	crovir-transcript-1631		√					
MAZ	crovir-transcript-2964	√						
MBD1	crovir-transcript-14772	√	√					
MBNL3	crovir-transcript-16112		√					
MED30	crovir-transcript-4885	√	√					
MEIS1	crovir-transcript-8589		√					
MEIS2	crovir-transcript-7256	√						
MLLT3	crovir-transcript-15356	√						
MXD1	crovir-transcript-12992		√					
NCOA3	crovir-transcript-6549	√						
NFATC1	crovir-transcript-5268	√						
NFE2	crovir-transcript-14649	√						
NFIA	crovir-transcript-5965		√		√			
NFIB	crovir-transcript-15370		√		√			
NFXL1	crovir-transcript-4728		√					
NOC3L	crovir-transcript-10347	√						
NOCT	crovir-transcript-4490	√						
NOLC1	crovir-transcript-10213	√						
NR1H2	crovir-transcript-5260	√						
NR4A1	crovir-transcript-14451	√	√			√		
NR4A2	crovir-transcript-8197	√	√		√			
NR4A3	crovir-transcript-2319	√						
OVOL1	crovir-transcript-12399		√					
PHTF1	crovir-transcript-6359	√						
PITX1	crovir-transcript-13885	√						
PITX2	crovir-transcript-4420		√					
PLAG1	crovir-transcript-5094		√					
PPARD	crovir-transcript-6233	√						
PPP1R13L	crovir-transcript-1260		√					
PRDM2	crovir-transcript-11363		√					
PSMD9	crovir-transcript-16386	√						
PTTG1	crovir-transcript-14040	√				√		
PURA	crovir-transcript-12918		√					
PURB	crovir-transcript-12918		√					
RARA	crovir-transcript-2559		√			√		
RBM14	crovir-transcript-12391	√						
RERE	crovir-transcript-1029		√					
RFX2	crovir-transcript-10114	√						
RFX7	crovir-transcript-11031	√						
RFXAP	crovir-transcript-3901	√						
RREB1	crovir-transcript-5349		√					
SALL2	crovir-transcript-3122	√						

SALL3	crovir-transcript-5264	√				
SCML2	crovir-transcript-3515	√				
SFPQ	crovir-transcript-669	√				√
SKI	crovir-transcript-2243	√				
SMAD9	crovir-transcript-3899	√				
SMARCA5	crovir-transcript-4474	√				
SOX17	crovir-transcript-5107	√				
SOX18	crovir-transcript-11609	√				
SOX9	crovir-transcript-14372		√			
SP1	crovir-transcript-15064		√			√
SPDEF	crovir-transcript-6220		√		√	
SREBF1	crovir-transcript-1605		√			
SSRP1	crovir-transcript-7902	√				
SUFU	crovir-transcript-10545		√			
SUPT20H	crovir-transcript-15280	√				
SUPT4H1	crovir-transcript-9971	√				
TAF1	crovir-transcript-16150	√				
TBX3	crovir-transcript-9066		√			√
TCFL5	crovir-transcript-6673			√		
TERF2	crovir-transcript-16196		√			
TFAP2A	crovir-transcript-5334		√			
TFAP4	crovir-transcript-1652		√			
TFCP2L1	crovir-transcript-8096		√			
TFDP1	crovir-transcript-3361			√		
TLE1	crovir-transcript-11392		√			
TOX4	crovir-transcript-2320	√				
TSC22D2	crovir-transcript-10819	√				
TSC22D3	crovir-transcript-16160		√			
TSC22D4	crovir-transcript-11156		√			
UHRF1	crovir-transcript-6698		√			
WWP2	crovir-transcript-16206	√				
XBP1	crovir-transcript-16464		√			
YEATS4	crovir-transcript-11866	√				
YY1	crovir-transcript-7062	√			√	
ZBTB26	crovir-transcript-1507		√			
ZBTB33	crovir-transcript-209			√		
ZBTB49	crovir-transcript-269	√				
ZBTB6	crovir-transcript-1506		√			
ZBTB7A	NA			√		
ZC3H15	crovir-transcript-8369	√				
ZC3H3	crovir-transcript-7268	√				
ZFP36L1	crovir-transcript-7440		√			
ZGPAT	crovir-transcript-1101	√				
ZNF217	crovir-transcript-6590		√			
ZNF341	crovir-transcript-6462		√			
ZNF451	crovir-transcript-8917	√				
ZNF462	crovir-transcript-15288		√			
ZNF511	crovir-transcript-6877		√			
ZNF512	crovir-transcript-9923	√				
ZNF574	crovir-transcript-710	√				
ZNF592	crovir-transcript-11170		√			
ZNF622	crovir-transcript-5422	√				
ZNF710	crovir-transcript-11397		√			
ZNF76	crovir-transcript-6234	√				

Supplementary Table S2. KEGG Pathway overrepresentation analysis results for candidate transcription factors. Results of KEGG pathway overrepresentation analysis using candidate TFs as the target set and all other annotated TFs in the Prairie Rattlesnake as the background. Only pathways with significant overrepresentation (FDR < 0.05) are shown.

Gene Set	Description	Size of Gene Set	Overlap Size	Expected Overlap Size	Enrichment Ratio	p-value	FDR	Candidate TFs in Gene Set
hsa05031	Amphetamine addiction	14	11	2.4211	4.5435	4.794E-07	5.800E-05	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS;FOSB;HDAC1;JUN
hsa04915	Estrogen signaling pathway	20	12	3.4586	3.4696	1.192E-05	4.672E-04	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS;JUN;NCOA3;RARA;SP1
hsa04925	Aldosterone synthesis and secretion	12	9	2.0752	4.3370	1.236E-05	4.672E-04	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;NR4A1;NR4A2
hsa04261	Adrenergic signaling in cardiomyocytes	10	8	1.7293	4.6261	1.880E-05	4.672E-04	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;CREM
hsa05034	Alcoholism	15	10	2.5940	3.8551	1.930E-05	4.672E-04	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOSB;HDAC1;HDAC3
hsa05030	Cocaine addiction	13	9	2.2481	4.0033	3.459E-05	6.976E-04	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOSB;JUN
hsa04926	Relaxin signaling pathway	14	9	2.4211	3.7174	8.342E-05	1.262E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS;JUN
hsa04927	Cortisol synthesis and secretion	14	9	2.4211	3.7174	8.342E-05	1.262E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;NR4A1;SP1
hsa04911	Insulin secretion	9	7	1.5564	4.4976	9.518E-05	1.280E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4
hsa04728	Dopaminergic synapse	12	8	2.0752	3.8551	1.534E-04	1.856E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS
hsa04918	Thyroid hormone synthesis	10	7	1.7293	4.0478	2.734E-04	3.007E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4
hsa04725	Cholinergic synapse	8	6	1.3835	4.3370	4.629E-04	4.668E-03	ATF4;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS
hsa04151	PI3K-Akt signaling pathway	17	9	2.9398	3.0614	6.467E-04	6.020E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOXO3;NR4A1
hsa04668	TNF signaling pathway	18	9	3.1128	2.8913	1.114E-03	9.625E-03	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS;JUN
hsa04928	Parathyroid hormone synthesis, secretion and action	22	10	3.8045	2.6285	1.460E-03	1.178E-02	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOS;NR4A2;SP1
hsa04010	MAPK signaling pathway	20	9	3.4586	2.6022	2.852E-03	2.157E-02	ATF2;ATF4;DDIT3;ELK1;ELK4;FOS;JUN;NFATC1;NR4A1
hsa04211	Longevity regulating pathway	17	8	2.9398	2.7212	3.571E-03	2.542E-02	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;FOXO3
hsa05163	Human cytomegalovirus infection	25	10	4.3233	2.3130	4.709E-03	3.165E-02	ATF2;ATF4;ATF6B;CREB3;CREB3L1;CREB3L2;CREB3L4;ELK1;NFATC1;SP1
hsa04024	cAMP signaling pathway	18	8	3.1128	2.5700	5.537E-03	3.526E-02	CREB3;CREB3L1;CREB3L2;CREB3L4;FOS;JUN;NFATC1;SOX9

Supplementary Table S3. *De novo* motif search results for venom promoter sequences from MEME. No motifs identified by MEME are significantly enriched in venom promoter sequences relative to non-venom gene promoter sequences.

MEME ID	Sequence	Width (bp)	# Sites	log likelihood ratio	p-value	E-value
MEME-1	AWTCMTKT	8	7	61	5.20E-01	6.70E-01
MEME-2	ATYCATGYMMDCRTW	15	12	149	1.00E-01	1.00E-01
MEME-3	TTGYTTCTHWWWDTY	15	10	102	5.10E-01	6.50E-01
MEME-4	AATCCT	6	14	81	1.50E-01	1.60E-01
MEME-5	GWTGTA	6	7	53	4.90E-01	6.20E-01
MEME-6	AGGAAATAA	9	4	47	7.70E-01	1.20E+00
MEME-7	CTTCAGCWTH	10	4	39	8.90E-01	1.70E+00
MEME-8	ATGGCTCT	8	2	22	5.20E-01	6.60E-01
MEME-9	AGGTGTTT	8	2	23	6.60E-01	9.50E-01
MEME-10	TTGATS	6	2	16	4.40E-01	5.30E-01
MEME-11	WWRSNARNDRH	11	24	62	9.10E-01	1.80E+00
MEME-12	GGAATGA	7	2	20	1.00E+00	4.00E+00
MEME-13	ACAATA	6	2	18	1.00E+00	3.60E+00
MEME-14	CAAACA	6	2	17	1.00E+00	3.90E+00
MEME-15	ACTTCWG	7	5	37	1.00E+00	4.00E+00
MEME-16	RSBTAAYK	8	3	24	9.60E-01	2.20E+00
MEME-17	ANDKSWK	7	24	31	1.00E+00	3.40E+00
MEME-18	GGRAYA	6	3	21	1.00E+00	3.80E+00
MEME-19	TKTGCA	6	2	16	1.00E+00	4.00E+00
MEME-20	YYDGDG	6	24	28	1.00E+00	2.00E+00

Chapter 6

MICROCHROMOSOMES EXHIBIT DISTINCT FEATURES OF VERTEBRATE CHROMOSOME STRUCTURE AND FUNCTION WITH UNDERAPPRECIATED RAMIFICATIONS FOR GENOME EVOLUTION

Blair W. Perry[1], Drew R. Schield[1,2], Richard H. Adams[1,3], Todd A. Castoe[1]

[1]Department of Biology, 501 S. Nedderman Drive, University of Texas at Arlington, Arlington, TX, 76019, USA.

[2]Current Address: Department of Ecology and Evolutionary Biology, 1900 Pleasant Street, 334 UCB, University of Colorado, Boulder, CO, 80309, USA

[3]Current Address: Department of Computer and Electrical Engineering and Computer Science, 777 Glades Road, EE 418, Florida Atlantic University, Boca Raton, FL, 33431, USA

Abstract

Microchromosomes are common yet poorly understood components of many vertebrate genomes. Recent studies have revealed that microchromosomes contain a high density of genes and possess other distinct characteristics compared to macrochromosomes. Whether distinctive characteristics of microchromosomes extend to features of genome structure and organization, however, remains an open question. Here we analyze Hi-C sequencing data from multiple vertebrate lineages and show that microchromosomes exhibit consistently high degrees of interchromosomal interaction (particularly with other microchromosomes), appear to be co-localized to a common central nuclear territory, and are comprised of a higher proportion of open chromatin than macrochromosomes. These findings highlight an unappreciated level of diversity in vertebrate genome structure and function, and raise important questions regarding the evolutionary origins and ramifications of microchromosomes and the genes that they house.

Introduction

The three-dimensional (3D) organization and interactions of the genome play fundamental roles in gene regulation and genome function (Cremer and Cremer, 2001; Cremer et al., 1993). Advances in functional genomics approaches such as Hi-C sequencing (Lieberman-Aiden et al., 2009) have broadened our understanding of 3D genomic interactions and organization in the nucleus, including how chromatin loops coordinate the regulation of genes and how chromosomes form discrete chromosome territories within the nucleus (Bolzer et al., 2005; Cremer et al., 1993; Habermann et al., 2001). Most studies of 3D genome organization and structure have focused on mammalian genomes that are exclusively comprised of macrochromosomes (Cremer and Cremer, 2001; Cremer et al., 1993; Kurz et al., 1996). However, many non-mammalian vertebrates possess microchromosomes - nuclear chromosomes generally smaller than 30 Mb in length - in addition to macrochromosomes (Axelsson et al., 2005; Burt, 2002; International Chicken Genome Sequencing Consortium, 2004; Ohno et al., 1969; Schield et al., 2019; Zhou and Gui, 2002). Microchromosome number is variable across vertebrates, ranging from 0 in macrochromosome-only lineages to greater than 40 in other lineages (Deakin and Ezaz, 2019; O'Connor et al., 2019). Vertebrate microchromosomes consistently exhibit many distinct features across lineages, including high gene density, low transposable element content, and high rates of recombination (Backström et al., 2010; International Chicken Genome Sequencing Consortium, 2004; Schield et al., 2019, 2020), and represent a functionally and evolutionarily unique fraction of the genomes of many vertebrates. However, it remains largely unknown how 3D genomic features manifest in nuclei of vertebrates containing both macro- and microchromosomes.

Recent Hi-C studies of vertebrates with microchromosomes have provided increasing evidence for distinct features of microchromosome organization and function. A study of the Prairie Rattlesnake (*Crotalus viridis*) found that microchromosomes exhibit higher degrees of interaction with other chromosomes than expected based on chromosome size (Schield et al., 2019). A similar trend was observed in chicken erythrocytes (*Gallus gallus*) (Fishman et al., 2018). This study also inferred AB compartments across the chicken genome, which broadly correspond to regions of open (A compartment) and closed (B compartment) chromatin (Lieberman-Aiden et al., 2009), and showed that microchromosomes exhibit a higher proportion of A compartment regions than macrochromosomes (Fishman et al., 2018). Together, these studies suggest that microchromosomes may be functionally and organizationally distinct compared to macrochromosomes. The extent to which these patterns represent universal characteristics of microchromosomes remains unexplored, and their evolutionary causes and ramifications largely unconsidered.

Here, we use recently published chromosome-level genome assemblies and Hi-C datasets for representatives of multiple vertebrate lineages to infer patterns of 3D interaction and organization of genomes that possess both macro- and microchromosomes. Based on these data, we demonstrate that high interchromosomal interaction and enrichment for A compartment regions are likely ubiquitous features of vertebrate microchromosomes, and find support for previous suggestions that microchromosomes co-inhabit the center of the nucleus. Collectively, these findings suggest that vertebrate genomes with microchromosomes may structurally, functionally, and evolutionarily operate in fundamentally distinct ways compared to macrochromosome-only genomes. This conclusion highlights the largely unexplored evolutionary relevance of the presence/absence of microchromosomes across vertebrate lineages, and the relevance of genes being encoded on microchromosomes.

Results

Our analyses of Hi-C data indicate that, for all species analyzed (Supp. Tables 1-2), interchromosomal contact frequency generally increases as chromosome size decreases (Fig. 1a_{i}-$_{ii}$). Microchromosomes therefore exhibit a higher degree of interchromosomal interaction, with all non-mammalian species exhibiting a significantly higher degree of interchromosomal interaction in microchromosomes than in macrochromosomes (Fig. 1d$_{ii}$-$_{iii}$). Interestingly, in the chicken, which possesses the smallest microchromosomes among all species we analyzed, there is an apparent inflection point in chromosome size at which interchromosomal activity begins to decrease as chromosome size continues to decrease (Fig. 1di, Supp. Fig. 1). This pattern is apparent in all three chicken tissues analyzed, and less pronounced inflection points near the smallest microchromosomes in the Prairie Chicken (Fig. 1e$_{i}$) and Sea Turtle (Fig. 1f$_{i}$).

