The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype

Mark J. Margres[a,b,c,1], Rhett M. Rautsaw[a], Jason L. Strickland[a,d], Andrew J. Mason[a,e], Tristan D. Schramer[a], Erich P. Hofmann[a,f], Erin Stiers[a], Schyler A. Ellsworth[g], Gunnar S. Nystrom[g], Michael P. Hogan[g], Daniel A. Bartlett[g], Timothy J. Colston[g,h], David M. Gilbert[g], Darin R. Rokyta[g], and Christopher L. Parkinson[a,i,1]

[a]Department of Biological Sciences, Clemson University, Clemson, SC 29634 ; [b]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138; [c]Department of Integrative Biology, University of South Florida, Tampa, FL 33620; [d]Department of Biology, University of South Alabama, Mobile, AL 36688; [e]Present address: Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, 43210; [f]Present address: Science Department, Cape Fear Community College, Wilmington, NC 28401; [g]Department of Biological Science, Florida State University, Tallahassee, FL 32306; [h]Present address: Department of Biology, University of Florida, Gainesville, FL 32611; [i]Department of Forestry, and Environmental Conservation, Clemson University, Clemson, SC 29634

Supporting Information (SI)

SI Materials and Methods.

HMW DNA isolation. Blood was extracted from the caudal vein of an adult male *C. tigris* (52 cm snout-vent length, 56.5 cm total length) collected from Santa Cruz County, Arizona. High-molecular weight (HMW) gDNA was then extracted using a pipette-free protocol. Briefly, fresh blood was centrifuged at 1500 rpm for five minutes, and the serum removed. The pellet was resuspended in 7 mL of lysis buffer (10 mM Tris-Cl pH 8.0, 25 mM EDTA pH 8.0, 0.5% SDS, 20 ug/mL RNase A) and incubated at 37°C for one hour. Fifty μL of 20 mg/mL Proteinase K were then added, and the sample was incubated for another three hours. Phenol-chloroform extraction was then performed in 15 mL phase-lock tubes. HMW gDNA was eluted using 3 M sodium acetate (pH 5.2) and hooked using a glass pipette tip. DNA was stored in 10 mM Tris-HCl (pH 8) at room temperature overnight before being moved to 4°C.

Genome assembly and statistics. CG content was determined with a sliding window approach estimating the proportion of Cs and Gs among non-ambiguous base calls. Gene density was similarly estimated with a sliding window approach, counting the total number of genes with a start site within the window. Repeat content was determined based on repeat masker annotations. Only full matches were considered, and overlapping repeat regions were merged before calculating the proportion of sequences annotated as repeats.

For the RaGOO scaffolded assembly, the proportion of methylated cytosines was calculated from the venom-gland WGBS library from the genome individual (Table S3). Here, reads were mapped to the RaGOO assembly using Bismark (1) as described below. We treated methylation levels (as determined by the Bismark alignment) as a proportion and calculated the sum for all positions in a given 100 kb window as described above. For ATAC-seq coverage, right and left venom-gland libraries for the genome individual (Table S4) were combined and aligned to the RaGOO assembly using Bowtie2 (2) as described below. ATAC-seq coverage was calculated as mean coverage for a given 100 kb window. Summary statistics for the RaGOO assembly shown were visualized in R using the circlize package (3).

RNA sequencing. Extracted total RNA was checked for quality on an Agilent 2100 Bioanalyzer. cDNA library preparation was performed from 1 μL total RNA using the New England Biolabs NEB-Next Ultra RNA library Prep Kit for Illumina (New England Biolabs) following the manufacturer's instructions. Libraries were quality-checked on an Agilent 2100 Bioanalyzer. Concentration was estimated via KAPA qPCR.

As stated in the main manuscript, we tested three different stimulation timepoints prior to gland removal to investigate how dependent venom expression was on last date of venom expulsion. Of the three main individuals investigated using RNA-seq and ATAC-seq, venom expression was stimulated four days prior to gland removal for the genome individual (CLP2741) as this is standard practice (see below; 4), at one day for CLP2752 (i.e., active state), and >30 days (i.e., resting state) for CLP2742. Four days is standard practice in snake venom-gland transcriptomics as venom transcription has been shown to be maximized at this time-point (4), and the four other individuals sequenced were processed at four days.

Transcriptome assembly for genome annotation. To facilitate genome annotation, 25 individual transcriptomes from the genome individual were *de novo* assembled using Extender (5), NGen (6), and Trinity (7) as previously described (8). Assemblies were combined, and the merged assembly produced 3,479,678 transcripts. We removed fragments and chimeric transcripts using ChimeraKiller (https://github.com/masonaj157/ChimeraKiller) and retained full length transcripts with $\geq$ 90% blast match (blastx v. 2.2.31+) to a sequence from the UniProt database (downloaded 22 January 2019); this approach produced 309,105 transcripts.

ATAC-seq, peak calling, and motif analyses. Nine tissues from three individuals (Table S4) were dissected immediately following euthanasia and individually prepped for ATAC-seq as previously described (9, 10) with minor modifications. Briefly, tissues were transferred to cold cell lysis buffer (1M NaCl, 1M MgCl2, Tris-HCL, 20% IGEPAL-360, ultrapure water) and tituated to form a cell suspension. Cells were then counted on an inverted light microscope using a hemocytometer.

Following transposition, clean-up was performed using Qiagen and Zymo MinElute Reaction kits following manufacturer's protocols. All libraries were then PCR amplified for six cycles followed by quality control qPCR. Whether additional PCR cycles were required was determined for each library by plotting the linear Rn∼cycle number and calculating the number of cycles corresponding to 1/4 the maximum florescent intensity as previously described (10); all libraries were PCR amplified for a total of 9–12 cycles. Bead purification and size selection were then used to remove both small (0.9×) and large (0.5×) fragments using Ampure XP beads (Beckman Coulter) following manufacturers protocol for each library. Prior to sequencing, libraries were visualized on a HS DNA Bioanalyzer Chip (Aligent) and quantified via a NGS Library Quantification Kit (KAPA Biosystems). Low quality bases were trimmed with Trim Galore! version 0.4.5 (11). Trimmed reads were quality assessed with FastQC and KAT (12) and aligned against the *de novo* reference genome (excluding the mitochondrial genome) using Bowtie2 (2) and the –very-sensitive flag.

Peaks were independently called for each library using Genrich (13) and MACS2 (14). In Genrich, peaks were called in ATAC-seq mode using the -y and -r flags for each library independently; number of peaks called per library varied from 1,668–14,358. Prior to peak calling in MACS2, duplicate read alignments were removed with Picardtools (15) (not necessary in Genrich). Peaks were then called assuming an effective genome size of 1.6e9; number of peaks called per library varied from 18,911–54,062.

To determine if candidate TFs were up-regulated in the venom gland compared to other body tissues, we used DESeq2 (16) to test for significant differential expression. Our candidate TFs included (1) the TFs identified in the motif analyses in this study, (2) TFs identified as up-regulated in the *C. viridis* venom gland (17), and (3) TF paralogs of (2) from the *C. viridis* genome. For the *C. viridis* TFs, we used the sequences from the *C. viridis* reference genome. For the TFs identified in our motif analysis, we confirmed/predicted coding-sequence gene annotations using the protein sequence from the genomes of *Protobothrops mucrosquamatus* (18) and/or *Pantherophis guttatus* (19) and FGENESH+. We used the coding sequence for each gene as the reference and aligned the merged transcriptome reads for all venom gland ($n = 7$) and body tissue ($n = 27$) transcriptomes using Bowtie2 (2); Bowtie2 was implemented in RSEM v.1.3.1. with default parameters (20). We used the count data from RSEM and imported the data into RStudio v.1.3.1056 with R v.4.0.2 (21) using the tximport function (22). We used a false discover rate of 0.05 and compared the seven venom glands as one treatment to the 27 other tissues as the second treatment. Log-fold change and an adjusted p-value threshold of 0.05 were used to determine (1) which genes were significantly differentially expressed, and (2) direction of the differential expression.

Whole-genome bisulfite sequencing. To determine if methylation differed between tissue types in different regions of the genome, we calculated mean methylation of open regions, 1 kb regions around toxin and nontoxin TSS, toxin genes, nontoxin genes, intergenic regions, and closed regions. Open regions were venom-specific ATAC-seq peaks as described in the main text (Genrich and Macs2 peaks combined). For toxin TSS, we calculated mean methylation levels separately for highly expressed toxin genes (TPM >1,000), lowly expressed toxin genes (TPM <1,000), and all toxin genes. For intergenic regions, we selected the same number and size of regions as the number of genes in the genome and used BEDTools shuffle to place these regions outside of genic regions. For closed regions, we similarly used BEDTools shuffle to select the same number and size of ATAC-seq peak calls.

To search for enriched GO terms among the differentially expressed non-toxin genes, we used ShinyGO (23) using the *Gallus gallus* and *Rattus norvegicus* genomes, a false-discovery rate cutoff of 0.05, and extracting the 30 most significant terms. No significant terms were identified using the *Gallus gallus* genome. All results shown used the *Rattus norvegicus* genome. Terms could be identified more than once as we accepted significant terms across the GO categorical databases and the KEGG database. The R package igraph (24) was used to identify and plot modules and networks, and the R package circlize (3) was used to plot connections of genes and enriched GO terms.