To further investigate patterns of interchromosomal contacts between macrochromosomes and microchromosomes, we compared empirical interchromosomal contact frequencies (ICFs) to ICFs predicted by a null model assuming uniform interactions between chromosomes, following (Zhang et al., 2012). In all non-mammalian species, we find an excess of ICFs between microchromosome pairs and fewer than expected ICFs between macrochromosomes and microchromosomes (Fig. 1d$_{iv}$-$_{iiv}$). Hierarchical clustering of chromosomes based on observed over expected ICFs distinguishes macrochromosomes from microchromosomes in nearly all species and tissues, with a small number of exceptions in the rattlesnake (Fig. 1$_{iv}$) and the three chicken tissues analyzed (Supp. Fig. 1).

For all species possessing microchromosomes, we inferred AB compartments based on patterns of interchromosomal contact frequencies at 50 kb resolution between all chromosomes and binned measures of GC content. We find that microchromosomes in all species are comprised of a significantly higher proportion of A compartment regions compared to macrochromosomes, which are predominately comprised of B compartment regions (Fig. 2, Supp. Fig. 2).

Genome-wide heatmaps of binned Hi-C contact frequency and 3D interpretations of interaction data both show evidence of well-defined chromosome territories for macrochromosomes (Fig. 3, Supp. Figs. 3-8). For microchromosomes, contact frequency heatmaps show elevated levels of intrachromosomal interaction (Supp. Fig. 3), and show an elevated degree of microchromosome-microchromosome interaction.

Furthermore, this high degree of microchromosome interaction results in a lack of obvious spatial distinction between microchromosomes in 3D interpretations of Hi-C interaction data, and independent microchromosome territories are not well defined (Fig. 3, Supp. Fig. 3-8). While 3D interpretations of Hi-C data should not be directly interpreted as biologically accurate models of the nucleus, they do provide fairly robust inferences regarding the degree of isolation of chromosomes based on patterns of 2D interaction. Note that 3D models were not generated for the three chicken tissues due to the data for several microchromosomes being too sparse to generate intrachromosomal contact maps at necessary resolution.

Discussion

Using Hi-C contact data from diverse vertebrate lineages, we demonstrate that microchromosomes consistently exhibit an elevated degree of interchromosomal interactivity compared to that of macrochromosomes. This pattern of elevated inter-chromosomal interaction for microchromosomes is consistent with previous studies of single species (chicken (Fishman et al., 2018) and rattlesnake (Schield et al., 2019)), and our expanded sampling indicate that these patterns are likely remarkably consistent across diverse vertebrate lineages. We consistently find that the high magnitude of microchromosome interactivity is dominated by microchromosome-to-microchromosome interactions, and additionally show that microchromosomes are consistently enriched for, and in many cases comprised almost exclusively of, A compartment regions. These findings emphasize the unique structural and functional features of vertebrate microchromosomes, and raise interesting questions about the relationships between microchromosome structure and genome function and organization.

Previous microscopy studies have suggested that bird microchromosomes inhabit the center of the nucleus with macrochromosomes arranged around them at the nuclear periphery (Berchtold et al., 2011; Habermann et al., 2001; Skinner et al., 2009). Similar studies have not yet, however, been conducted for other species with microchromosomes (i.e. fish, non-avian reptiles), and the degree to which this chromosomal arrangement is conserved across vertebrates with microchromosomes remains unknown. Our findings of consistently elevated microchromosome-microchromosome interactions is consistent with a model in which microchromosomes are localized in the center of the nucleus across diverse vertebrate lineages. This arrangement of microchromosomes is also supported by our inference that microchromosomes are primarily comprised of A compartment (open chromatin) regions, which tend to be concentrated at the center of the nucleus (Kosak et al., 2007; Misteli, 2007). Taken together, our Hi-C based inferences and previous studies tentatively support a model of nuclear organization in which A-rich microchromosomes occupy the center of the nucleus, surrounded by A-rich regions of macrochromosomes that inhabit the nuclear periphery (Fig. 3g). Interestingly, somewhat analogous examples exist in insect chromosomes (e.g., *Drosophila* dot chromosome), in which these chromosomes with distinct compositional characteristics (heterochromatic, gene dense, transposon-rich) occupy distinct regions of the nucleus (Riddle and Elgin, 2018), implying broad links between nuclear chromosome organization and chromosome composition, structure and function. Future studies that utilize 3D fluorescence *in situ* hybridization for multiple vertebrates with microchromosomes would be particularly valuable for testing our hypotheses for nuclear organization, and the degree to which it is conserved across species and cell types.

Available evidence suggests that microchromosomes collectively exhibit features that are distinct from typical macrochromosomes, in that they are closely associated in the nucleus and interact more frequently with other microchromosomes than to macrochromosomes. This argues for the presence of a microchromosome-specific territory in the nucleus that features a higher degree of interchromosomal interaction than typically observed for macrochromosomes (Fig. 3f). However, the degree to which microchromosomes inhabit well-defined individual territories within this encompassing microchromosome territory remains an open question; it is possible that the lack of defined microchromosome territories in our 3D interpretations of Hi-C data may result from variable positioning of microchromosomes across sampled cells (i.e., a merged ‘average’ of relative position). It also remains an open question how such an arrangement of microchromosomes may influence the formation and position of the nucleolus in the nucleus. Regardless, the high degree of interaction among microchromosomes raises the possibility of interchromosomal regulatory interactions between microchromosomes, a phenomenon thought to be rare in macrochromosomes (Bashkirova and Lomvardas, 2019; Maass et al., 2019) that should be explore further in microchromosomes.

While our findings show notably similar characteristics between microchromosomes of multiple vertebrate lineages, it is worth noting that our current sampling is remarkably sparse in the context of vertebrate diversity, and lacks representatives from several important lineages that also possess microchromosomes (i.e. fish) for which Hi-C contact information data is not currently available. While we do observe consistent patterns across many of the tissue and cell types sampled here (whole blood, venom gland, erythrocytes) that may represent common features of microchromosome biology and organization, we expect variation and exceptions to these patterns to exist in various cell types, tissues, and developmental stages within species. Indeed, we observed evidence of variation in interchromosomal contact patterns when various chicken cell types are compared, with some of these variations being particularly distinct in chicken embryonic fibroblast cells (Supp. Fig. 1). The degree to which patterns of microchromosome interaction and structure observed here are broadly present and/or consistent across the full diversity of vertebrate lineages, tissue, and cell types therefore remains an open question for future studies, as additional data for diverse vertebrates becomes available.

A major consideration emphasized by our findings is how unique features of microchromosomes may affect the evolution of genes housed on microchromosomes. Unlike macrochromosomes, microchromosomes tend to share a common nuclear territory, and have high levels of interchromosomal interaction, and consist of mainly A compartment active chromatin. Intriguingly, despite this unusually high level of interchromosomal interaction, which may suggest functional interactions among microchromosomes, they segregate independently and consistently exhibit among the highest genome-wide recombination rates

(Backström et al., 2010; International Chicken Genome Sequencing Consortium, 2004; Schield et al., 2019). This has profound implications for the evolution of genes on microchromosomes, and suggests that the rate and efficiency of selection, and the effects of drift, would be distinct on microchromosomes compared to macrochromosomes. For example, high recombination rates in microchromosomes would be very effective at breaking down linkage disequilibrium, breaking associations among selected alleles, and thereby increasing the efficacy of selection. These features suggest that microchromosomes possess ideal characteristics for housing genes underlying adaptation. Anecdotal support for this comes from the Prairie Rattlesnake genome, in which microchromosomes contain the majority of important venom genes, which are generally known to be under strong local selection (Casewell et al., 2013; Mackessy, 2010; Schield et al., 2019), although more extensive systematic studies of additional vertebrate lineages would be necessary to test hypotheses for the special relevance of microchromosomes in adaptation. Continued accumulation of chromosome-level genome resources for diverse vertebrates will provide new opportunities to test hypotheses related to the roles of microchromosomes in genome evolution, investigate the relevance of genes and gene families being located on microchromosomes, and elucidate the factors that drive shifts from macrochromosome-only systems to those containing both chromosome types.

Methods

Hi-C data were downloaded from the NCBI Sequence Read Archive for the Prairie Rattlesnake (*Crotalus viridis*), Burmese Python (*Python bivittatus*), Argentine Black and White Tegu (*Salvator merianae*), Green Sea Turtle (*Chelonia mydas*), Greater Prairie Chicken (*Tympanuchus cupido*), chicken (*Gallus gallus*), Rhesus Macaque (*Macaca mulatta*), Patski Mouse (*Mus musculus* x *Mus spretus*), and human (*Homo sapiens*). See Supplementary Table 1 for details. Hi-C reads for each species were mapped to genome assemblies and processed using the Juicer pipeline (Durand et al., 2016a). For each species, inter- and intrachromosomal contact matrices were extracted from the resulting Hi-C map using the dump command in Juicer Tools v1.9.9 at 50kb, 100kb, and 1mb resolutions using KR-normalization and only reads that mapped with MAPQ > 30. The size at which a chromosome is designated a microchromosome is not well defined, and most previous studies have defined microchromosomes and macrochromosomes largely based on visual dichotomies apparent in chromosome squashes (ex. Habermann et al., 2001). In this study, avian microchromosomes were defined as chromosomes shorter than 30Mb. Visual inspection of linear chromosome length for non-avian reptiles revealed a more apparent natural break between larger and smaller chromosomes around 50Mb, and we therefore defined chromosomes shorter than this as microchromosomes. Downstream analyses of observed versus expected interchromosomal contact frequencies (described below) lend support to this breakpoint, as chromosomes defined herein as macrochromosomes and microchromosomes based on these criteria cluster strongly with others of the same type, with few exceptions (see Fig. 1).

The sum of all interchromosomal contacts per chromosome was divided by chromosome length to produce a relative measure of interchromosomal contact density per chromosome, and the relationship between this normalized contact frequency and chromosome length was tested using linear regression in R (R Core Team, 2014). Differences between macrochromosome and microchromosome interchromosomal contact frequencies were tested using student's t-tests. Observed contact frequencies were compared to the expected interchromosomal contact frequency for each chromosome pair assuming uniform interactions between chromosomes following (Zhang et al., 2012). The log2 ratio of observed over expected interchromosomal contact frequency was plotted as a heatmap in R using pheatmap v1.0.12 (https://github.com/raivokolde/pheatmap). Heatmaps of Hi-C contact frequency were generated with Juicebox (Durand et al., 2016b).

miniMDS (Rieber and Mahony, 2017) was used to generate 3D interpretations of Hi-C data using 1Mb resolution interchromosomal contact data and 50kb resolution intrachromosomal contact data. miniMDS was run using full partitioning with minimum partition size 0.08 and the default smoothing parameter. The resulting 3D models were visualized using Mayavi (Ramachandran and Varoquaux, 2011). Note that Hi-C data for the three chicken tissues was too sparse to generate 50kb intrachromosomal contact maps for input into miniMDS, and therefore these samples were excluded from 3D modeling.

Juicer Hi-C matrices were converted to the cooler format (Abdennur and Mirny 2020) at 50 kb resolution using hic2cool v0.8.3 (https://github.com/4dn-dcic/hic2cool) and normalized using 'balance' within the cooler CLI package v0.8.7 (Abdennur and Mirny, 2020). GC content was measured in 50 kb bins using the 'nuc' program within bedtools v2.29.0 (Quinlan and Hall, 2010). AB compartments were determined with 'call-compartments' within cooltools v0.3.2 (https://github.com/mirnylab/cooltools) using trans (interchromosomal) contacts and binned measures of GC content as the reference track. The proportion of A compartment regions per chromosome was calculated as the number of 50 kb bins determined to belong to the A compartment divided by the total number of bins representing the chromosome and plotted in R. A student's t-test was used to test for enrichment of A compartments on microchromosomes.

Acknowledgments

We thank the Texas Advanced Computing Center for access to computational resources, and support to T.A.C from NSF (DEB-1655571, IOS-1655735).

Figures

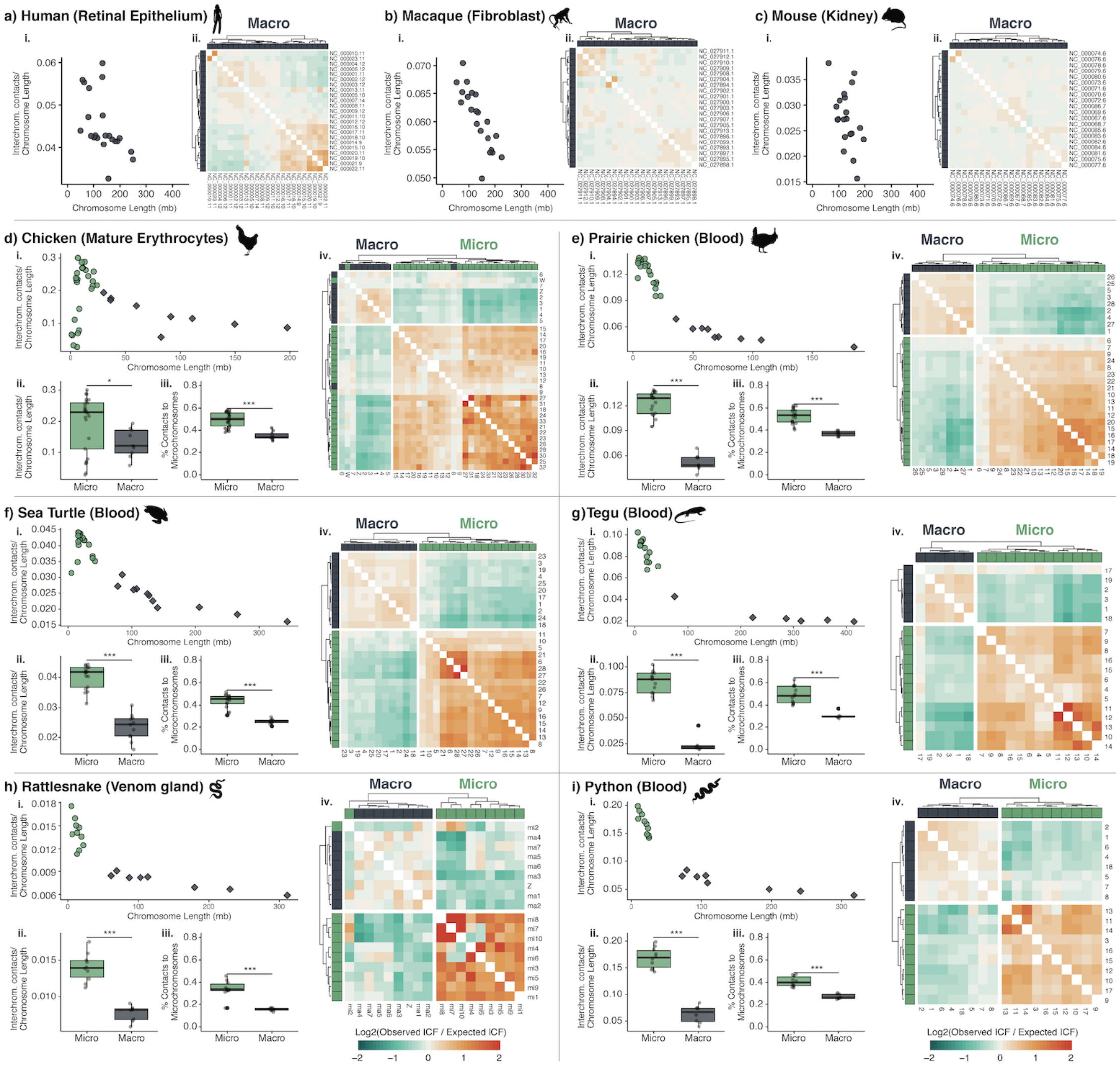


Figure 1. Microchromosomes exhibit elevated interchromosomal contact frequencies and interact preferentially with other microchromosomes. a_i-i_i) Sums of interchromosomal contact frequencies per chromosome normalized by chromosome length plotted over chromosome length. d_{ii}-i_{ii}) Comparisons of interchromosomal contact frequency normalized by chromosome length for macro and microchromosomes (*: p-value < 0.05, ***: p-value < 0.001, Student's t-test). d_{iii}-i_{iii}) Comparison of the proportion of interchromosomal contacts that involve a microchromosome for macrochromosomes and microchromosomes (*** denotes $p < 0.001$, Student's t-test). a_{ii}-c_{ii}, d_{iv}-i_{iv}) Heatmaps of the ratio of observed to expected interchromosomal contact frequency (ICF) between all chromosome pairs, with hierarchical clustering and chromosome type annotated above and to the left of each heatmap.

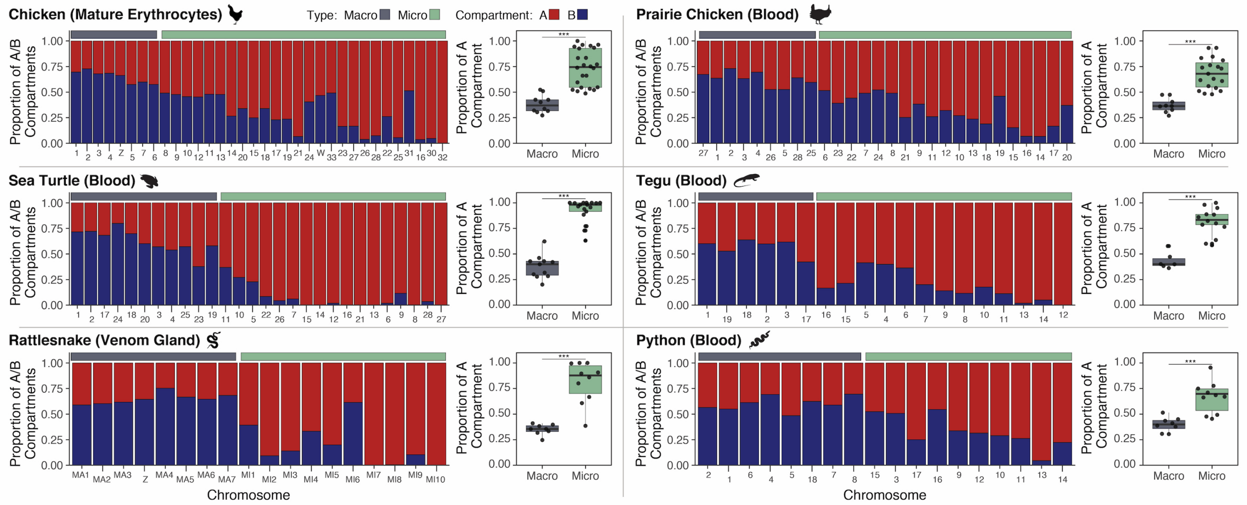


Figure 2. Microchromosomes are enriched for the A compartment. Bar plots indicate the proportion of 50 kb bins for each chromosome that were determined to be A (red) and B (blue) compartment. In all species, microchromosomes exhibit a higher proportion of A compartment bins than macrochromosomes (boxplots on right; *** denotes p < 0.001, Student's t-test).

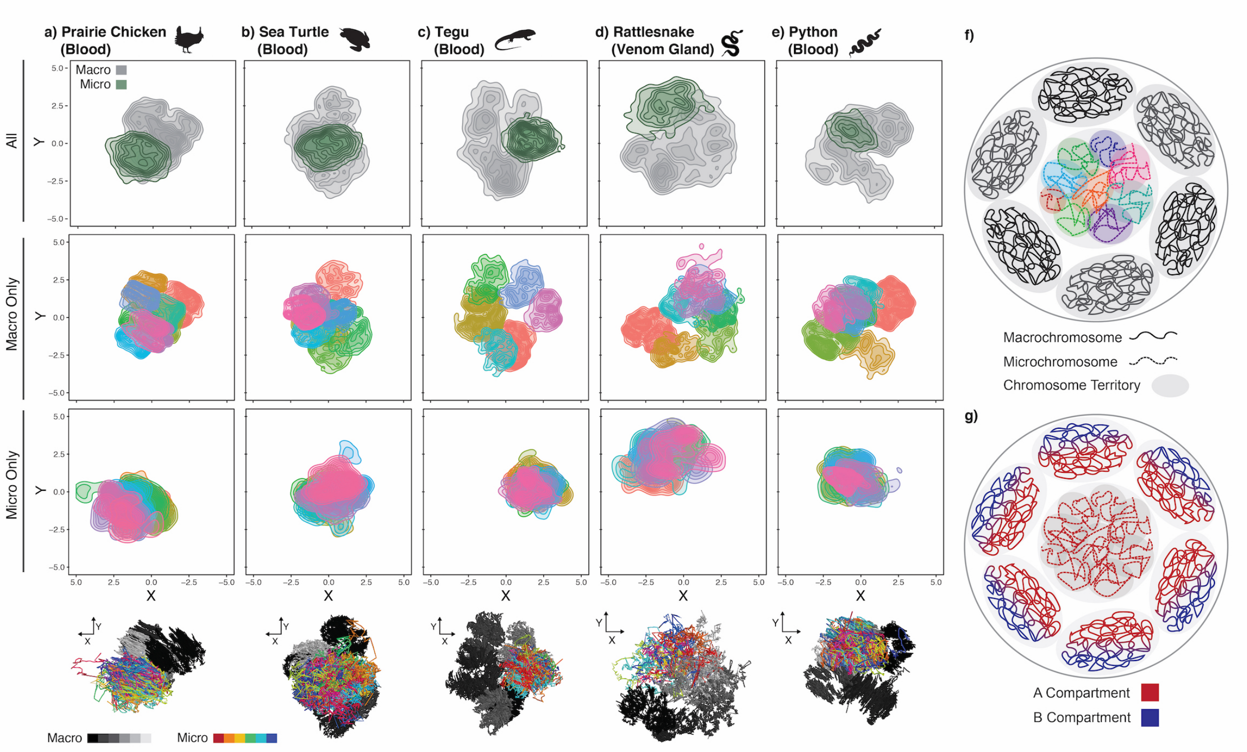
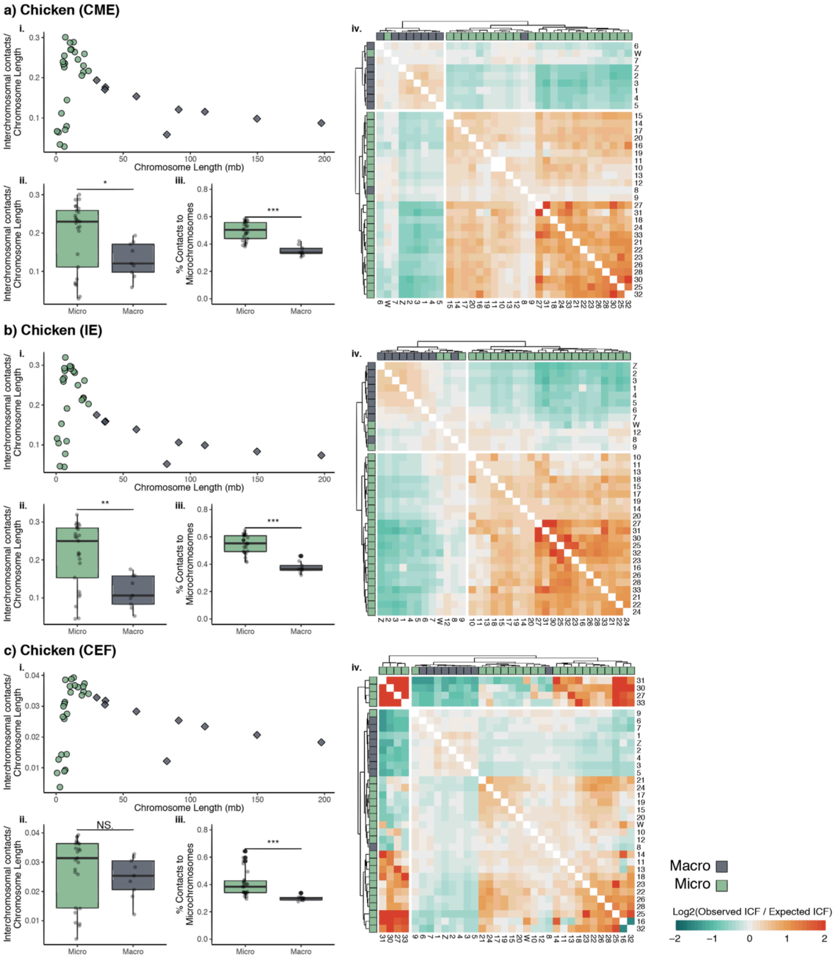
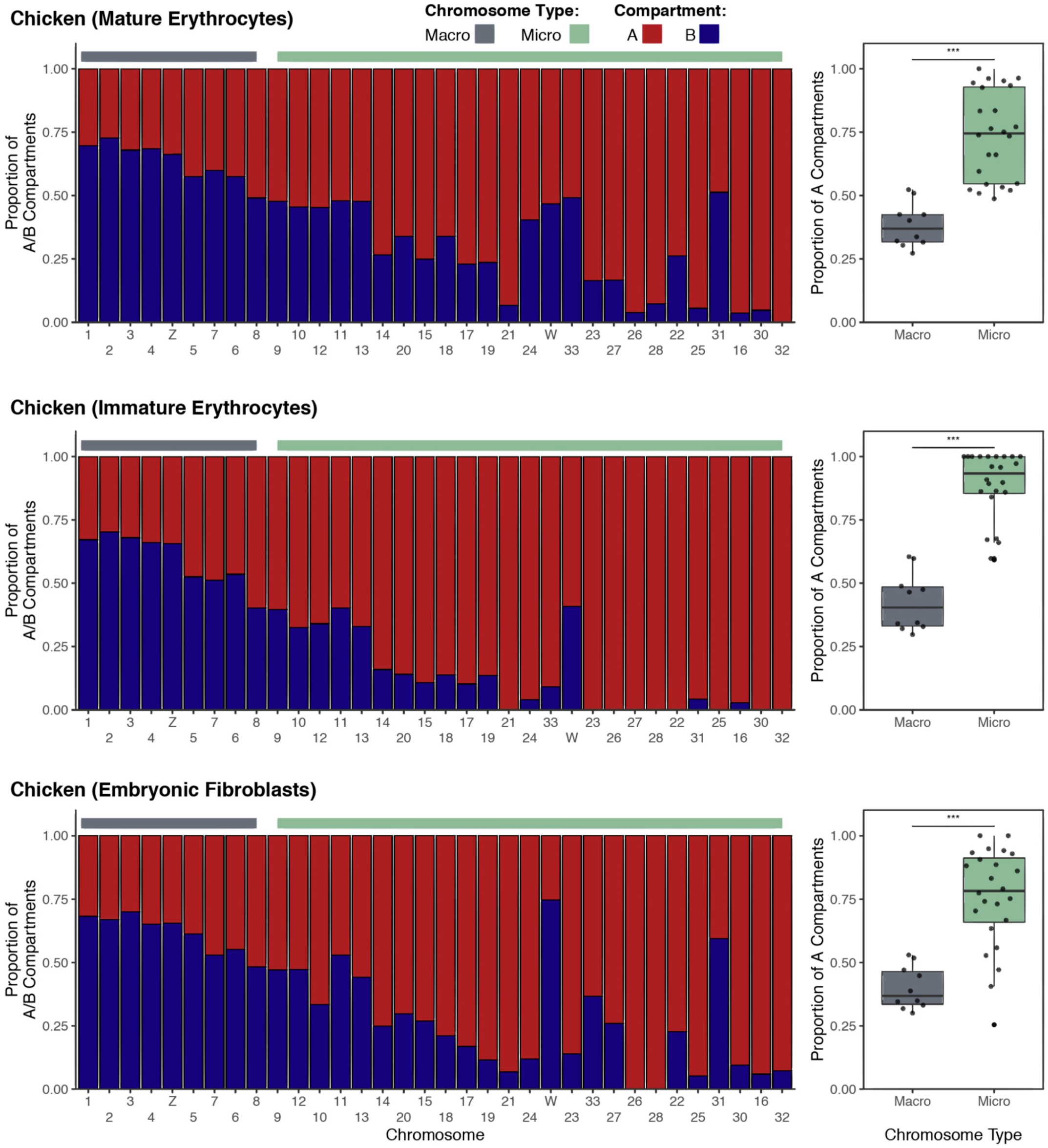


Figure 3. Microchromosomes are likely co-localized in the 3D nucleus. a-e) 3D interpretations of Hi-C interaction data shown as 2D point density plots from three distinct orientations for all chromosomes, macrochromosomes only, and microchromosomes only. For macro and micro- only plots, different colors represent different chromosomes. Shown at the bottom are 3D interpretations of all chromosomes, with macrochromosomes in greyscale and microchromosomes in color. Additional orientations for each species are available in Supp. Figs. 4-8. f-g) cartoon representations of a nucleus illustrating the hypotheses that f) microchromosomes are centrally located in the nucleus and collectively inhabit a "microchromosome territory" and g) that of spatial organization of A and B compartments in a nucleus containing A-rich microchromosomes.