Venom RP-HPLC. Reversed-phase high-performance liquid chromatography (RP-HPLC) was performed on venom samples for the eight individuals shown in Fig. 1a using the Shimadzu Prominence HPLC system. Briefly, ∼15 μg of protein was injected for each venom sample onto an Aeris 3.6 μm C18 column (Phenomenex, Torrance, CA) using a detection wavelength of 220 nm. A standard solvent system of A (0.1% trifluoroacetic acid (TFA) in water) and B (0.06% TFA in acetonitrile) was used. All samples were run at a flow rate of 0.2 mL/min over a 125-minute gradient. The gradient started at 10% B for five minutes, increased from 10% to 55% B over 110 minutes, and again increased to 75% B over five minutes. The gradient was kept at 75% B for an additional five minutes before a 15-minute wash step using 10% B. Distinct peaks were identified and quantified using the manual peak integration tools in the Shimadzu LabSolutions software. Shannon's diversity index (H) was calculated using the vegan package (25) in R and the percent area under each peak as input. Diversity indices were then corrected to obtain the effective number of peaks as previously described (26). 95% confidence intervals were then calculated for the effective number of peak estimates for all samples except *C. tigris*. To determine whether *C. tigris* possessed a significantly simpler venom phenotype, we determined whether the effective number of peaks for *C. tigris* fell outside the 95% CI of the distribution.

Quantitative venom proteomics. We used quantitative mass spectrometry (qMS) to determine which toxins identified in the genome and transcriptome were present in the venom sample of the genome individual (CLP2741) as previously described (27). Briefly, 11 μg of venom and 150 μL of 100 mM Ammonium Bicarbonate were digested for twenty minutes. Thirty μL of 10 mM DTT

(Dithiothreitol) were then added and incubated for 1 hour at 60°C. After incubation, 30 μL of 50 mM IAA (Iodoacetoamine) were added; thirty minutes later, 150 μL of 50 mM Ammonium Bicarbonate were added. Trypsin (Promega V511 at 0.5 μg/2.5μL) diluted in 50 mM Ammonium Bicarbonate was added and incubated for 18 hours at 37°C. We then added 1% TFA (Trifluoroacetic acid) at 5% volume of the solution to stop digestion. Following digestion, the venom was lyophilized and resuspended in 0.1% formic acid to 250 ng/μL. We used three digested *E. coli* proteins (Abcam) as standards: 25 fmol of Beta-lactamase ampC (P00811), 250 fmol of Protein deglycase 1 (P31658), and 2500 fmol of Chaperone protein FimC (P31697). For qMS, 2 μL of digested venom were run on a Thermo Q Exactive HF high-resolution electrospray tandem mass spectrometer (Thermo Scientific) with an UltiMate3000 RSLCnano System (Thermo Scientific) in triplicate at room temperature. We used an Acclaim PepMap RSLC 75 μM 15 cm nanoviper column (Thermo Scientific) and a flow rate of 300 nl/min to separate the protein fragments. We used a one hour linear gradient from 3% to 45% B [solvent A: 99.9% H_2O (EMD Omni Solvent) and 0.1% formic acid; solvent B: 99.9% acetonitrile and 0.1% formic acid]. We operated the detector in data-dependent mode with the Thermo Excalibur 3.1.66 data system software (Thermo Scientific) and nanosprayed the effluent into the spectrometer. We scanned from 350-1700 m/z at a resolution of 60,000 in profile mode and targeted ions with a value of 10^5 during high energy collisional dissociation.

We compared the qMS results to all toxin and non-toxin protein-coding sequences identified during genome annotation using Proteome Discover 2.2 software (Thermo Scientific) and SequestHT as the search engine. We used the following parameters: enzyme name = Trypsin, precursor mass tolerance = 10 ppm, minimum peptide length = 6, maximum peptide length = 144, maximum δCn = 0.05, dynamic modifications, carbamidomethyl +57.021 Da(C), oxidation +15.995 Da(M), maximum missed cleavage = 2, and fragment mass tolerance = 0.2 Da. We confirmed the identified proteins in Scaffold v. 4.10.0 (Proteome Software Inc., Portland, OR, USA) with a false discovery rate of 1.0%. We considered a protein verified if it passed quality checks and was detected in one of the three replicates. We excluded a cysteine-rich secretory protein (CRISP-1) due to low quality assignment of peptides to the sequence. Proteins grouped by Scaffold due to shared peptide evidence were treated individually rather than as a cluster to allow us to directly assign the peptide to the corresponding transcript.

We compared molar estimates of toxin proteins in the *C. tigris* venom to StringTie TPM estimates from the combined venom-gland transcriptome (i.e., right and left venom-gland transcriptomes combined) to visualize the correlation between toxin transcript expression and abundance in the proteome. For each replicate, we calculated separate conversion factors for each of the three control proteins by calculating the slope of the best fit line of the known concentration and the observed normalized spectral counts (as determined by Scaffold), with an intercept at the origin. We then used those conversion factors to convert normalized spectral counts for each toxin protein in each replicate to concentrations, averaging final concentrations across the three replicates. Prior to comparing estimates, we applied the centered log-ratio (clr) transformation to all abundances as previously described (27). Equivalent to a log transformation for linear relationships, the clr transformation preserves rank and was used here to compare relative quantities of components.

SI Results.

Estimating phenotypic complexity. To determine whether *Crotalus tigris* possessed a significantly simpler venom phenotype than other rattlesnakes, we first estimated Shannon's diversity index and then converted this metric into effective number of peaks (see Methods for details). The Tiger Rattlesnake venom possessed the fewest effective peaks (5.15; Table S1). Excluding the Tiger Rattlesnake, the mean effective number of peaks was 15.80 (SD=3.21, 95% CI=9.50–22.10; Table S1). Because the Tiger Rattlesnake estimate fell outside the 95% CI, we concluded that the Tiger Rattlesnake venom was significantly less complex than other rattlesnake species.

Genomic organization of toxin genes. To determine the chromosomal location of the 51 putative toxin genes in the Tiger Rattlesnake genome, we analyzed the mapping results of the scaffolded assembly and compared the locations of Tiger Rattlesnake toxin genes to the locations of orthologous toxin genes in the Prairie Rattlesnake (Fig. 1b; Table S3). Most major toxin gene families shared similar chromosomal locations across species with the largest toxin gene families being found on microchromosomes, consistent with previous results in the Prairie Rattlesnake (Table S3). For example, the SVMP gene family ($n = 5$ paralogs in *C. tigris*) was located on microchromosome 1 (confidence score = 0.98; scores range 0–1), the PLA_2 gene family ($n = 3$) was located on microchromosome 7 (confidence score = 0.56), and the snake venom serine proteinase (SVSP; $n = 15$) gene family was located on microchromosome 2 (confidence score = 0.62) in both species. Not all toxin genes, however, were similarly located across species. C-type lectins (CTL; $n = 5$), for example, were located on microchromosome 5 in the Prairie Rattlesnake, but the Tiger Rattlesnake CTL scaffold mapped to unassigned scaffold 47 (i.e., chromosomal location unknown). Vascular endothelial growth factor (VEGF), a highly expressed toxin in the Tiger Rattlesnake (Fig. S3), mapped to microchromosome 3 (confidence score = 0.68) whereas VEGF in the Prairie Rattlesnake was located on macrochromosome 1. Similarly, nucleotidase mapped with high confidence to macrochromosome 1 in the Tiger Rattlesnake (confidence score = 0.94) but was proposed to be located on microchromosome 1 in the Prairie Rattlesnake. All data are shown in Fig. 1b and Table S3.

1. Krueger F, Andrews S (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27(11):1571–1572.
2. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359.
3. Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize implements and enhances circular visualization in R. *Bioinformatics* 30(19):2811–2812.
4. Rotenberg D, Bamberger ES, Kochva E (1971) Studies on ribonucleic acid synthesis in the venom glands of *Vipera palaestinae* (Ophidia, Reptilia). *Biochem. J.* 121:609–612.
5. Rokyta DR, Lemmon AR, Margres MJ, Aronow K (2012) The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* 13:312.
6. Margres M, Aronow K, Loyacano J, Rokyta D (2013) The venom-gland transcriptome of the eastern coral snake (*Micrurus fulvius*) reveals high venom complexity in the intragenomic evolution of venoms. *BMC Genomics* 14:531.
7. Grabherr M, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29(7):644–652.

8. Holding M, Margres M, Mason A, Parkinson C, Rokyta D (2018) Evaluating the performance of *de novo* assembly methods for venom-gland transcriptomics. *Toxins* 10(6):249.
9. Buenrostro J, Giresi P, Zaba L, Chang H, Greenleaf W (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10(12):1213.
10. Buenrostro J, Wu B, Chang H, Greenleaf W (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protocc Mol. Biol.* 109(1):21–29.
11. Krueger F (2015) Trim galore. a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.
12. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo B (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33(4):574–576.
13. Genrich. https://github.com/jsh58/Genrich.git.
14. MACS2. https://github.com/taoliu/MACS.git.
15. Picard. https://github.com/broadinstitute/picard.git.
16. Love M, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.
17. Schield D, et al. (2019) The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res.* 29(4):590–601.
18. Aird S, et al. (2017) Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol. Evol.* 9(10):2640–2649.
19. Ullate-Agote A, Milinkovitch M, Tzika A (2015) The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. *Int. J. Dev. Biol.* 58(10-11-12):881–888.
20. Li B, Dewey C (2011) RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323.
21. R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*.
22. Soneson C, Love M, Robinson M (2015) Differential analyses for RNA-seq: transcript level estimates improve gene-level inferences. *F1000Research* 4.
23. Ge S, Jung D, Yao R (2019) ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* pp. 1–2.
24. Csardi G, Nepusz T (2006) The igraph software package for complex network search.
25. Oksanen J, et al. (2007) The vegan package. *Community Ecology Package* pp. 631–637.
26. Jost L (2006) Entropy and diversity. *Oikos* 113(2):363–375.
27. Rokyta D, Margres M, Calvin K (2015) Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. *G3: Genes, Genom. Genet.* g3:115.