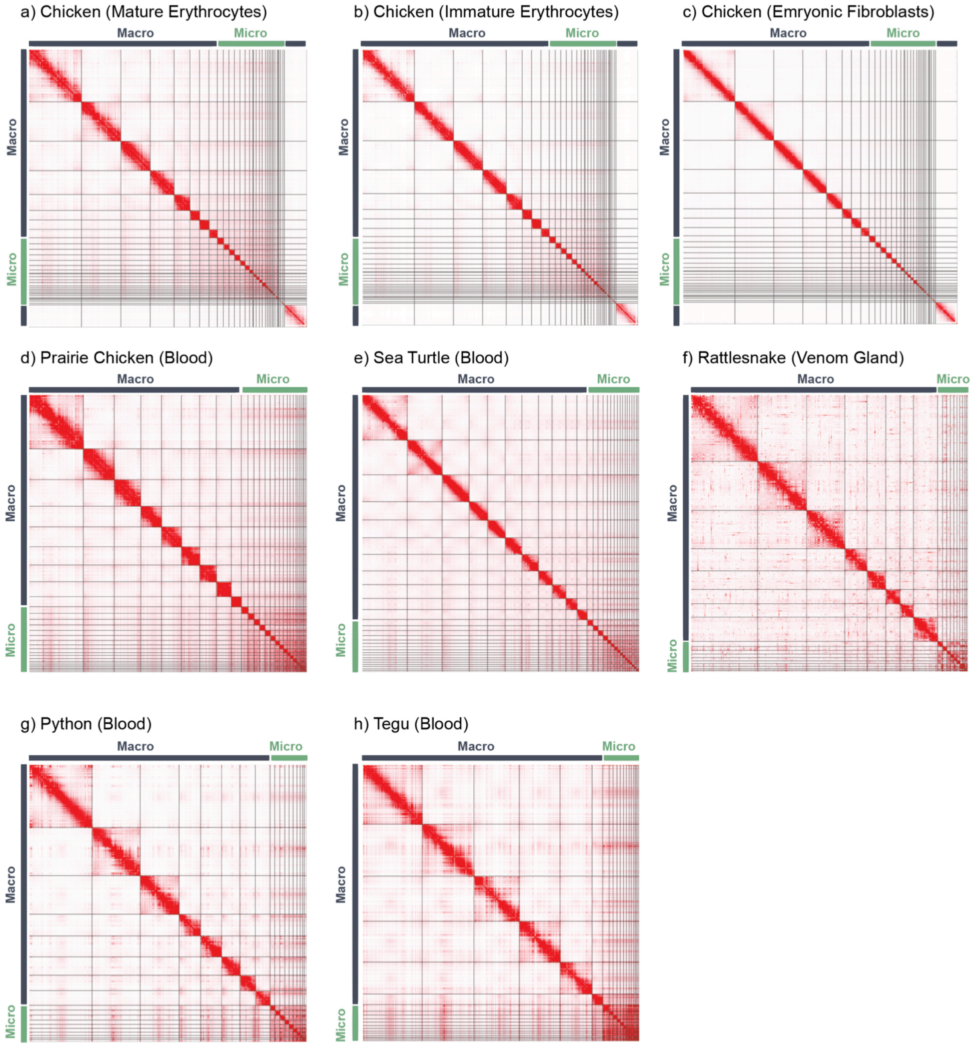
Supplementary Figures



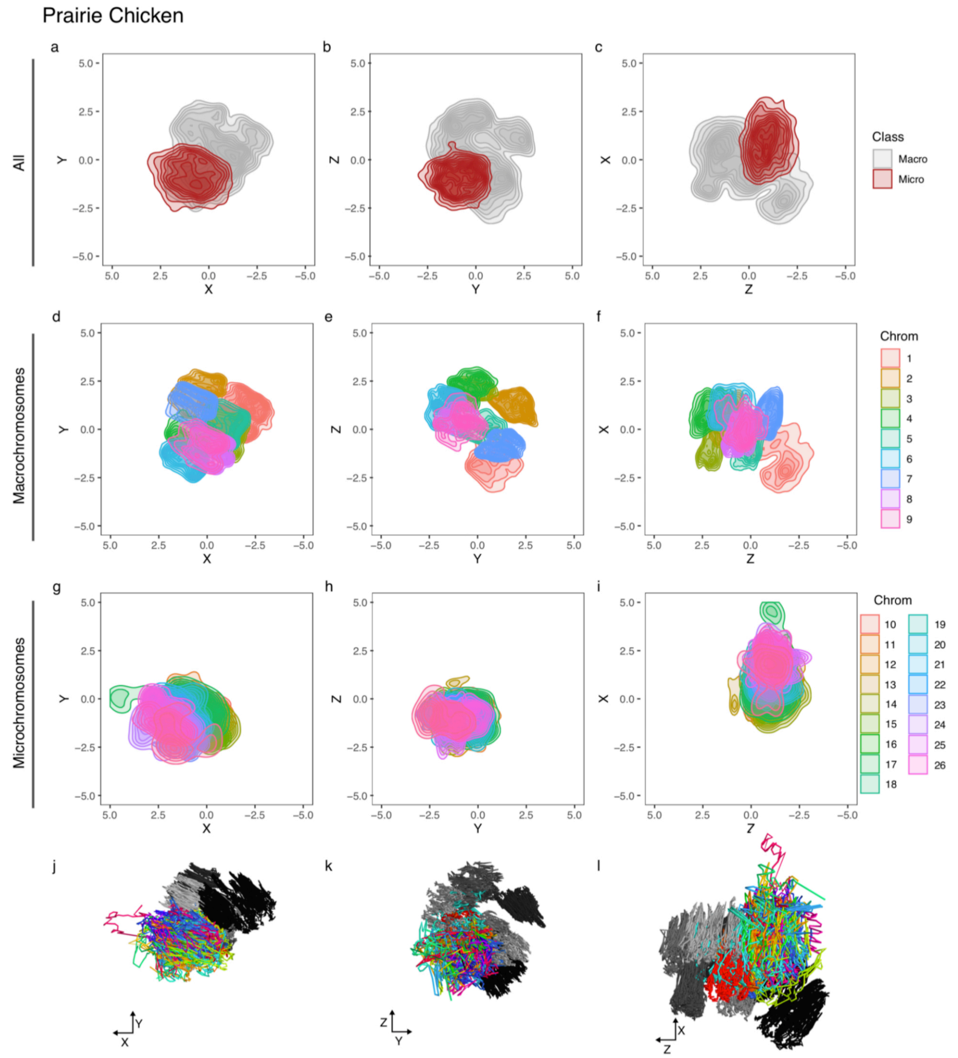
Supplementary Figure 1. Microchromosomes interactions in three chicken tissues. Patterns of interchromosomal interaction for chicken mature erythrocyte (CME), immature erythrocyte (IE), and embryonic fibroblast (CEF) cells. a.i-c.i) Sums of interchromosomal contact frequencies per chromosome normalized by chromosome length plotted over chromosome length. a.ii-c.ii) Comparisons of interchromosomal contact frequency normalized by chromosome length for macro and microchromosomes (*: p-value < 0.05, ***: p-value < 0.001, Student's t-test). a.iii-c.iii) Comparison of the proportion of interchromosomal contacts that involve a microchromosome for macrochromosomes and microchromosomes (*** denotes p < 0.001, Student's t-test). a.iv-c.iv) Heatmaps of the ratio of observed to expected interchromosomal contact frequency (ICF) between all chromosome pairs, with hierarchical clustering and chromosome type annotated above and to the left of each heatmap.



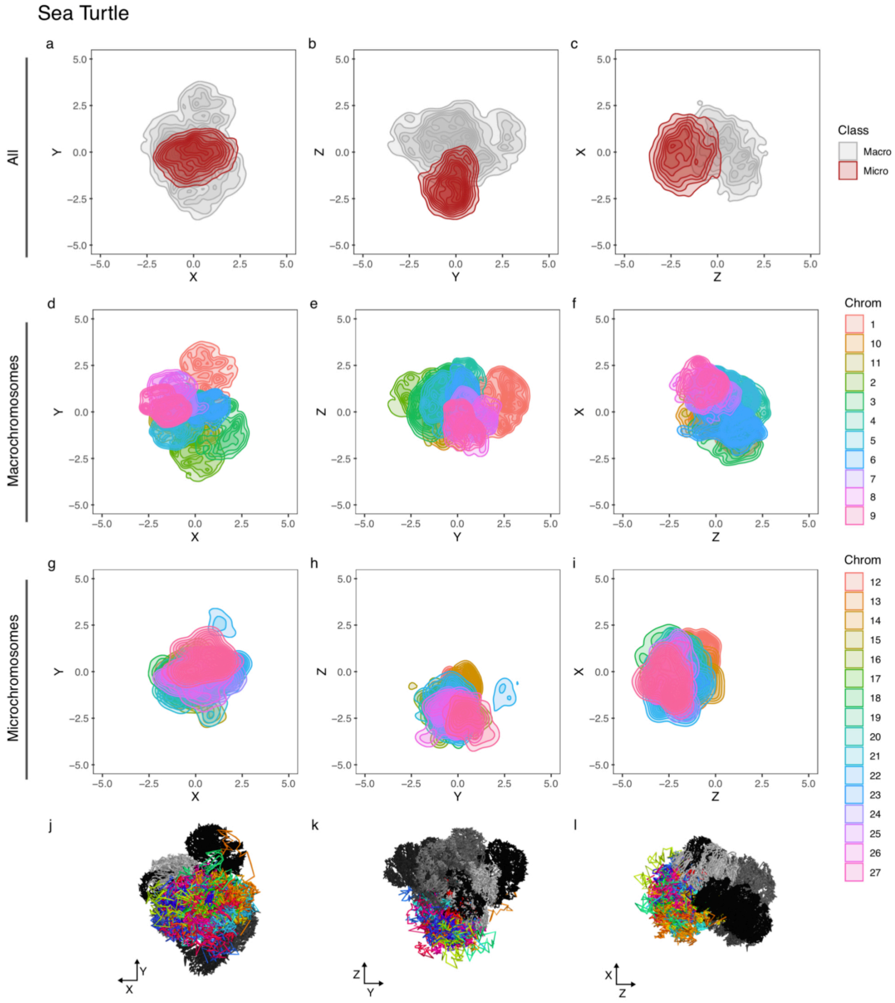
Supplementary Figure 2. Microchromosomes are enriched for the A compartment in all three chicken tissues. Bar plots indicate the proportion of 50 kb bins for each chromosome that were determined to be A (red) and B (blue) compartment. In all tissues, microchromosomes exhibit a higher proportion of A compartment bins than macrochromosomes (boxplots on right; *** denotes $p < 0.001$, Student's t-test).



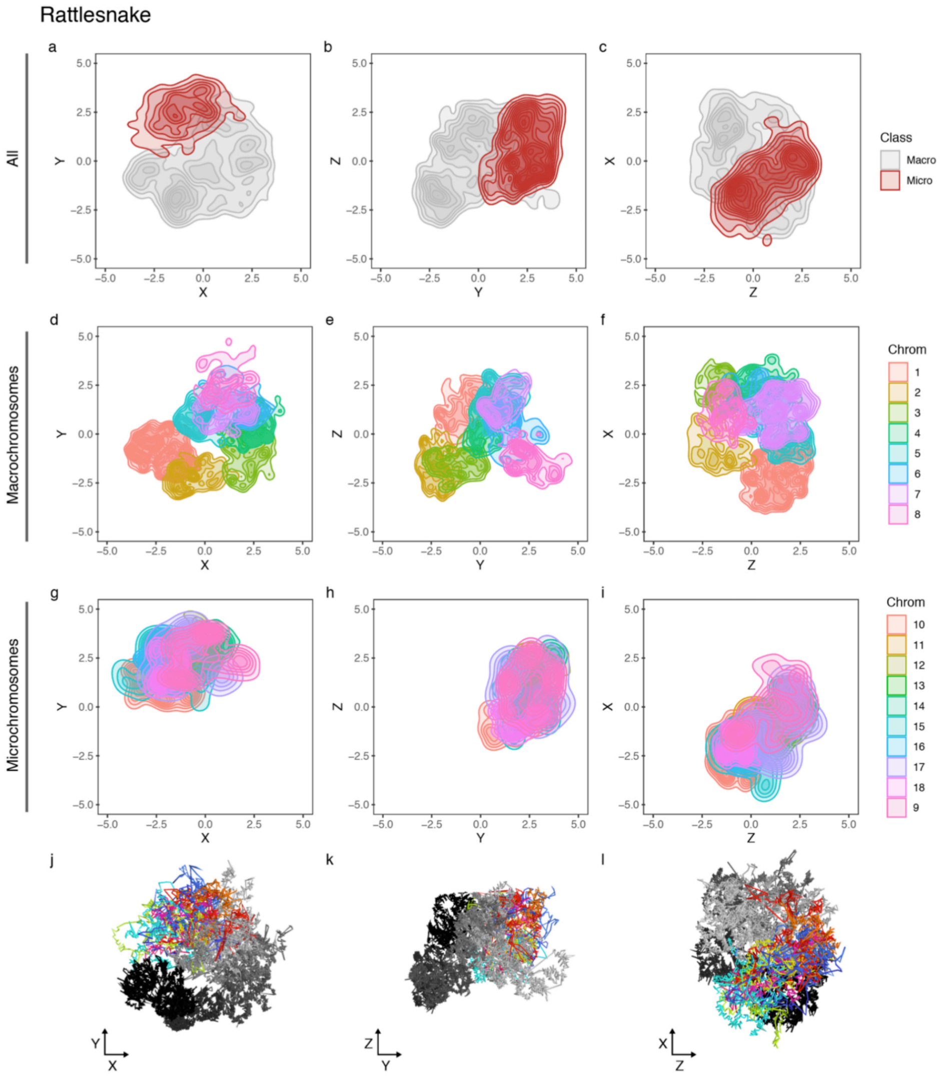
Supplementary Figure 3. Hi-C contact frequency heatmaps at 50kb resolution for all focal species possessing both macrochromosomes and microchromosomes. Darker red indicates higher contact frequency. Chromosome territories are evidenced by defined "blocks" of interaction frequency corresponding to chromosomes that indicate a high degree of self- interaction and lesser degree of interaction with other chromosomes.



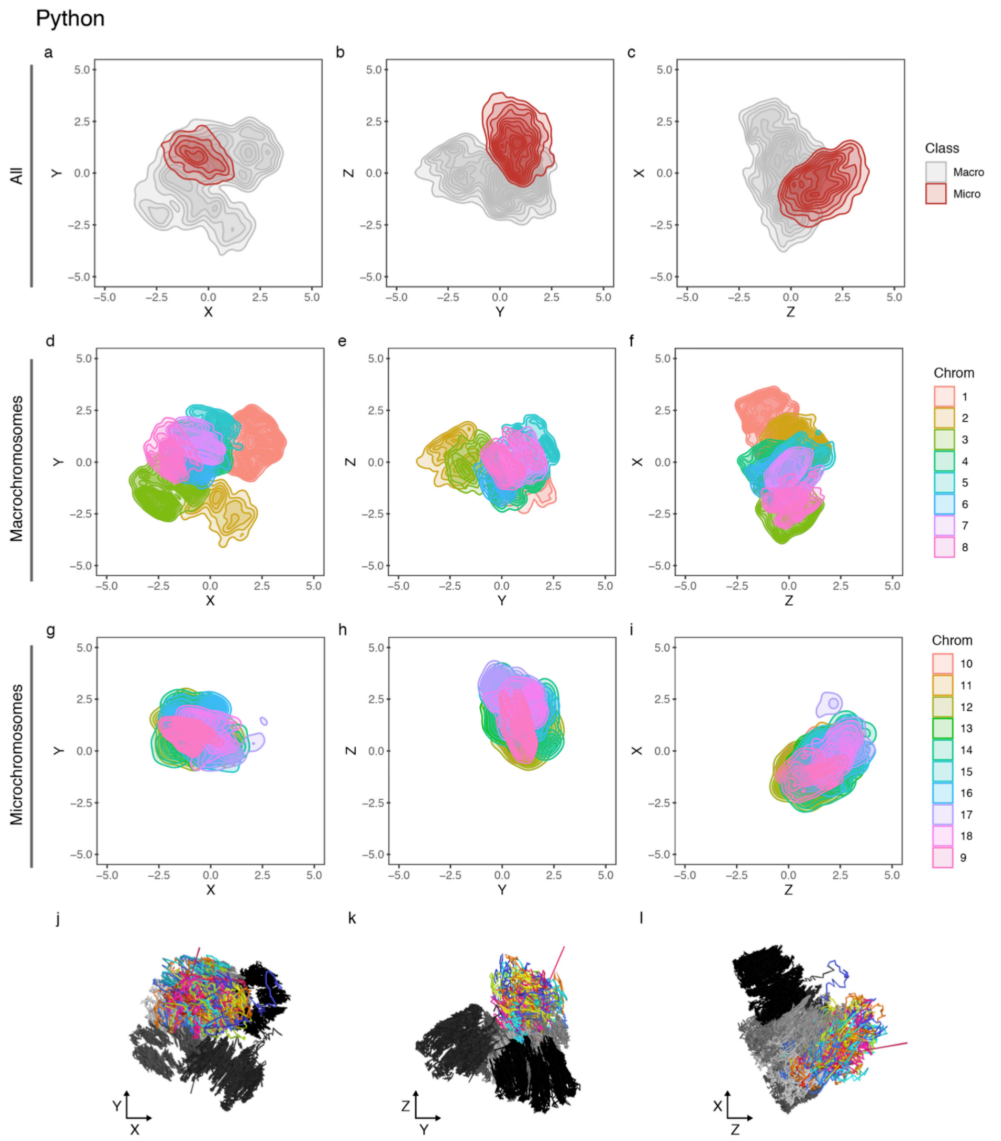
Supplementary Figure 4. 3D interpretation of Prairie Chicken Hi-C interaction data is shown at three distinct orientations (left, center, and right columns), with plots of 2D point density of 3D chromosome models. A-C) 2D point density of all microchromosomes (red) and macrochromosomes (grey). D-F) 2D point density of macrochromosomes only, with each macrochromosome shown as a different color. G-I) 2D point density of microchromosomes only, with each macrochromosome shown as a different color. J-L) 3D models of all chromosomes, with macrochromosomes shown in greyscale and microchromosomes in color.



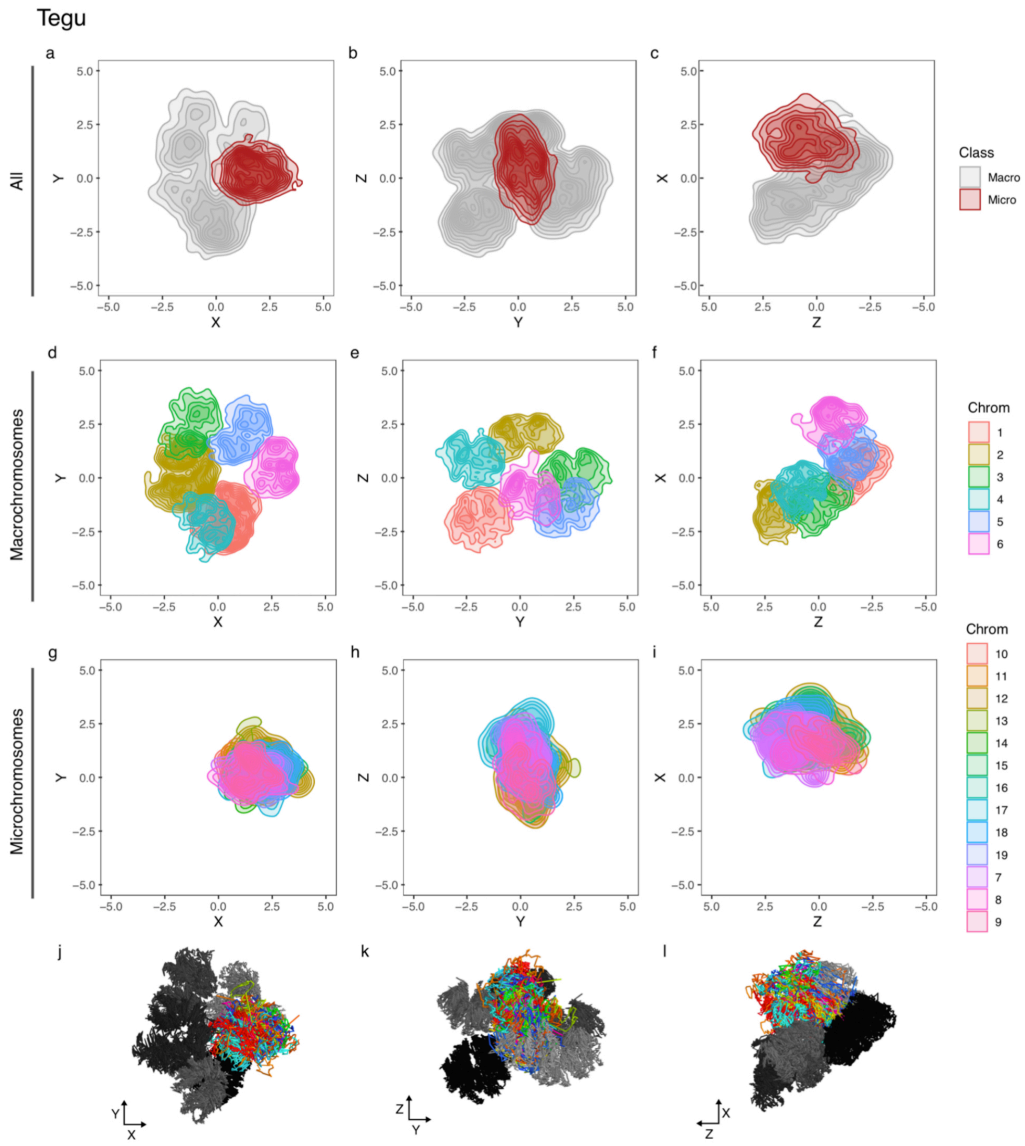
Supplementary Figure 5. 3D interpretation of Sea Turtle Hi-C interaction data is shown at three distinct orientations (left, center, and right columns), with plots of 2D point density of 3D chromosome models. A-C) 2D point density of all microchromosomes (red) and macrochromosomes (grey). D-F) 2D point density of macrochromosomes only, with each macrochromosome shown as a different color. G-I) 2D point density of microchromosomes only, with each macrochromosome shown as a different color. J-L) 3D models of all chromosomes, with macrochromosomes shown in greyscale and microchromosomes in color.



Supplementary Figure 6. 3D interpretation of Rattlesnake Hi-C interaction data is shown at three distinct orientations (left, center, and right columns), with plots of 2D point density of 3D chromosome models. A-C) 2D point density of all microchromosomes (red) and macrochromosomes (grey). D-F) 2D point density of macrochromosomes only, with each macrochromosome shown as a different color. G-I) 2D point density of microchromosomes only, with each macrochromosome shown as a different color. J-L) 3D models of all chromosomes, with macrochromosomes shown in greyscale and microchromosomes in color.



Supplementary Figure 7. 3D interpretation of Python Hi-C interaction data is shown at three distinct orientations (left, center, and right columns), with plots of 2D point density of 3D chromosome models. A-C) 2D point density of all microchromosomes (red) and macrochromosomes (grey). D-F) 2D point density of macrochromosomes only, with each macrochromosome shown as a different color. G-I) 2D point density of microchromosomes only, with each macrochromosome shown as a different color. J-L) 3D models of all chromosomes, with macrochromosomes shown in greyscale and microchromosomes in color.



Supplementary Figure 8. 3D interpretation of Tegu Hi-C interaction data is shown at three distinct orientations (left, center, and right columns), with plots of 2D point density of 3D chromosome models. A-C) 2D point density of all microchromosomes (red) and macrochromosomes (grey). D-F) 2D point density of macrochromosomes only, with each macrochromosome shown as a different color. G-I) 2D point density of microchromosomes only, with each macrochromosome shown as a different color. J-L) 3D models of all chromosomes, with macrochromosomes shown in greyscale and microchromosomes in color.

Supplementary Tables

Supplementary Table 1. Hi-C datasets used in this study.

Species	Common Name	Tissue	Microchromosomes	Source	NCBI Accession
Homo sapiens	Human	Retinal Epithelium	No	(Rao et al., 2014)	GEO: GSE63525
Mus musculus x Mus spretus	Mouse (Patski cell line)	Kidney	No	(Darrow et al., 2016)	GEO: GSE71831
Macaca mulatta	Rhesus Macaque	Fibroblast	No	(Darrow et al., 2016)	GEO: GSE71831
Gallus gallus	Chicken	Mature Erythrocytes	**Yes**	(Fishman et al., 2018)	BioSample: SAMN06555414, SAMN06555414
		Immature Erythrocytes	**Yes**	(Fishman et al. 2018)	BioSample: SAMN10291560, SAMN10291559
		Embryonic Fibroblasts	**Yes**	(Fishman et al. 2018)	BioSample: SAMN06555417, SAMN06555416
Tympanuchus cupido	Greater Prairie Chicken	Blood	**Yes**	(Dudchenko et al., 2017, 2018; Johnson et al.)	BioSample: SAMN10973758
Chelonia mydas	Green Sea Turtle	Blood	**Yes**	(Dudchenko et al., 2017, 2018; Wang et al., 2013)	BioSample: SAMN10973717
Salvator merianae	Argentine Black and White Tegu	Blood	**Yes**	(Dudchenko et al., 2017, 2018; Roscito et al., 2018)	BioSample: SAMN10973771
Crotalus viridis	Prairie Rattlesnake	Venom gland	**Yes**	(Schield et al., 2019)	BioSample: SAMN07738522
Python bivittatus	Burmese Python	Blood	**Yes**	(Castoe et al., 2013; Dudchenko et al., 2017, 2018)	BioSample: SAMN0973752

Supplementary Table 2. Hi-C mapping and contact statistics output from Juicer Hi-C analysis pipeline.

	Human (Retinal Epithelium)	**Mouse (Kidney)**	**Macaque (Fibroblast)**	**Chicken (Mature Erythrocytes)**	**Chicken (Immature Erythrocytes)**	**Chicken (Embryonic Fibroblasts)**	**Prairie Chicken (Blood)**	**Sea Turtle (Blood)**	**Tegu (Blood)**	**Rattlesnake (Venom Gland)**	**Python (Blood)**
Sequenced Read Pairs	626,007,610	580,083,428	701,157,628	261,175,632	358,570,027	276,677,613	118,671,947	135,271,653	140,870,003	195,378,673	274,127,665
Normal Paired (% Sequenced Reads)	305,144,884 (48.74%)	379,398,498 (65.40%)	579,963,191 (82.72%)	173,992,150 (66.62%)	174,107,133 (48.56%)	217,640,966 (78.66%)	100,090,013 (84.34%)	49,344,486 (36.48%)	73,786,879 (52.38%)	124,506,379 (63.73%)	129,656,423 (47.30%)
Chimeric Paired (% Sequenced Reads)	242,920,080 (38.80%)	0 (0.00%)	81,842,940 (11.67%)	77,562,745 (29.70%)	146,084,058 (40.74%)	44,525,771 (16.09%)	10,283,192 (8.67%)	63,913,102 (47.25%)	56,100,941 (39.82%)	44,190,474 (22.62%)	123,903,833 (45.20%)
Chimeric Ambiguous (% Sequenced Reads)	71,331,847 (11.39%)	0 (0.00%)	18,541,357 (2.64%)	3,632,523 (1.39%)	8,650,141 (2.41%)	2,650,906 (0.96%)	1,372,378 (1.16%)	18,123,962 (13.40%)	7,560,451 (5.37%)	18,225,522 (9.33%)	16,051,407 (5.86%)
Unmapped (% Sequenced Reads)	6,610,799 (1.06%)	200,684,930 (34.60%)	20,810,140 (2.97%)	5,988,214 (2.29%)	29,728,695 (8.29%)	11,859,970 (4.29%)	6,926,364 (5.84%)	3,890,103 (2.88%)	3,421,732 (2.43%)	8,456,298 (4.33%)	4,516,002 (1.65%)
Alignable (% Sequenced Reads)	548,064,964 (87.55%)	379,398,498 (65.40%)	661,806,131 (94.39%)	251,554,895 (96.32%)	320,191,191 (89.30%)	262,166,737 (94.76%)	110,373,205 (93.01%)	113,257,588 (83.73%)	129,887,820 (92.20%)	168,696,853 (86.34%)	253,560,256 (92.50%)
Hi-C Contacts (% Sequenced Reads; % Unique Reads)	395,343,162 (63.15% / 79.27%)	194,327,529 (33.50% / 51.91%)	407,127,169 (58.06% / 63.19%)	202,114,957 (77.39% / 89.56%)	240,961,183 (67.20% / 80.68%)	135,149,878 (48.85% / 66.30%)	98,041,793 (82.62% / 90.40%)	74,709,784 (55.23% / 74.71%)	98,990,997 (70.27% / 84.18%)	74,918,770 (38.35% / 66.61%)	192,482,126 (70.22%; 86.76%)
Inter-chromosomal (% Sequenced Reads; % Unique Reads)	66,570,270 (10.63% / 13.35%)	36,578,807 (6.31% / 9.77%)	86,740,307 (12.37% / 13.46%)	72,419,139 (27.73% / 32.09%)	67,550,997 (18.84% / 22.62%)	12,407,240 (4.48% / 6.09%)	34,532,138 (29.10% / 31.84%)	28,883,745 (21.35% / 28.88%)	32,069,488 (22.77% / 27.27%)	5,809,147 (2.97% / 5.16%)	59,111,119 (21.56%; 26.64%)
Intra-chromosomal (% Sequenced Reads; % Unique Reads)	328,772,892 (52.52% / 65.92%)	157,748,722 (27.19% / 42.14%)	320,386,862 (45.69% / 49.73%)	129,695,818 (49.66% / 57.47%)	173,410,186 (48.36% / 58.06%)	122,742,638 (44.36% / 60.21%)	63,509,655 (53.52% / 58.56%)	45,826,039 (33.88% / 45.82%)	66,921,509 (47.51% / 56.91%)	69,109,623 (35.37% / 61.45%)	133,371,007 (48.65%; 60.12%)
Short Range (<20Kb) (% Sequenced Reads; % Unique Reads)	113,928,677 (18.20% / 22.84%)	40,521,898 (6.99% / 10.82%)	95,758,183 (13.66% / 14.86%)	29,548,271 (11.31% / 13.09%)	43,758,213 (12.20% / 14.65%)	63,552,780 (22.97% / 31.18%)	20,021,749 (16.87% / 18.46%)	17,276,086 (12.77% / 17.28%)	20,976,086 (14.89% / 17.84%)	54,160,982 (27.72% / 48.15%)	36,099,196 (13.17%; 16.27%)
Long Range (>20Kb) (% Sequenced Reads; % Unique Reads)	214,843,996 (34.32% / 43.08%)	117,226,658 (20.21% / 31.31%)	224,628,009 (32.04% / 34.87%)	100,147,540 (38.34% / 44.38%)	129,651,970 (36.16% / 43.41%)	59,189,746 (21.39% / 29.04%)	43,487,829 (36.65% / 40.10%)	28,549,941 (21.11% / 28.55%)	45,945,389 (32.62% / 39.07%)	14,948,351 (7.65% / 13.29%)	97,271,750 (35.48%; 43.84%)

REFERENCES

Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311–316.

Aguirre, V., Werner, E.D., Giraud, J., Lee, Y.H., Shoelson, S.E., and White, M.F. (2002). Phosphorylation of Ser307 in insulin receptor substrate-1 blocks interactions with the insulin receptor and inhibits insulin action. J. Biol. Chem. *277*, 1531–1537.

Aird, S.D., Arora, J., Barua, A., Qiu, L., Terada, K., and Mikheyev, A.S. (2017). Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. Genome Biol. Evol. *9*, 2640–2649.

Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Banet, J.F., Billis, K., Girón, C.G., and Hourlier, T. (2016). The Ensembl gene annotation system. Database *2016*, baw093.