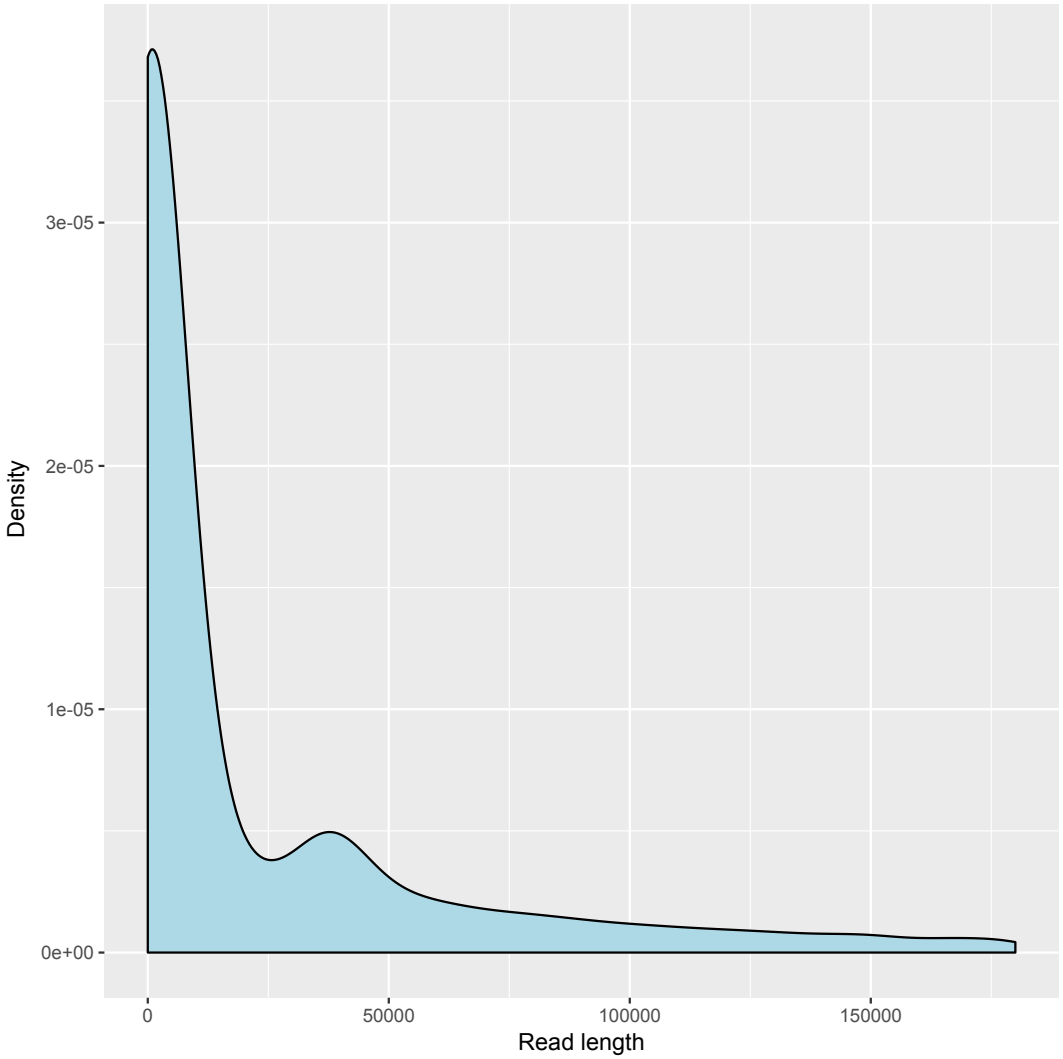

Fig. S1. Fragment size distribution of the PacBio reads. Read length in bp is shown on the x-axis. Frequency is on the y-axis.

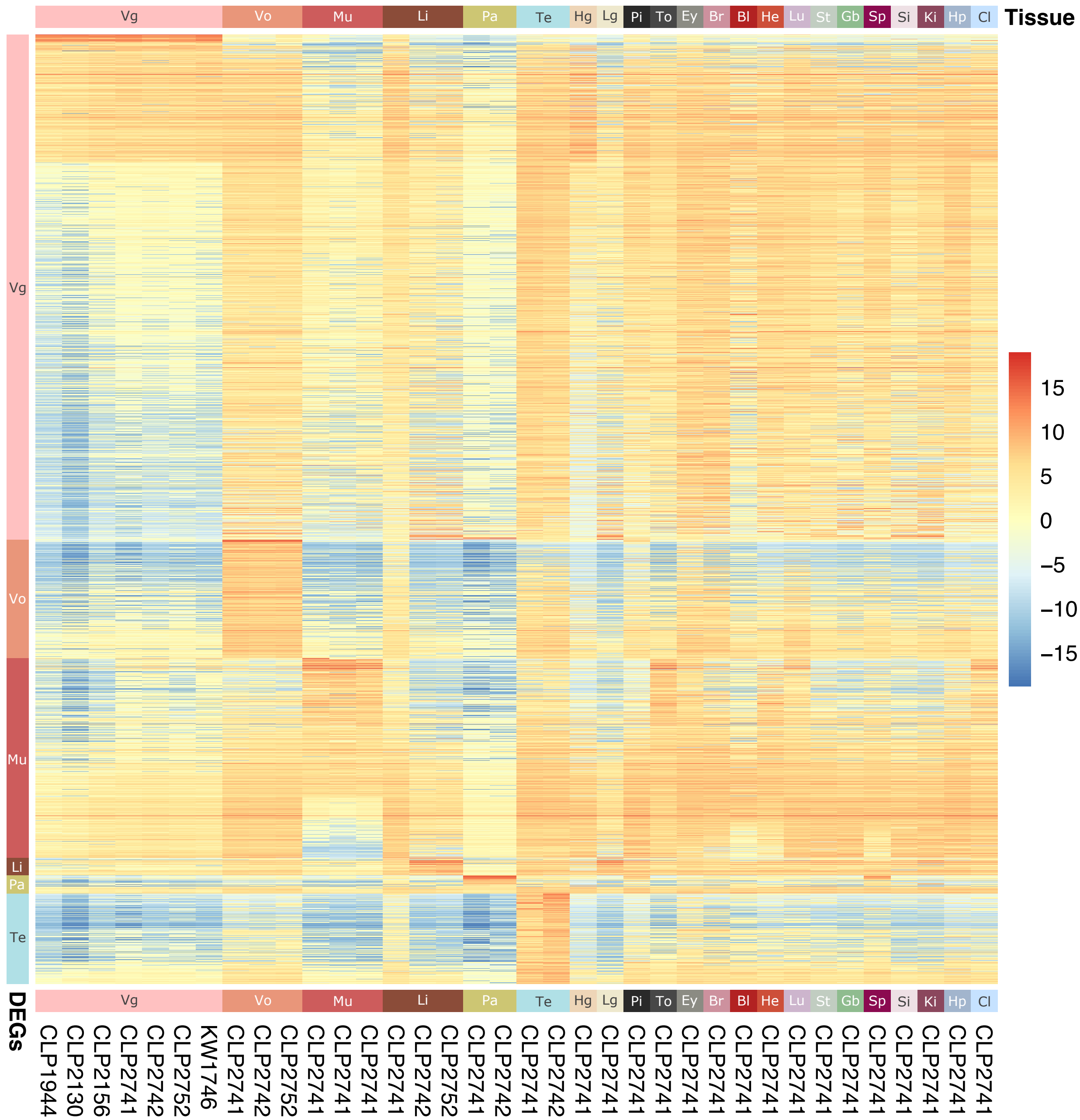

Fig. S2. Significant differential expression across 22 tissue types in the Tiger Rattlesnake. Heatmap of significantly differentially expressed genes (DEGs; $n = 3,551$) across all sequenced tissues. Statistical comparisons were made by comparing tissues with replicates (as identified on the y-axis) against all other tissues. DEGs were identified as genes with $\log_2$(counts per million reads mapped) >5 for at least two replicates, FDR $\leq 1\%$, and log-fold change ≥ 2. We identified 1,889, 442, 747, 66, 68, and 339 DEGs for Vg, Vo, Mu, Li, Pa, and Te, respectively. Warmer colors indicate higher expression. Individual snake IDs are shown on the x-axis within each tissue group; CLP2741 is the genome animal. Vg for CLP2741, CLP2742, and CLP2752 represent combined left and right venom-gland transcriptomes; the other four individuals had both glands sequenced in a single transcriptome. Abbreviations: Vg, venom glands; Vo, vomeronasal organ; Mu, muscle (venom-gland compressor muscle, body muscle, tail muscle); Li, liver; Pa, pancreas; Te, testes; Hg, harderian gland; Lg, labial gland; Pi, pit; To, tongue; Ey, eye; Br, brain; Bl, blood; He, heart; Lu, lung; St, stomach; Gb, gall bladder; Sp, spleen; Ki, kidney; Hp, hemipenis; Cl, cloaca.

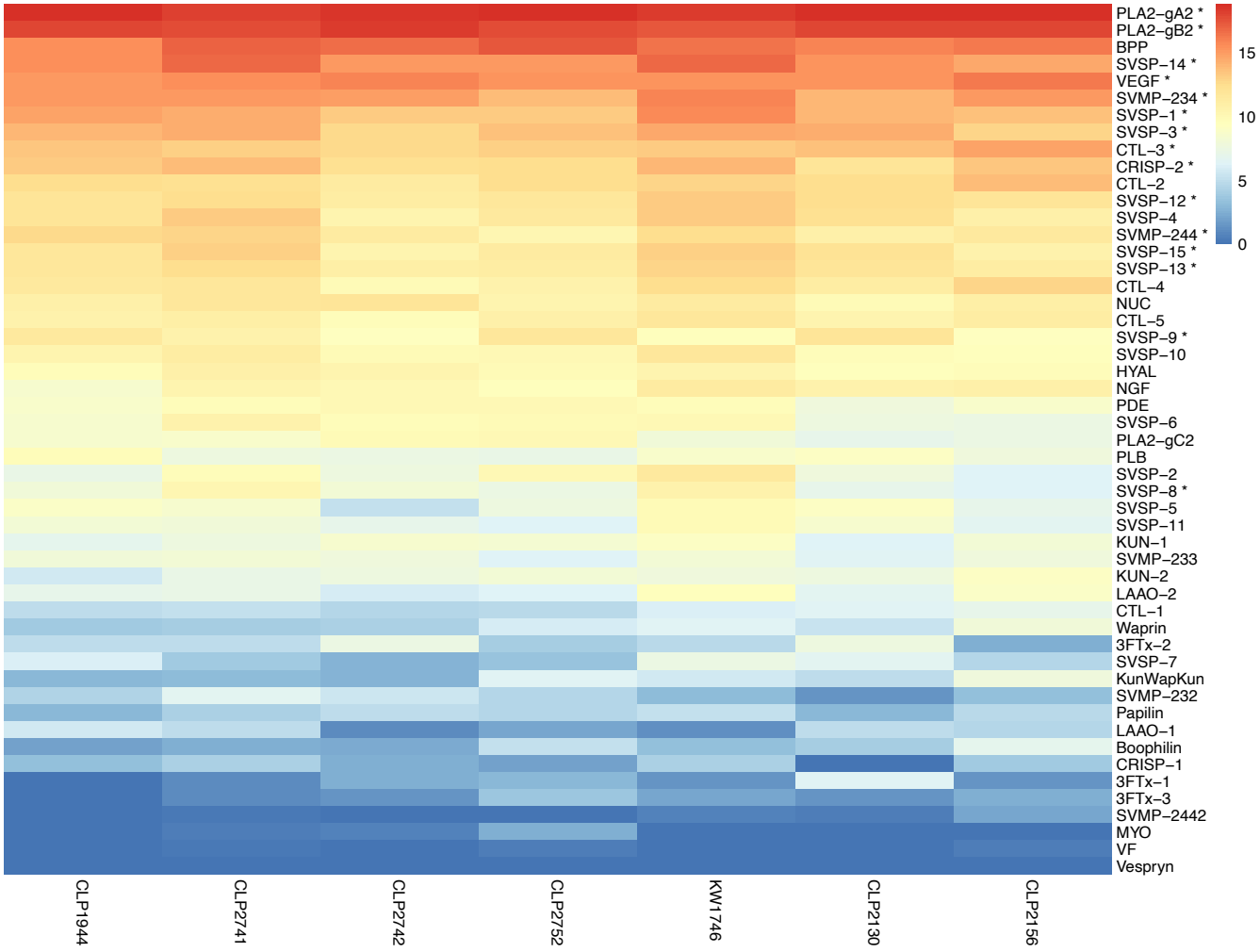

Fig. S3. Expression of the 51 putative toxin genes in the venom-gland transcriptomes of seven Tiger Rattlesnakes. Heatmap of toxins is ordered from greatest to least average expression using $\log_2$(counts per million reads mapped). Warmer colors indicate higher expression. Asterisks indicate the 15 toxins that were detected proteomically in the venom of the genome animal (CLP2741). Individual snake IDs are shown on the x-axis. For CLP2741, CLP2742, and CLP2752, left and right venom-gland transcriptomes were combined to generate the plot; the other four individuals had both glands sequenced in a single transcriptome. Abbreviations: BPP, Bradykinin-potentiating peptide; CRISP, Cysteine-rich secretory protein; CTL, C-type lectin; HYAL, Hyaluronidase; KUN, Kunitz-type toxin; KunWapKun, Kunitz-Waprin-Kunitz fused toxin; LAAO, L-Amino acid oxidase; MYO, Myotoxin; NGF, Nerve growth factor; NUC, Nucleotidase; PDE, Phosphodiesterase; PLA_2, Phospholipase A_2; PLB, Phospholipase B; SVMP, Snake venom metalloproteinase; SVSP, Snake venom serine proteinase; 3FTx, Three-finger toxin; VEGF, Vascular endothelial growth factor; VF, Venom factor.

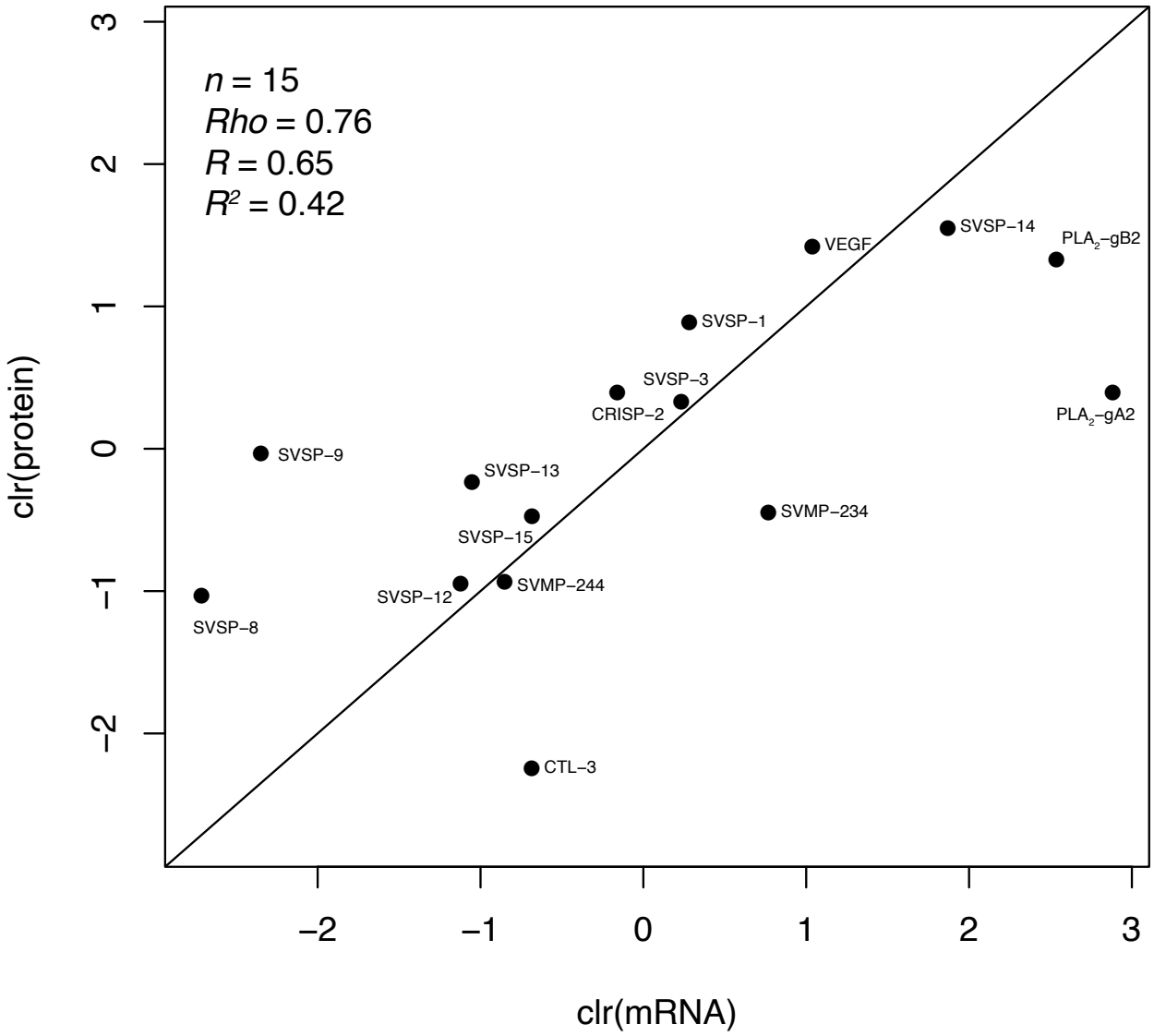

Fig. S4. Protein and mRNA abundances were correlated between venom proteome and venom-gland transcriptome in the Tiger Rattlesnake genome animal. We compared StringTie estimates of transcripts per million (TPM) from the venom-gland transcriptome (left and right venom glands combined; x-axis) to molar estimates of protein abundances in the venom (y-axis) for the 15 toxins identified in the venom proteome of the genome individual. All data were centered log-ratio (clr) transformed. Abbreviations: CRISP, Cysteine-rich secretory protein; CTL, C-type lectin; PLA_2, phospholipase A_2; SVMP, snake venom metalloproteinase; SVSP, snake venom serine proteinase; VEGF, vascular endothelial growth factor.