Aksamitiene, E., Kiyatkin, A., and Kholodenko, B.N. (2012). Cross-talk between mitogenic Ras/MAPK and survival PI3K/Akt pathways: a fine balance. Biochem. Soc. Trans. *40*, 139–146.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Andrew, A.L., Card, D.C., Ruggiero, R.P., Schield, D.R., Adams, R.H., Pollock, D.D., Secor, S.M., and Castoe, T.A. (2015). Rapid changes in gene expression direct rapid shifts in intestinal form and function in the Burmese python after feeding. Physiol Genomics *47*, 147–157.

Andrew, A.L., Perry, B.W., Card, D.C., Schield, D.R., Ruggiero, R.P., McGaugh, S.E., Choudhary, A., Secor, S.M., and Castoe, T.A. (2017). Growth and stress response mechanisms underlying post-feeding regenerative organ growth in the Burmese python. BMC Genomics *18*, 338.

Arai, K., Lee, S., and Lo, E.H. (2003). Essential role for ERK mitogen-activated protein kinase in matrix metalloproteinase-9 regulation in rat cortical astrocytes. Glia *43*, 254–264.

Axelsson, E., Webster, M.T., Smith, N.G.C., Burt, D.W., and Ellegren, H. (2005). Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res. *15*, 120–125.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. *14*, 283–291.

Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., Webster, M.T., Öst, T., Schneider, M., and Kempenaers, B. (2010). The recombination landscape of the zebra finch Taeniopygia guttata genome. Genome Res. *20*, 485–495.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in bipolymers.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Barash, H., Gross, E.R., Edrei, Y., Ella, E., Israel, A., Cohen, I., Corchia, N., Ben-Moshe, T., Pappo, O., and Pikarsky, E. (2010). Accelerated carcinogenesis following liver regeneration is associated with chronic inflammation-induced double-strand DNA breaks. Proc. Natl. Acad. Sci. *107*, 2207–2212.

Bashkirova, E., and Lomvardas, S. (2019). Olfactory receptor genes make the case for inter-chromosomal interactions. Curr. Opin. Genet. Dev. *55*, 106–113.

Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., and Braun, T. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat. Commun. *11*, 1–11.

Berchtold, D., Fesser, S., Bachmann, G., Kaiser, A., Eilert, J.-C., Frohns, F., Sadoni, N., Muck, J., Kremmer, E., and Eick, D. (2011). Nuclei of chicken neurons in tissues and three-dimensional cell cultures are organized into distinct radial zones. Chromosom. Res. *19*, 165–182.

Beyer, T.A., and Werner, S. (2008). The cytoprotective Nrf2 transcription factor controls insulin receptor signaling in the regenerating liver. Cell Cycle *7*, 874–878.

Biddie, S.C., John, S., Sabo, P.J., Thurman, R.E., Johnson, T.A., Schiltz, R.L., Miranda, T.B., Sung, M.-H., Trump, S., and Lightman, S.L. (2011). Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. Mol. Cell *43*, 145–155.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., and Hochreiter, S. (2015). msa: an R package for multiple sequence alignment. Bioinformatics *31*, 3997–3999.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., and

Speicher, M.R. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol. *3*, e157.

Boyce, M., Bryant, K.F., Jousse, C., Long, K., Harding, H.P., Scheuner, D., Kaufman, R.J., Ma, D.W., Coen, D.M., Ron, D., et al. (2005). A selective inhibitor-of eIF2 alpha dephosphorylation protects cells from ER stress. Science (80-.). *307*, 935–939.

Braun, S., Keller, U.A.D., Steiling, H., and Werner, S. (2004). Fibroblast growth factors in epithelial repair and cytoprotection. Philos. Trans. R. Soc. London Ser. B-Biological Sci. *359*, 753–757.

Breton, S., and Brown, D. (2007). New insights into the regulation of V-ATPase-dependent proton secretion. Am. J. Physiol. Physiol. *292*, F1–F10.

Brockes, J.P. (1997). Amphibian limb regeneration: rebuilding a complex structure. Science (80-.). *276*, 81–87.

Brockes, J.P., and Kumar, A. (2002). Plasticity and reprogramming of differentiated cells in amphibian regeneration. Nat. Rev. Mol. Cell Biol. *3*, 566–574.

Brown, D., Hirsch, S., and Gluck, S. (1988). Localization of a proton-pumping ATPase in rat kidney. J. Clin. Invest. *82*, 2114–2126.

Brown, D., Lui, B., Gluck, S., and Sabolic, I. (1992). A plasma membrane proton ATPase in specialized cells of rat epididymis. Am. J. Physiol. Physiol. *263*, C913–C916.

Brunson, J.C. (2018). ggalluvial: Alluvial Diagrams in "ggplot2."

Burt, D.W. (2002). Origin and evolution of avian microchromosomes. Cytogenet. Genome Res. *96*, 97–112.

Butts, C.T. (2008). network: a Package for Managing Relational Data in R. J. Stat. Softw. *24*, 1–36.

Carneiro, S.M., Pinto, V.R., Jared, C., Lula, L., Faria, F.P., and Sesso, A. (1991). Morphometric studies on venom secretory cells from Bothrops jararacussu (jararacuçu) before and after venom extraction. Toxicon *29*, 569–580.

Casewell, N.R., Wagstaff, S.C., Harrison, R.A., Renjifo, C., and Wüster, W. (2011). Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Mol. Biol. Evol. *28*, 2637–2649.

Casewell, N.R., Huttley, G.A., and Wuster, W. (2012). Dynamic evolution of venom proteins in squamate reptiles. Nat Commun *3*, 1066.

Casewell, N.R., Wuster, W., Vonk, F.J., Harrison, R.A., and Fry, B.G. (2013). Complex cocktails: the evolutionary novelty of venoms. Trends Ecol. Evol. *28*, 219–229.

Casewell, N.R., Wagstaff, S.C., Wüster, W., Cook, D.A.N., Bolton, F.M.S., King, S.I., Pla, D., Sanz, L., Calvete, J.J., and Harrison, R.A. (2014). Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. Proc. Natl. Acad. Sci. *111*, 9205–9210.

Casewell, N.R., Jackson, T.N.W., Laustsen, A.H., and Sunagar, K. (2020). Causes and consequences of snake venom variation. Trends Pharmacol. Sci.

Castoe, T.A., Jiang, Z.J., Gu, W., Wang, Z.O., and Pollock, D.D. (2008). Adaptive evolution and functional redesign of core metabolic proteins in snakes. PLoS One *3*, e2201.

Castoe, T.A., de Koning, A.P., Kim, H.M., Gu, W., Noonan, B.P., Naylor, G., Jiang, Z.J., Parkinson, C.L., and Pollock, D.D. (2009). Evidence for an ancient adaptive episode of convergent molecular evolution. Proc. Natl. Acad. Sci. U. S. A. *106*, 8986–8991.

Castoe, T.A., de Koning, A.P.J., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero, R.P., Degner, J.F., Daza, J.M., Gu, W., et al. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. *110*, 20645–20650.

Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G.M., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., and Doench, J.G. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell *24*, 777–790.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science (80-.). *351*, 1083–1087.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71.

Conesa, A., Nueda, M.J., Ferrer, A., and Talon, M. (2006). maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. Bioinformatics *22*, 1096–1102.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940.

Cox, C.L., and Secor, S.M. (2008). Matched regulation of gastrointestinal performance in the Burmese python, Python molurus. J. Exp. Biol. *211*, 1131–1140.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. *2*, 292.

Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schröck, E., Speicher, M.R., Mathieu, U., Jauch, A., and Emmerich, P. (1993). Role of chromosome territories in the functional compartmentalization of the cell nucleus. In Cold Spring Harbor Symposia on Quantitative Biology, (Cold Spring Harbor Laboratory Press), pp. 777–792.

Cuadrado, A., Martín-Moldes, Z., Ye, J., and Lastres-Becker, I. (2014). Transcription factors NRF2 and NF-κB are coordinated effectors of the Rho family, GTP-binding protein RAC1 during inflammation. J. Biol. Chem. *289*, 15244–15258.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. Nucleic Acids Res *43*, D662-9.

Currier, R.B., Calvete, J.J., Sanz, L., Harrison, R.A., Rowley, P.D., and Wagstaff, S.C. (2012). Unusual stability of messenger RNA in snake venom reveals gene expression dynamics of venom replenishment. PLoS One *7*.

Darrow, E.M., Huntley, M.H., Dudchenko, O., Stamenova, E.K., Durand, N.C., Sun, Z., Huang, S.-C., Sanborn, A.L., Machol, I., and Shamim, M. (2016). Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc. Natl. Acad. Sci. *113*, E4504–E4512.

Davis, A.C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes Dev *7*, 671–682.

Deakin, J.E., and Ezaz, T. (2019). Understanding the evolution of reptile chromosomes through applications of combined cytogenetics and genomics approaches. Cytogenet. Genome Res. *157*, 7–20.

Delorme, S.L., Lungu, I.M., and Vickaryous, M.K. (2012). Scar-free wound healing and regeneration following tail loss in the leopard gecko, Eublepharis macularius. Anat. Rec. Adv. Integr. Anat. Evol. Biol. *295*, 1575–1595.

Desbois-Mouthon, C., Wendum, D., Cadoret, A., Rey, C., Leneuve, P., Blaise, A., Housset, C., Tronche, F., Le Bouc, Y., and Holzenberger, M. (2006). Hepatocyte proliferation during liver regeneration is impaired in mice with liver-specific IGF-1R knockout. FASEB J. *20*, 773–775.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. J. Veg. Sci. *14*, 927–930.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dowell, N.L., Giorgianni, M.W., Kassner, V.A., Selegue, J.E., Sanchez, E.E., and Carroll, S.B. (2016). The deep origin and recent loss of venom toxin genes in rattlesnakes. Curr. Biol. *26*, 2434–2445.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., and Aiden, A.P. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science (80-.). *356*, 92–95.

Dudchenko, O., Shamim, M.S., Batra, S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., St Hilaire, B.G., Yao, W., and Stamenova, E. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. BioRxiv 254797.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016a). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016b). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101.

Effron, M., Griner, L., and Benirschke, K. (1977). Nature and rate of neoplasia found in captive wild mammals, birds, and reptiles at necropsy. J. Natl. Cancer Inst. *59*, 185–198.

Ellison, C.E., and Bachtrog, D. (2013). Dosage compensation via transposable element mediated rewiring of a regulatory network. Science (80-.). *342*, 846–850.

English, J.M., Pearson, G., Baer, R., and Cobb, M.H. (1998). Identification of substrates and regulators of the mitogen-activated protein kinase ERK5 using chimeric protein kinases. J Biol Chem *273*, 3854–3860.

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics *7*, 191.

Espenshade, P.J. (2006). SREBPs: sterol-regulated transcription factors. J. Cell Sci. *119*, 973–976.

Fane, M., Harris, L., Smith, A.G., and Piper, M. (2017). Nuclear factor one transcription factors as epigenetic regulators in cancer. Int. J. Cancer *140*, 2634–2641.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. *9*, 397–405.

Fisher, F.M., and Maratos-Flier, E. (2016). Understanding the physiology of FGF21. Annu. Rev. Physiol. *78*, 223–241.

Fishman, V., Battulin, N., Nuriddinov, M., Maslova, A., Zlotina, A., Strunov, A., Chervyakova, D., Korablev, A., Serov, O., and Krasikova, A. (2018). 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. Nucleic Acids Res. *47*, 648–665.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., and Patwardhan, T.A. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nat. Genet. *51*, 1664–1669.

Gearing, L.J., Cumming, H.E., Chapman, R., Finkel, A.M., Woodhouse, I.B., Luu, K., Gould, J.A., Forster, S.C., and Hertzog, P.J. (2019). CiiiDER: A tool for predicting and analysing transcription factor binding sites. PLoS One *14*, e0215495.

Giorgianni, M.W., Dowell, N.L., Griffin, S., Kassner, V.A., Selegue, J.E., and Carroll, S.B. (2020). The origin and diversification of a novel protein family in venomous snakes. Proc. Natl. Acad. Sci. *117*, 10911–10920.

Gomez, N., Erazo, T., and Lizcano, J.M. (2016). ERK5 and Cell Proliferation: Nuclear Localization Is What Matters. Front. Cell Dev. Biol. *4*.

Granger, D., Marsolais, M., Burry, J., and Laprade, R. (2002). V-type H+-ATPase in the human eccrine sweat duct: immunolocalization and functional demonstration. Am. J. Physiol. Physiol. *282*, C1454–C1460.

Grossman, S.R., Engreitz, J., Ray, J.P., Nguyen, T.H., Hacohen, N., and Lander, E.S. (2018). Positional specificity of different transcription factor classes within enhancers. Proc. Natl. Acad. Sci. *115*, E7222–E7230.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. *8*, 1–9.

Gurtner, G.C., Werner, S., Barrandon, Y., and Longaker, M.T. (2008). Wound repair and regeneration. Nature *453*, 314–321.

Gutiérrez, J.M., Calvete, J.J., Habib, A.G., Harrison, R.A., Williams, D.J., and Warrell, D.A. (2017). Snakebite envenoming. Nat. Rev. Dis. Prim. *3*, 1–21.

Habermann, F.A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., and Solovei, I. (2001). Arrangements of macro-and microchromosomes in chicken cells. Chromosom. Res. *9*, 569–584.

Handcock, M.S., Hunter, D.R., Butts, C.T., Goodreau, S.M., and Morris, M. (2008). statnet: Software tools for the representation, visualization, analysis and simulation of network data. J. Stat. Softw. *24*, 1548.

Hargreaves, A.D., Swain, M.T., Hegarty, M.J., Logan, D.W., and Mulley, J.F. (2014a). Genomic and transcriptomic insights into the regulation of snake venom production. BioRxiv 8474.

Hargreaves, A.D., Swain, M.T., Hegarty, M.J., Logan, D.W., and Mulley, J.F. (2014b). Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol.

6, 2088–2095.

Hayes, J.D., and Ashford, M.L.J. (2012). Nrf2 Orchestrates Fuel Partitioning for Cell Proliferation. Cell Metab. *16*, 139–141.

Helmstetter, C., Reix, N., T'Flachebba, M., Pope, R.K., Secor, S.M., Le Maho, Y., and Lignot, J.H. (2009). Functional changes with feeding in the gastro-intestinal epithelia of the Burmese python (Python molurus). Zool. Sci *26*, 632–638.

Hirosumi, J., Tuncman, G., Chang, L., Görgün, C.Z., Uysal, K.T., Maeda, K., Karin, M., and Hotamisligil, G.S. (2002). A central role for JNK in obesity and insulin resistance. Nature *420*, 333.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Holding, M.L., Biardi, J.E., and Gibbs, H.L. (2016). Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. Proc. R. Soc. B Biol. Sci. *283*, 20152841.

Ignatiadis, N., Klaus, B., Zaugg, J.B., and Huber, W. (2016). Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. Nat. Methods *13*, 577.

International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature *432*, 695–716.

Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., and Taskiran, I.I. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat. Genet. *50*, 1011–1020.

Jia, Y., Chng, W.-J., and Zhou, J. (2019). Super-enhancers: critical roles and therapeutic targets in hematologic malignancies. J. Hematol. Oncol. *12*, 77.

Johnson, J., Novak, B.J., Athrey, G., Shapiro, B., and Phelan, R. Whole genome sequence analysis reveals evolutionary history of extinct Heath Hen. Unpublished.

Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., and Belmonte, J.C.I. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. Nature *464*, 606–609.

Jorge, R.J.B., Monteiro, H.S.A., Gonçalves-Machado, L., Guarnieri, M.C., Ximenes, R.M., Borges-Nojosa, D.M., Karla, P. de O., Zingali, R.B., Corrêa-Netto, C., and Gutiérrez, J.M. (2015). Venomics and antivenomics of Bothrops erythromelas from five geographic populations within the Caatinga ecoregion of northeastern Brazil. J. Proteomics *114*, 93–114.