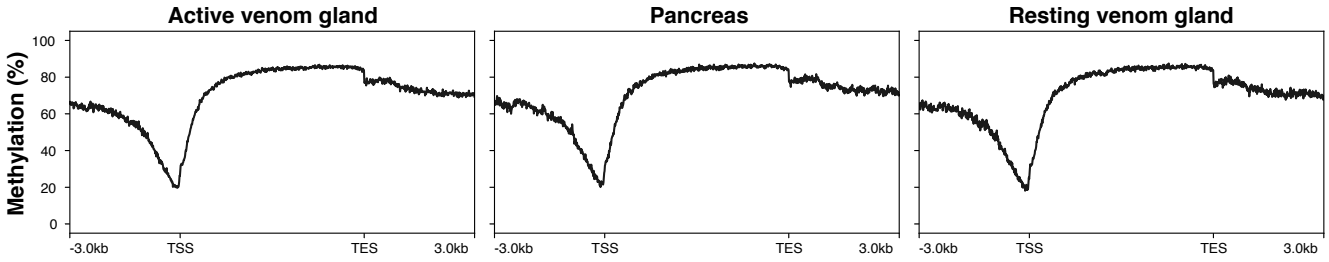

Fig. S5. **Reductions in methylation at TSS across all non-venom genes.** Methylation percentage is shown on the y-axis. Genomic position ± 3kb relative to the transcription start site (TSS) and transcription end site (TES) is shown on the x-axis. Black lines represent mean methylation level for all 18,189 non-venom genes annotated in the genome.

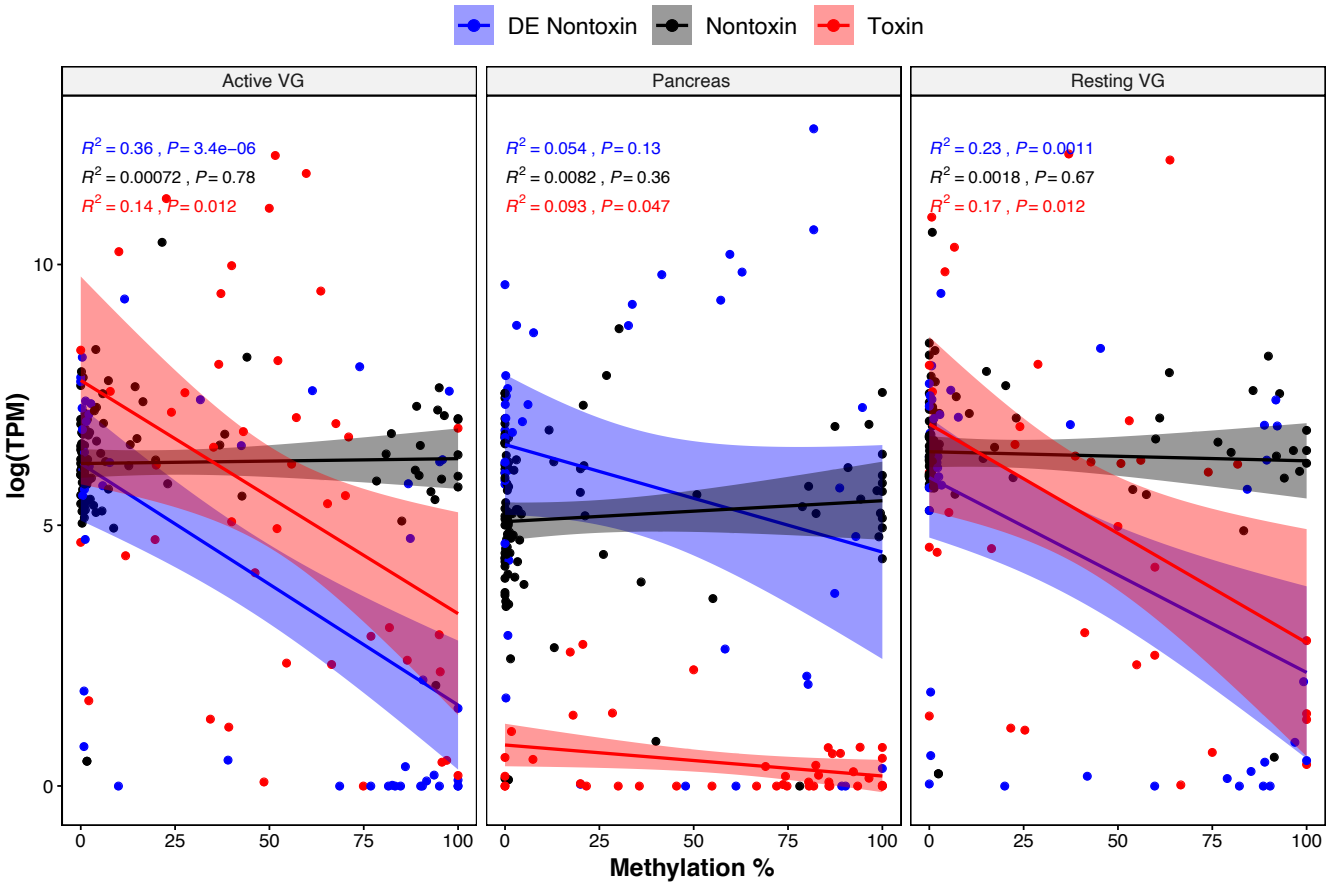

Fig. S6. Expression and methylation levels were significantly correlated. Each point represents an individual transcript. Blue points represent nontoxins that were significantly differentially expressed (DE) across venom glands (VGs) and pancreas (mean TPM >300; Padj <0.01). Black points represent nontoxins that were not significantly DE across VGs and pancreas (mean TPM >300; Padj >0.1). Red points represent toxins. Expression level is on the y-axis. Methylation percentage is on the x-axis.

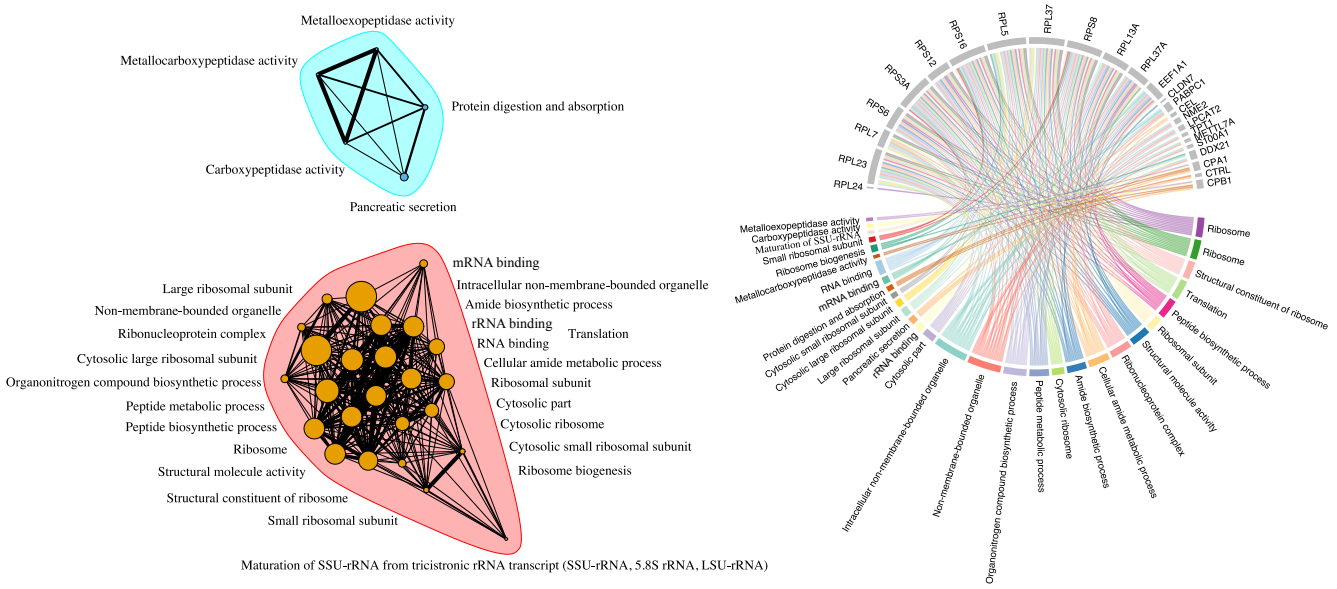

Fig. S7. Gene ontology (GO) term enrichment for nontoxins that were significantly differentially expressed across venom glands and pancreas. Left: Network generated by ShinyGO of all significantly enriched GO terms for differentially expressed nontoxin genes (TPM >300; Padj <0.01). Two nodes are connected if they share greater than 20% of their genes. Node size corresponds to size of the gene sets and edge thickness correspond to the number of overlapped genes. Right: Chord diagram connecting GO terms to their associated genes.