Juárez, P., Comas, I., González-Candelas, F., and Calvete, J.J. (2008). Evolution of snake venom disintegrins by positive Darwinian selection. Mol. Biol. Evol. *25*, 2391–2407.

Junqueira-de-Azevedo, I.L.M., Bastos, C.M.V., Ho, P.L., Luna, M.S., Yamanouye, N., and Casewell, N.R. (2015). Venom-related transcripts from Bothrops jararaca tissues provide novel molecular insights into the production and evolution of snake venom. Mol. Biol. Evol. *32*, 754–766.

Kamata, H., Honda, S., Maeda, S., Chang, L., Hirata, H., and Karin, M. (2005). Reactive oxygen species promote TNFα-induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. Cell *120*, 649–661.

Kannan, S., Whitehead, K.J., Wang, L., Gomes, A. V, Litwin, S.E., Kensler, T.W., Abel, E.D., Hoidal, J.R., and Soorappan, R.N. (2013). Nrf2 Deficiency Prevents Reductive Stress Induced Hypertrophic Cardiomyopathy. Free Radic. Biol. Med. *65*, S83–S83.

Kaser, A., Adolph, T.E., and Blumberg, R.S. (2013). The unfolded protein response and gastrointestinal disease. In Seminars in Immunopathology, (Springer), pp. 307–319.

Kato, Y., Chao, T.H., Hayashi, M., Tapping, R.I., and Lee, J.D. (2000). Role of BMK1 in regulation of growth factor-induced cellular responses. Immunol Res *21*, 233–237.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kensler, T.W., Wakabayash, N., and Biswal, S. (2007). Cell survival responses to environmental stresses via the Keap1-Nrf2-ARE pathway. Annu. Rev. Pharmacol. Toxicol. *47*, 89–116.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.

Kerchove, C.M., Carneiro, S.M., Markus, R.P., and Yamanouye, N. (2004). Stimulation of the α-adrenoceptor triggers the venom production cycle in the venom gland of Bothrops jararaca. J. Exp. Biol. *207*, 411–416.

Kerchove, C.M., Luna, M.S.A., Zablith, M.B., Lazari, M.F.M., Smaili, S.S., and Yamanouye, N. (2008a). α 1-Adrenoceptors trigger the snake venom production cycle in secretory cells by activating phosphatidylinositol 4, 5-bisphosphate hydrolysis and ERK signaling pathway. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. *150*, 431–437.

Kerchove, C.M., Luna, M.S.A., Zablith, M.B., Lazari, M.F.M., Smaili, S.S., and Yamanouye, N. (2008b). α1-adrenoceptors trigger the snake venom production cycle in secretory cells by activating phosphatidylinositol 4, 5-bisphosphate hydrolysis and ERK signaling pathway. Comp. Biochem. Physiol. Part A Mol. Integr. Physiol. *150*, 431–437.

Khan, H.A., and Margulies, C.E. (2019). The role of mammalian Creb3-like transcription factors in response to nutrients. Front. Genet. *10*, 591.

Kikuchi, K. (2014). Advances in understanding the mechanism of zebrafish heart regeneration. Stem Cell Res. *13*, 542–555.

Kochva, E., Oron, U., Ovadia, M., Simon, T., and Bdolah, A. (1980). Venom glands, venom synthesis, venom secretion and evolution BT - Natural toxins. In Natural Toxins, pp. 3–12.

Kochva, E., Tönsing, L., Louw, A.I., Liebenberg, N., and Visser, L. (1982). Biosynthesis, secretion and in vivo isotopic labelling of venom of the Egyptian cobra, Naja haje annulifera. Toxicon *20*, 615–635.

Kolde, R. (2012). Pheatmap: pretty heatmaps. R Packag. Version *61*.

Koludarov, I., Jackson, T.N.W., Suranse, V., Pozzi, A., Sunagar, K., and Mikheyev, A.S. (2020). Reconstructing the evolutionary history of a functionally diverse gene 2 family reveals complexity at the genetic origins of novelty. BioRxiv 583344.

Kosak, S.T., Scalzo, D., Alworth, S. V, Li, F., Palmer, S., Enver, T., Lee, J.S.J., and Groudine, M. (2007). Coordinate gene regulation during hematopoiesis is related to genomic organization. PLoS Biol. *5*, e309.

Krämer, A., Green, J., Pollard Jr, J., and Tugendreich, S. (2013). Causal analysis approaches in ingenuity pathway analysis. Bioinformatics *30*, 523–530.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. *34*, 1812–1819.

Kurz, A., Lampel, S., Nickolenko, J.E., Bradl, J., Benner, A., Zirbel, R.M., Cremer, T., and Lichter, P. (1996). Active and inactive genes localize preferentially in the periphery of chromosome territories. J. Cell Biol. *135*, 1195–1205.

Laplante, M., and Sabatini, D.M. (2009). mTOR signaling at a glance. J. Cell Sci. *122*, 3589–3594.

Laplante, M., and Sabatini, D.M. (2012). mTOR Signaling in Growth Control and Disease. Cell *149*, 274–293.

Layer, R.M., Pedersen, B.S., DiSera, T., Marth, G.T., Gertz, J., and Quinlan, A.R. (2018). GIGGLE: a search engine for large-scale integrated genome analysis. Nat. Methods *15*, 123.

Lee, J., Giordano, S., and Zhang, J.H. (2012). Autophagy, mitochondria and oxidative stress: cross-talk and redox signalling. Biochem. J. *441*, 523–540.

Lewis, K.N., Mele, J., Hayes, J.D., and Buffenstein, R. (2010). Nrf2, a guardian of healthspan and gatekeeper of species longevity. Integr. Comp. Biol. *50*, 829–843.

Lewis, K.N., Wason, E., Edrey, Y.H., Kristan, D.M., Nevo, E., and Buffenstein, R. (2015). Regulation of Nrf2 signaling and longevity in naturally long-lived rodents. Proc. Natl. Acad. Sci. *112*, 3722–3727.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

Liao, Y., Smyth, G.K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., and Dorschner, M.O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (80-.). *326*, 289–293.

Lignot, J.H., Helmstetter, C., and Secor, S.M. (2005). Postprandial morphological response of the intestinal epithelium of the Burmese python (Python molurus). Comp Biochem Physiol A Mol Integr Physiol *141*, 280–291.

LoPiccolo, J., Blumenthal, G.M., Bernstein, W.B., and Dennis, P.A. (2008). Targeting the PI3K/Akt/mTOR pathway: effective combinations and clinical considerations. Drug Resist Updat *11*, 32–50.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

Lozito, T.P., and Tuan, R.S. (2017). Lizard tail regeneration as an instructive model of enhanced healing capabilities in an adult amniote. Connect. Tissue Res. *58*, 145–154.

De Lucca, F.L., Haddad, A., Kochva, E., Rothschild, A.M., and Valeri, V. (1974). Protein synthesis and morphological changes in the secretory epithelium of the venom gland of Crotalus durissus terrificus at different times after manual extraction of venom. Toxicon *12*, 361–368.

Luna, M.S.A., Hortencio, T.M.A., Ferreira, Z.S., and Yamanouye, N. (2009). Sympathetic outflow activates the venom gland of the snake Bothrops jararaca by regulating the activation of transcription factors and the synthesis of venom gland proteins. J. Exp. Biol. *212*, 1535–1543.

Lynch, V.J., Nnamani, M.C., Kapusta, A., Brayer, K., Plaza, S.L., Mazur, E.C., Emera, D., Sheikh, S.Z., Grützner, F., and Bauersachs, S. (2015). Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. Cell Rep. *10*, 551–561.

Maass, P.G., Barutcu, A.R., and Rinn, J.L. (2019). Interchromosomal interactions: A genomic love story of kissing chromosomes. J Cell Biol *218*, 27–38.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697.

Mackessy, S.P. (1991). Morphology and ultrastructure of the venom glands of the northern Pacific rattlesnake Crotalus viridis oreganus. J. Morphol. *208*, 109–128.

Mackessy, S.P. (2010). Evolutionary trends in venom composition in the western rattlesnakes (Crotalus viridis sensu lato): toxicity vs. tenderizers. Toxicon *55*, 1463–1474.

Mackessy, S.P. (2021). Handbook of Venoms and Toxins of Reptiles, 2nd Ed. S.P. Mackessy, ed. (Boca Raton, FL: CRC Press/Taylor & Francis Group), p. 652.

Mackessy, S.P., and Baxter, L.M. (2006). Bioweapons synthesis and storage: the venom gland of front-fanged snakes. Zool. Anzeiger-A J. Comp. Zool. *245*, 147–159.

Margres, M.J., Rautsaw, R.M., Strickland, J.L., Mason, A.J., Schramer, T.D., Hofmann, E.P., Stiers, E., Ellsworth, S.A., Nystrom, G.S., and Hogan, M.P. (2021). The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. Proc. Natl. Acad. Sci. *118*.

Mateyak, M.K., Obaya, A.J., Adachi, S., and Sedivy, J.M. (1997). Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination. Cell Growth Differ *8*, 1039–1048.

Matsuda, M., Korn, B.S., Hammer, R.E., Moon, Y.A., Komuro, R., Horton, J.D., Goldstein, J.L., Brown, M.S., and Shimomura, I. (2001). SREBP cleavage-activating protein (SCAP) is required for increased lipid synthesis in liver induced by cholesterol deprivation and insulin elevation. Genes Dev. *15*, 1206–1216.

McGaugh, S.E., Bronikowski, A.M., Kuo, C.-H., Reding, D.M., Addis, E.A., Flagel, L.E., Janzen, F.J., and Schwartz, T.S. (2015). Rapid molecular evolution across amniotes of the IIS/TOR network. Proc. Natl. Acad. Sci. USA *112*, 7055–7060.

Menzel, E.J., Nicholas, B., Denardo, D.F., and Secor, S.M. (2012). Adaptive regulation of gastrointestinal form and function for the diamondback rattlesnake. In Integrative and Comparative Biology, (OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA), pp. E294–E294.

Mercier, F., Bayle, D., Besancon, M., Joys, T., Shin, J.M., Lewin, M.J.M., Prinz, C., Reuben, M.A., Soumarmon, A., and Wong, H. (1993). Antibody epitope mapping of the gastric H+/K+-ATPase. Biochim. Biophys. Acta (BBA)-Biomembranes *1149*, 151–165.

Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. Cell *128*, 787–800.

Mitsuishi, Y., Taguchi, K., Kawatani, Y., Shibata, T., Nukiwa, T., Aburatani, H., Yamamoto, M., and Motohashi, H. (2012). Nrf2 redirects glucose and glutamine into anabolic pathways in metabolic reprogramming. Cancer Cell *22*, 66–79.

Moshage, H. (1997). Cytokines and the hepatic acute phase response. J. Pathol. *181*, 257–266.

Nakamura, H., Murakami, T., Hattori, S., Sakaki, Y., Ohkuri, T., Chijiwa, T., Ohno, M., and Oda-Ueda, N. (2014). Epithelium specific ETS transcription factor, ESE-3, of Protobothrops flavoviridis snake venom gland transactivates the promoters of venom phospholipase A2 isozyme genes. Toxicon *92*, 133–139.

Niederreiter, L., Fritz, T.M.J., Adolph, T.E., Krismer, A.-M., Offner, F.A., Tschurtschenthaler, M., Flak, M.B., Hosomi, S., Tomczak, M.F., and Kaneider, N.C. (2013). ER stress transcription factor Xbp1 suppresses intestinal tumorigenesis and directs intestinal stem cells. J. Exp. Med. *210*, 2041–2056.

Nishi, T., and Forgac, M. (2002). The vacuolar (H+)-ATPases—nature’s most versatile proton pumps. Nat. Rev. Mol. Cell Biol. *3*, 94–103.

O’Connor, R.E., Kiazim, L., Skinner, B., Fonseka, G., Joseph, S., Jennings, R., Larkin, D.M., and Griffin, D.K. (2019). Patterns of microchromosome organization remain highly conserved throughout avian evolution. Chromosoma *128*, 21–29.

Ohno, S., Muramoto, J., Stenius, C., Christian, L., Kittrell, W.A., and Atkin, N.B. (1969). Microchromosomes in holocephalian, chondrostean and holostean fishes. Chromosoma *26*, 35–40.

Okouchi, M., Okayama, N., Alexander, J.S., and Aw, T.Y. (2006). NRF2-dependent glutamate-L-cysteine ligase catalytic subunit expression mediates insulin protection against hyperglycemia-induced brain endothelial cell apoptosis. Curr. Neurovasc. Res. *3*, 249–261.

Ott, B.D., and Secor, S.M. (2007). Adaptive regulation of digestive performance in the genus Python. J Exp Biol *210*, 340–356.

Pamarthy, S., Kulshrestha, A., Katara, G.K., and Beaman, K.D. (2018). The curious case of vacuolar ATPase: regulation of signaling pathways. Mol. Cancer *17*, 41.

Papaiahgari, S., Zhang, Q., Kleeberger, S.R., Cho, H.Y., and Reddy, S.P. (2006). Hyperoxia stimulates an Nrf2-ARE transcriptional response via ROS-EGFR-PI3K-Akt/ERK MAP kinase signaling in pulmonary epithelial cells. Antioxid Redox Signal *8*, 43–52.

Pappalardo, F., Russo, G., Candido, S., Pennisi, M., Cavalieri, S., Motta, S., McCubrey, J.A., Nicoletti, F., and Libra, M. (2016). Computational Modeling of PI3K/AKT and MAPK Signaling Pathways in Melanoma Cancer. PLoS One *11*.

Pasquesi, G.I.M., Adams, R.H., Card, D.C., Schield, D.R., Corbin, A.B., Perry, B.W., Reyes-Velasco, J., Ruggiero, R.P., Vandewege, M.W., and Shortt, J.A. (2018). Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat. Commun. *9*, 1–11.

Perez-Moreno, M., Jamora, C., and Fuchs, E. (2003). Sticky business: orchestrating cellular signals at adherens junctions. Cell *112*, 535–548.

Perry, B.W., Card, D.C., McGlothlin, J.W., Pasquesi, G.I.M., Adams, R.H., Schield, D.R., Hales, N.R., Corbin, A.B., Demuth, J.P., and Hoffmann, F.G. (2018). Molecular adaptations for sensing and securing prey and insight into amniote genome diversity from the garter snake genome. Genome Biol. Evol. *10*, 2110–2129.

Perry, B.W., Andrew, A.L., Mostafa Kamal, A.H., Card, D.C., Schield, D.R., Pasquesi, G.I.M., Pellegrino, M.W., Mackessy, S.P., Chowdhury, S.M., and Secor, S.M. (2019). Multi-species comparisons of snakes identify coordinated signalling networks underlying post-feeding intestinal regeneration. Proc. R. Soc. B *286*, 20190910.

Perry, B.W., Schield, D.R., Westfall, A.K., Mackessy, S.P., and Castoe, T.A. (2020). Physiological demands and signaling associated with snake venom production and storage illustrated by transcriptional analyses of venom glands. Sci. Rep. *10*, 1–10.

Phanstiel, D.H., Boyle, A.P., Araya, C.L., and Snyder, M.P. (2014). Sushi. R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics *30*, 2808–2810.

Porrello, E.R., Mahmoud, A.I., Simpson, E., Hill, J.A., Richardson, J.A., Olson, E.N., and Sadek, H.A. (2011).