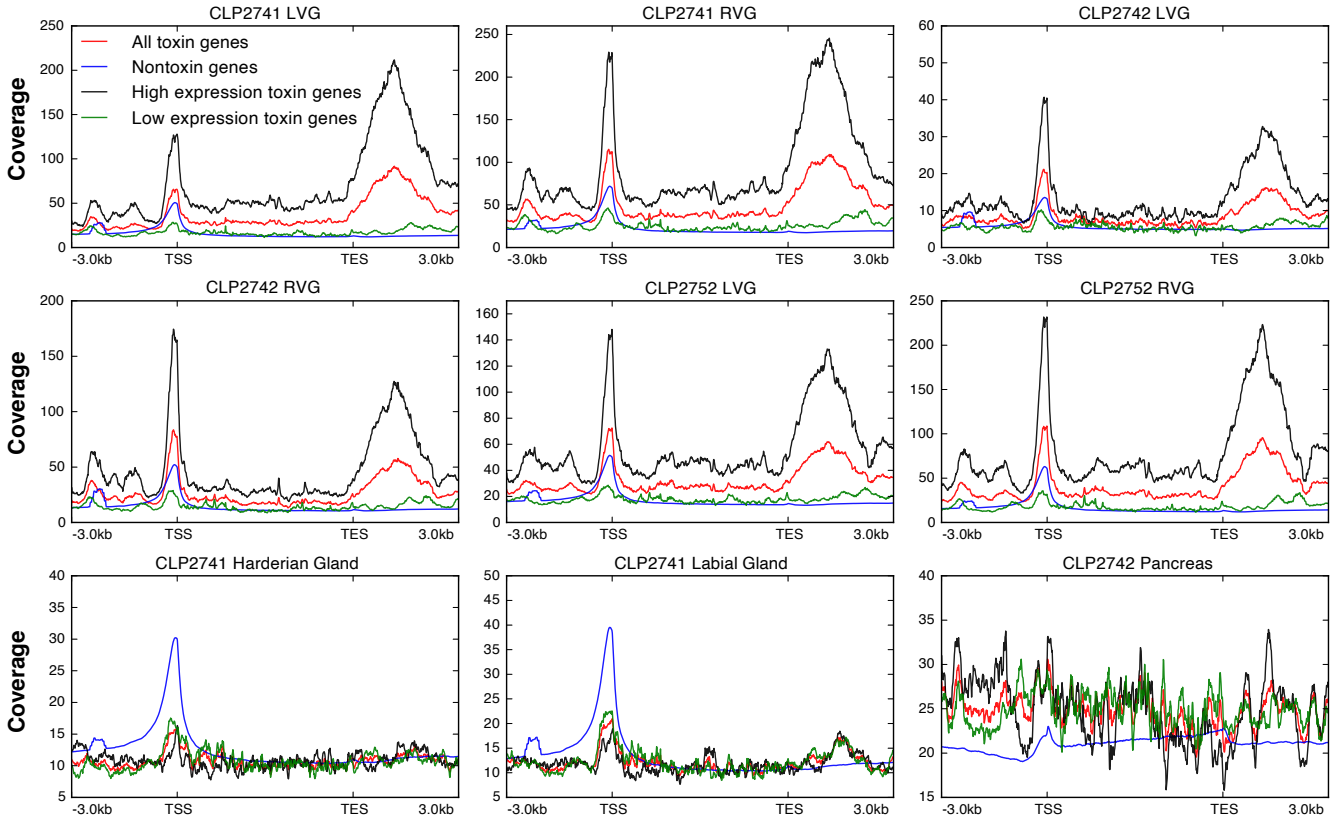

Fig. S8. ATAC-seq coverage near TSS per tissue. ATAC-seq coverage is shown on the y-axis. Genomic position ± 3kb relative to the transcription start site (TSS) and transcription end site (TES) is shown on the x-axis. Left (LVG) and right (RVG) venom glands for three individuals are shown as well as three control tissues: Harderian gland, labial gland, and pancreas. Lines represent mean ATAC-seq coverage for all 51 toxin genes (red), all 18,189 non-venom genes annotated in the genome (blue), the 20 most highly expressed (TPM >1,000) toxin genes (black), and the 32 most lowly expressed (TPM <1,000) toxin genes (green). Abbreviations: TPM, transcript per million.

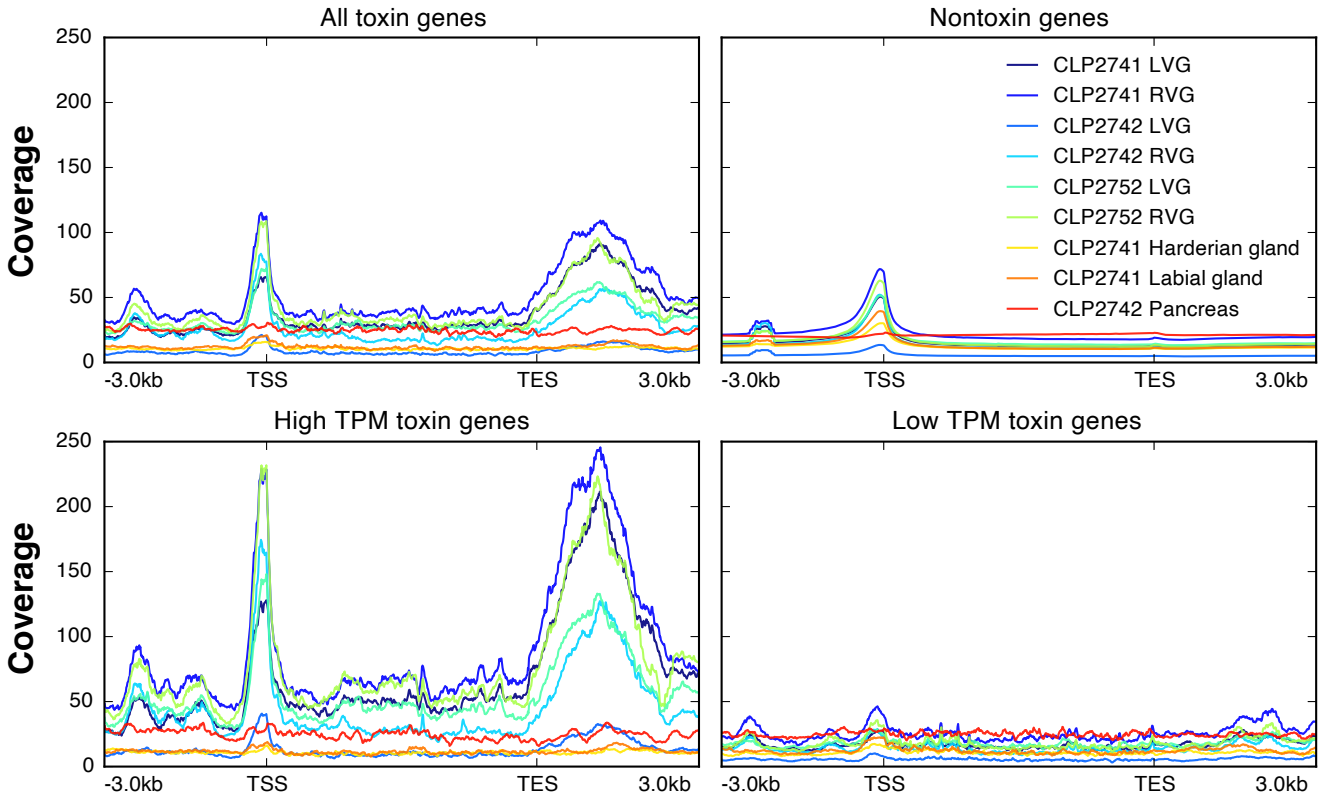

Fig. S9. ATAC-seq coverage near TSS across all genes. ATAC-seq coverage is shown on the y-axis. Genomic position ± 3kb relative to the transcription start site (TSS) and transcription end site (TES) is shown on the x-axis. Left (LVG) and right (RVG) venom glands for three individuals are shown as well as three control tissues: Harderian gland, labial gland, and pancreas. Lines represent mean ATAC-seq coverage for each sample for all 51 toxin genes (top left), all 18,189 non-venom genes annotated in the genome (top right), the 20 most highly expressed (TPM >1,000) toxin genes (bottom right), and the 32 most lowly expressed (TPM <1,000) toxin genes (bottom left). Abbreviations: TPM, transcript per million.