Transient regenerative potential of the neonatal mouse heart. Science (80-.). *331*, 1078–1080.

Poss, K.D., Keating, M.T., and Nechiporuk, A. (2003). Tales of regeneration in zebrafish. Dev. Dyn. *226*, 202–210.

Post, Y., Puschhof, J., Beumer, J., Kerkkamp, H.M., de Bakker, M.A.G., Slagboom, J., de Barbanson, B., Wevers, N.R., Spijkers, X.M., Olivier, T., et al. (2020). Snake venom gland organoids. Cell *180*, 233-247.e21.

Puente, B.N., Kimura, W., Muralidhar, S.A., Moon, J., Amatruda, J.F., Phelps, K.L., Grinsfelder, D., Rothermel, B.A., Chen, R., and Garcia, J.A. (2014). The oxygen-rich postnatal environment induces cardiomyocyte cell-cycle arrest through DNA damage response. Cell *157*, 565–579.

Puschhof, J., Post, Y., Beumer, J., Kerkkamp, H.M., Bittenbinder, M., Vonk, F.J., Casewell, N.R., Richardson, M.K., and Clevers, H. (2021). Derivation of snake venom gland organoids for in vitro venom production. Nat. Protoc. *16*, 1494–1510.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

R Core Team (2014). R: A language and environment for statistical computing.

Rabon, E.C., and Reuben, M.A. (1990). The mechanism and structure of the gastric H, K-ATPase. Annu. Rev. Physiol. *52*, 321–344.

Raghow, R., Yellaturu, C., Deng, X., Park, E.A., and Elam, M.B. (2008). SREBPs: the crossroads of physiological and pathological lipid homeostasis. Trends Endocrinol Metab *19*, 65–73.

Ramachandran, P., and Varoquaux, G. (2011). Mayavi: 3D visualization of scientific data. Comput. Sci. Eng. *13*, 40–51.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., and Lander, E.S. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Reddy, N.M., Kleeberger, S.R., Kensler, T.W., Yamamoto, M., Hassoun, P.M., and Reddy, S.P. (2009). Disruption of Nrf2 impairs the resolution of hyperoxia-induced acute lung injury and inflammation in mice. J. Immunol. *182*, 7264–7271.

Reding, D.M., Addis, E.A., Palacios, M.G., Schwartz, T.S., and Bronikowski, A.M. (2016). Insulin-like signaling (IIS) responses to temperature, genetic background, and growth variation in garter snakes with divergent life histories. Gen. Comp. Endocrinol. *233*, 88–99.

Rees, B.B., Sudradjat, F.A., and Love, J.W. (2001). Acclimation to hypoxia increases survival time of zebrafish, Danio rerio, during lethal hypoxia. J. Exp. Zool. *289*, 266–272.

Reif, M.S., Fisher, C.L., Mackessy, S.P., and Secor, S.M. (2015). Testing the adaptive correlation between feeding habits and digestive physiology for snakes. In Integrative and Comparative Biology, (OXFORD UNIV PRESS INC JOURNALS DEPT, 2001 EVANS RD, CARY, NC 27513 USA), pp. E319–E319.

Riddle, N.C., and Elgin, S.C.R. (2018). The Drosophila dot chromosome: where genes flourish amidst repeats. Genetics *210*, 757–772.

Rieber, L., and Mahony, S. (2017). miniMDS: 3D structural inference from high-resolution Hi-C data. Bioinformatics *33*, i261–i266.

Riquelme, C.A., Magida, J.A., Harrison, B.C., Wall, C.E., Marr, T.G., Secor, S.M., and Leinwand, L.A. (2011). Fatty acids identified in the Burmese python promote beneficial cardiac growth. Science (80-.). *334*, 528–531.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Roesner, A., Hankeln, T., and Burmester, T. (2006). Hypoxia induces a complex response of globin expression in zebrafish (Danio rerio). J. Exp. Biol. *209*, 2129–2137.

Rokyta, D.R., Lemmon, A.R., Margres, M.J., and Aronow, K. (2012). The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus). BMC Genomics *13*, 312.

Rokyta, D.R., Margres, M.J., and Calvin, K. (2015). Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. G3 Genes, Genomes, Genet. *5*, 2375–2382.

Roscito, J.G., Sameith, K., Pippel, M., Francoijs, K.-J., Winkler, S., Dahl, A., Papoutsoglou, G., Myers, G., and Hiller, M. (2018). The genome of the tegu lizard Salvator merianae: combining Illumina, PacBio, and optical mapping data to generate a highly contiguous assembly. Gigascience *7*, giy141.

Rose, R., Golosova, O., Sukhomlinov, D., Tiunov, A., and Prosperi, M. (2019). Flexible design of multiple metagenomics classification pipelines with UGENE. Bioinformatics *35*, 1963–1965.

Rotenberg, D., Bamberger, E.S., and Kochva, E. (1971). Studies on ribonucleic acid synthesis in the venom glands of Vipera palaestinae (Ophidia, Reptilia). Biochem. J. *121*, 609–612.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91–D94.

Saviola, A.J., Pla, D., Sanz, L., Castoe, T.A., Calvete, J.J., and Mackessy, S.P. (2015). Comparative venomics of the Prairie Rattlesnake (Crotalus viridis viridis) from Colorado: Identification of a novel pattern of ontogenetic changes in venom composition and assessment of the immunoreactivity of the commercial antivenom CroFab®. J. Proteomics *121*, 28–43.

Schäfer, M., and Werner, S. (2008). Oxidative stress in normal and impaired wound repair. Pharmacol. Res. *58*, 165–171.

Schield, D.R., Card, D.C., Hales, N.R., Perry, B.W., Pasquesi, G.M., Blackmon, H., Adams, R.H., Corbin, A.B., Smith, C.F., and Ramesh, B. (2019). The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. *29*, 590–601.

Schield, D.R., Pasquesi, G.I.M., Perry, B.W., Adams, R.H., Nikolakis, Z.L., Westfall, A.K., Orton, R.W., Meik, J.M., Mackessy, S.P., and Castoe, T.A. (2020). Snake recombination landscapes are concentrated in functional regions despite PRDM9. Mol. Biol. Evol. *37*, 1272–1294.

Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E., Elberg, A., and Larmarange, J. (2018). Ggally: Extension to ggplot2. R Packag. Version *1*.

Schoonjans, K., Staels, B., and Auwerx, J. (1996). Role of the peroxisome proliferator-activated receptor (PPAR) in mediating the effects of fibrates and fatty acids on gene expression. J Lipid Res *37*, 907–925.

Schwartz, T.S., and Bronikowski, A.M. (2014). Gene expression of components of the insulin/insulin-like signaling pathway in response to heat stress in the garter snake, Thamnophis elegans. J. Iowa Acad. Sci. *121*, 1–4.

Schwartz, T.S., and Bronikowski, A.M. (2016). Evolution and function of the insulin and insulin-like signaling network in ectothermic reptiles: some answers and more questions. Integr. Comp. Biol. *56*, 171–184.

Secor, S.M. (2003). Gastric function and its contribution to the postprandial metabolic response of the Burmese python Python molurus. J. Exp. Biol. *206*, 1621–1630.

Secor, S.M. (2005). Evolutionary and cellular mechanisms regulating intestinal performance of amphibians and reptiles. Integr Comp Biol *45*, 282–294.

Secor, S.M. (2008). Digestive physiology of the Burmese python: broad regulation of integrated performance. J. Exp. Biol. *211*, 3767–3774.

Secor, S.M., and Diamond, J. (1995). Adaptive responses to feeding in Burmese pythons: pay before pumping. J. Exp. Biol. *198*, 1313–1325.

Secor, S.M., and Diamond, J. (1997). Effects of meal size on postprandial responses in juvenile Burmese pythons (Python molurus). Am J Physiol *272*, R902-12.

Secor, S.M., and Diamond, J. (1998). A vertebrate model of extreme physiological regulation. Nature *395*, 659–662.

Secor, S.M., and Diamond, J.M. (2000). Evolution of regulatory responses to feeding in snakes. Physiol Biochem Zool *73*, 123–141.

Secor, S.M., and Ott, B.D. (2007). Adaptive correlation between feeding habits and digestive physiology for boas and pythons. Biol. Boas Pythons 257–268.

Secor, S.M., and White, S.E. (2010). Prioritizing blood flow: cardiovascular performance in response to the competing demands of locomotion and digestion for the Burmese python, Python molurus. J. Exp. Biol. *213*, 78–88.

Secor, S., Choudhary, A., Lundh, M., and Wagner, B. (2014). Is extreme physiology of Burmese pythons relevant to diabetes?(1108.8). FASEB J. *28*, 1108.8.

Secor, S.M., Fehsenfeld, D., Diamond, J., and Adrian, T.E. (2001). Responses of python gastrointestinal regulatory peptides to feeding. Proc. Natl. Acad. Sci. U. S. A. *98*, 13637–13642.

Secor, S.M., Lane, J.S., Whang, E.E., Ashley, S.W., and Diamond, J. (2002). Luminal nutrient signals for intestinal adaptation in pythons. Am J Physiol Gastrointest Liver Physiol *283*, G1298-309.

Seifert, A.W., Monaghan, J.R., Voss, S.R., and Maden, M. (2012). Skin regeneration in adult axolotls: a blueprint for scar-free healing in vertebrates. PLoS One *7*, e32875.

Sen, C.K., and Roy, S. (2008). Redox signals in wound healing. Biochim. Biophys. Acta (BBA)-General Subj. *1780*, 1348–1361.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shao, Y., Qu, Y., Dang, S., Yao, B., and Ji, M. (2013). MiR-145 inhibits oral squamous cell carcinoma (OSCC) cell growth by targeting c-Myc and Cdk6. Cancer Cell Int *13*, 51.

Shibata, H., Chijiwa, T., Oda-Ueda, N., Nakamura, H., Yamaguchi, K., Hattori, S., Matsubara, K., Matsuda, Y., Yamashita, A., and Isomoto, A. (2018). The habu genome reveals accelerated evolution of venom protein genes. Sci. Rep. *8*, 11300.

Shibata, T., Saito, S., Kokubu, A., Suzuki, T., Yamamoto, M., and Hirohashi, S. (2010). Global Downstream

Pathway Analysis Reveals a Dependence of Oncogenic NF-E2-Related Factor 2 Mutation on the mTOR Growth Signaling Pathway. Cancer Res. *70*, 9095–9105.

Skinner, B.M., Völker, M., Ellis, M., and Griffin, D.K. (2009). An appraisal of nuclear organisation in interphase embryonic fibroblasts of chicken, turkey and duck. Cytogenet. Genome Res. *126*, 156–164.

Smith, C.F., and Mackessy, S.P. (2016). The effects of hybridization on divergent venom phenotypes: characterization of venom from Crotalus scutulatus scutulatus× Crotalus oreganus helleri hybrids. Toxicon *120*, 110–123.

Smith, M.H., Ploegh, H.L., and Weissman, J.S. (2011). Road to ruin: targeting proteins for degradation in the endoplasmic reticulum. Science (80-.). *334*, 1086–1090.

Solovyev, V. (2004). Statistical approaches in eukaryotic gene prediction. Handb. Stat. Genet.

Suryamohan, K., Krishnankutty, S.P., Guillory, J., Jevit, M., Schröder, M.S., Wu, M., Kuriakose, B., Mathew, O.K., Perumal, R.C., and Koludarov, I. (2020). The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat. Genet. 1–12.

Sykiotis, G.P., and Bohmann, D. (2008). Keap1/Nrf2 signaling regulates oxidative stress tolerance and lifespan in Drosophila. Dev. Cell *14*, 76–85.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., and Bork, P. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564–577.

Teichmann, S.A., and Babu, M.M. (2004). Gene regulatory network growth by duplication. Nat. Genet. *36*, 492–496.

Tremblay, B.J.-M. (2020). universalmotif: Import, Modify, and Export Motifs with R.

Vinarsky, V., Atkinson, D.L., Stevenson, T.J., Keating, M.T., and Odelberg, S.J. (2005). Normal newt limb regeneration requires matrix metalloproteinase function. Dev. Biol. *279*, 86–98.

Wagner, G.P., and Lynch, V.J. (2010). Evolutionary novelties. Curr. Biol. *20*, R48–R52.

Wall, C.E., Cozza, S., Riquelme, C.A., McCombie, W.R., Heimiller, J.K., Marr, T.G., and Leinwand, L.A. (2011). Whole transcriptome analysis of the fasting and fed Burmese python heart: insights into extreme physiological cardiac adaptation. Physiol. Genomics *43*, 69–76.

Wang, X., and Tournier, C. (2006). Regulation of cellular functions by the ERK5 signalling pathway. Cell Signal *18*, 753–760.

Wang, L., Chen, Y., Sternberg, P., and Cai, J. (2008). Essential roles of the PI3 kinase/Akt pathway in regulating Nrf2-dependent antioxidant functions in the RPE. Invest Ophthalmol Vis Sci *49*, 1671–1678.

Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., Li, C., White, S., Xiong, Z., and Fang, D. (2013). The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat. Genet. *45*, 701–706.

Wardyn, J.D., Ponsford, A.H., and Sanderson, C.M. (2015). Dissecting molecular cross-talk between Nrf2 and NF-κB response pathways. Biochem. Soc. Trans. *43*, 621–626.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Wek, R.C., Jiang, H.Y., and Anthony, T.G. (2006). Coping with stress: eIF2 kinases and translational control. Biochem. Soc. Trans. *34*, 7–11.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wickham, H. (2011). ggplot2. Wiley Interdiscip. Rev. Comput. Stat. *3*, 180–185.

Wong, M.H., Xue, A., Baxter, R.C., Pavlakis, N., and Smith, R.C. (2016). Upstream and Downstream Co-inhibition of Mitogen-Activated Protein Kinase and PI3K/Akt/mTOR Pathways in Pancreatic Ductal Adenocarcinoma. Neoplasia *18*, 425–435.

World Health Organization (2019). Snakebite envenoming: a strategy for prevention and control.

Yamanouye, N., Carneiro, S.M., Scrivano, C.N., and Markus, R.P. (2000). Characterization of β-adrenoceptors responsible for venom production in the venom gland of the snake Bothrops jararaca. Life Sci. *67*, 217–226.

Yang, T., Espenshade, P.J., Wright, M.E., Yabe, D., Gong, Y., Aebersold, R., Goldstein, J.L., and Brown, M.S. (2002). Crucial step in cholesterol homeostasis: Sterols promote binding of SCAP to INSIG-1, a membrane protein that facilitates retention of SREBPs in ER. Cell *110*, 489–500.

Yin, W., Wang, Z., Li, Q., Lian, J., Zhou, Y., Lu, B., Jin, L., Qiu, P., Zhang, P., and Zhu, W. (2016). Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. Nat. Commun. *7*.

Yu, W., He, B., and Tan, K. (2017). Identifying topologically associating domains and subdomains by Gaussian Mixture model And Proportion test. Nat. Commun. *8*, 535.

Zancolli, G., and Casewell, N.R. (2020). Venom systems as models for studying the origin and regulation of evolutionary novelties. Mol. Biol. Evol.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. *33*, W741–W748.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., and Li, W. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, 1–9.

Zhang, Y., McCord, R.P., Ho, Y.-J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908–921.

Zheng, Y., and Wiens, J.J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Mol. Phylogenet. Evol. *94*, 537–547.

Zhou, L., and Gui, J.F. (2002). Karyotypic diversity in polyploid gibel carp, Carassius auratus gibelio Bloch. Genetica *115*, 223–232.