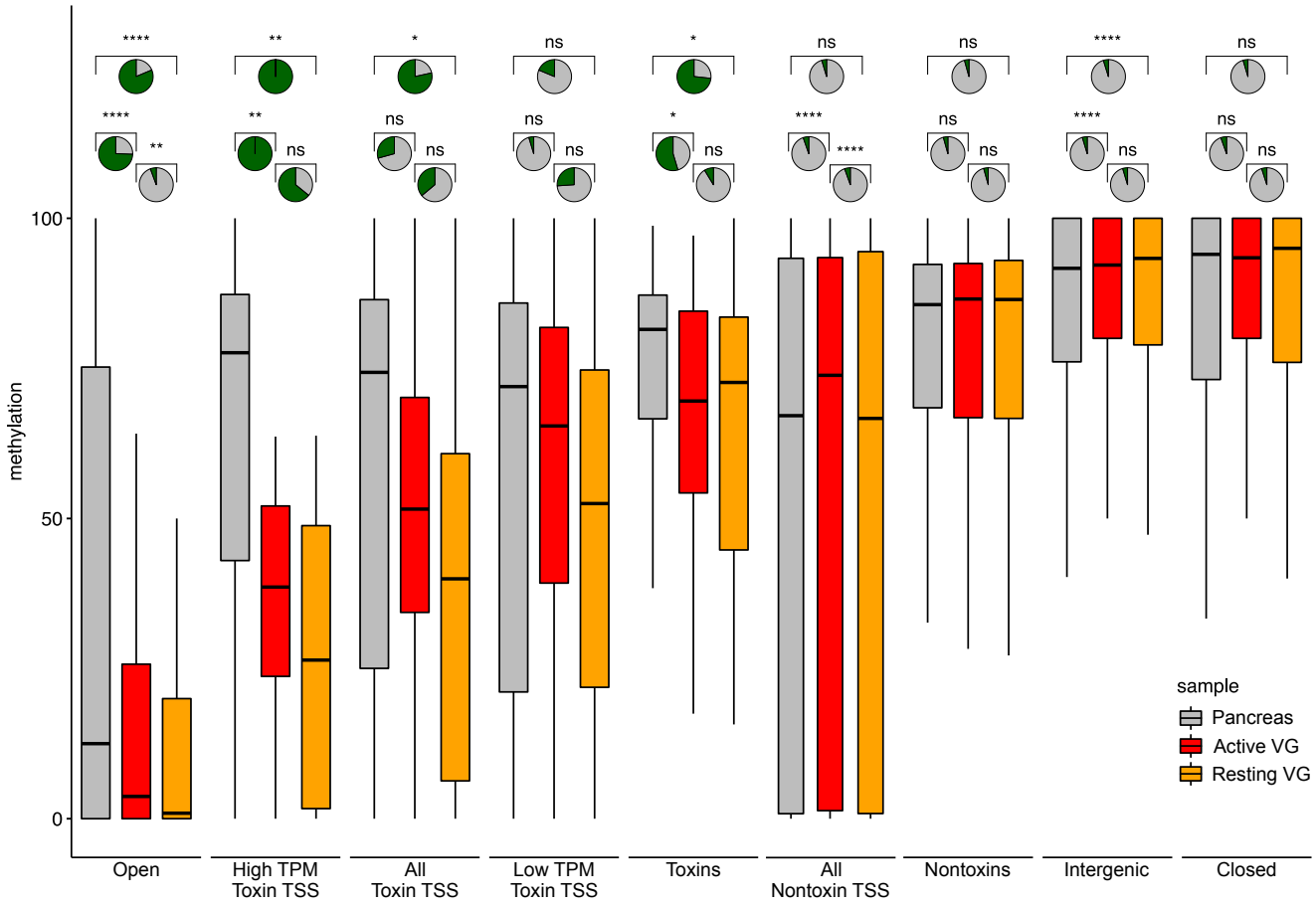

Fig. S10. Distributions of methylation levels for pancreas, active venom glands (VG), and resting VGs. Genomic regions are from Figure 3 and are as follows: chromatin accessible regions (Open), high-expression toxin gene transcription start sites (High TPM Toxin TSS), all toxin gene TSS, low-expression (Low TPM) toxin gene TSS, entire toxin genic regions including introns (Toxin), all nontoxin gene TSS, entire nontoxin genic regions including introns (Nontoxin), intergenic regions, and inaccessible chromatin regions (Closed). TSS regions included ± 500bp around the estimated start site. High TPM toxins represent the 20 most highly expressed toxins (TPM >1,000). Low TPM toxins represent the 32 transcripts with TPM <1,000. Nontoxin genes represent genes actively expressed in the VGs that do not produce toxic proteins. Statistical significance was assessed using t-tests. Pie charts represent the proportion of 10,000 bootstrapped t-tests that were significant (P <0.05) after subsampling; green represents significant tests and gray represents non-significant tests.

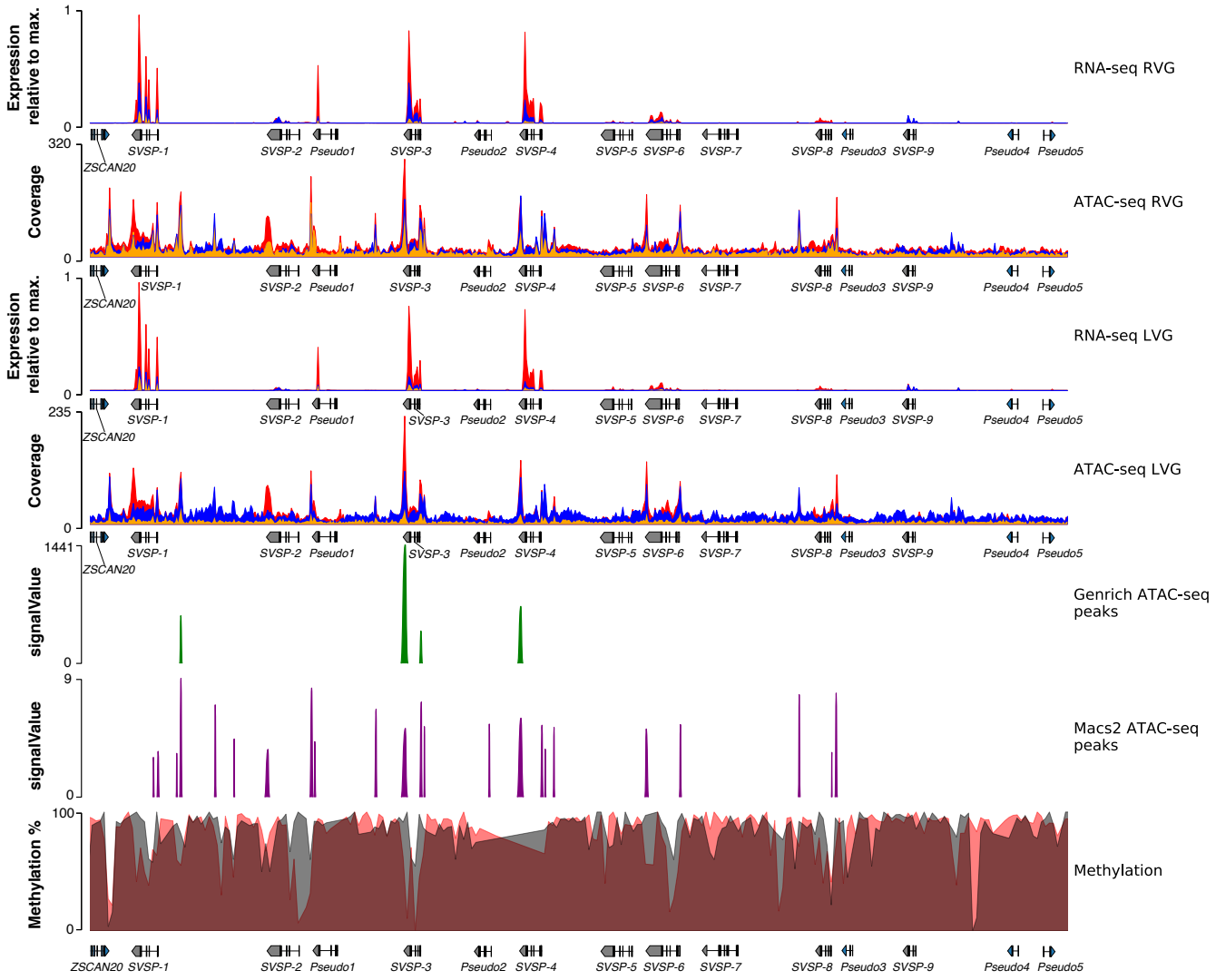

Fig. S11. Regulatory landscape for the SVSP toxin-gene family in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of SVSP toxin genes. Only part of the SVSP tandem array is shown here due to size constraints; remaining region is shown in Figure S2. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: SVSP, snake venom serine proteinase.

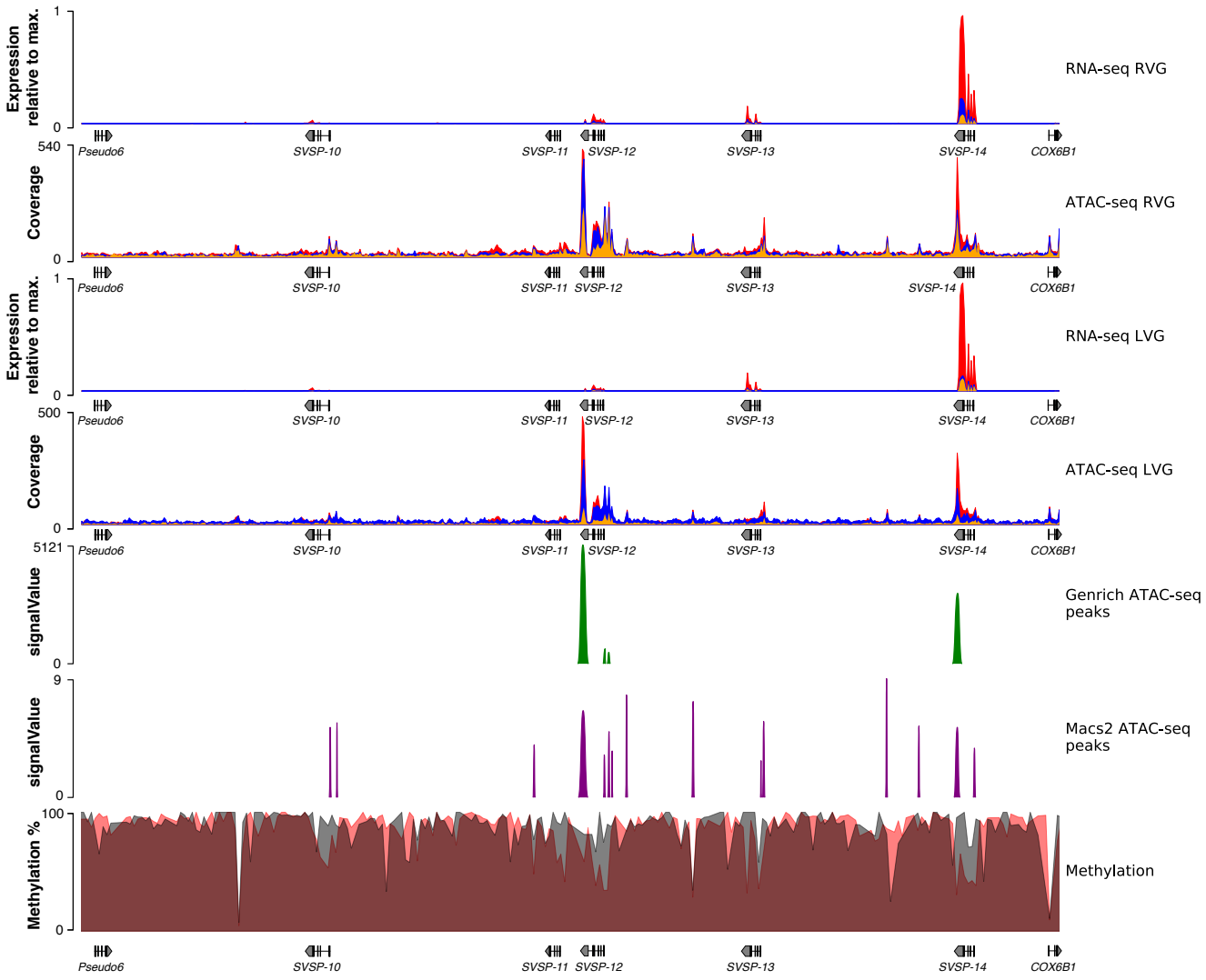

Fig. S12. Regulatory landscape for the SVSP toxin-gene family in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of SVSP toxin genes. Only part of the SVSP tandem array is shown here due to size constraints; remaining region is shown in Figure S1. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: SVSP, snake venom serine proteinase; Pseudo, pseudogenized SVSP.

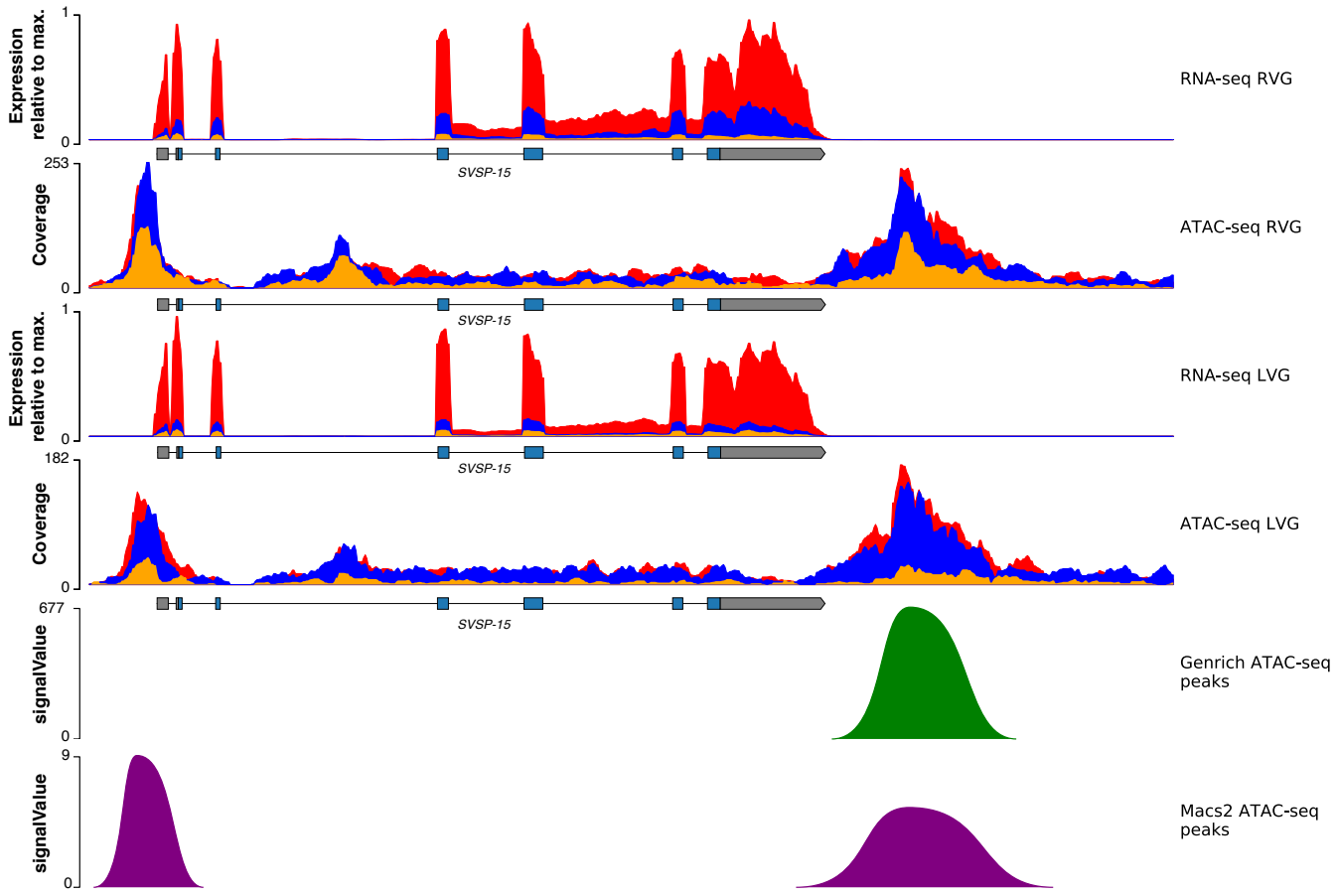

Fig. S13. Regulatory landscape for the SVSP-15 gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of the SVSP-15 toxin gene. SVSP-15 represents the only SVSP paralog not assembled onto the same scaffold as the large SVSP tandem array. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, and venom-gland specific ATAC-seq peaks identified by MACS2. No bisulfite reads mapped to this genomic region. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. Gene annotations are shown under tracks where appropriate. Abbreviations: SVSP, snake venom serine proteinase.

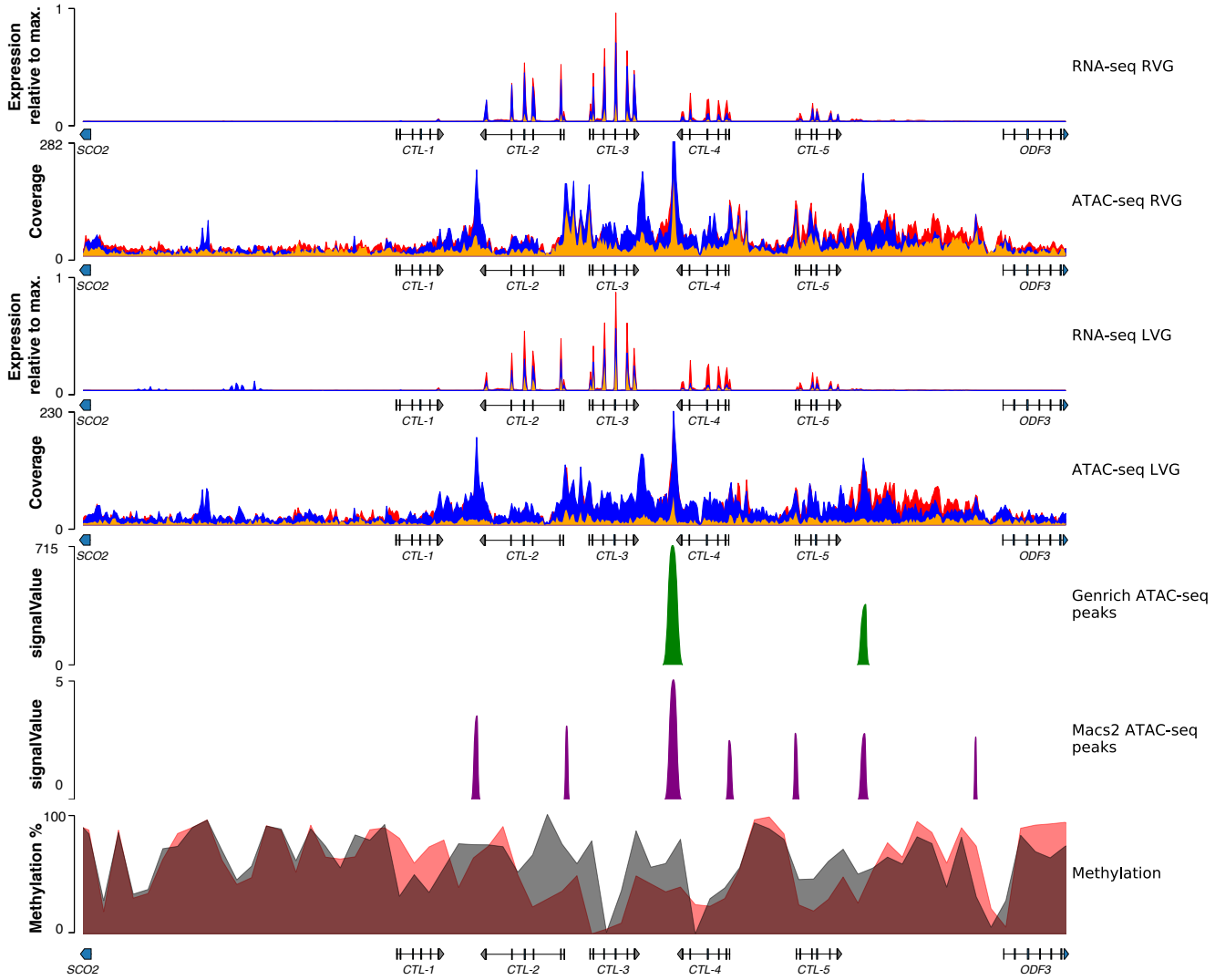

Fig. S14. Regulatory landscape for the CTL toxin-gene family in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of CTL toxin genes. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: CTL, C-type lectin.

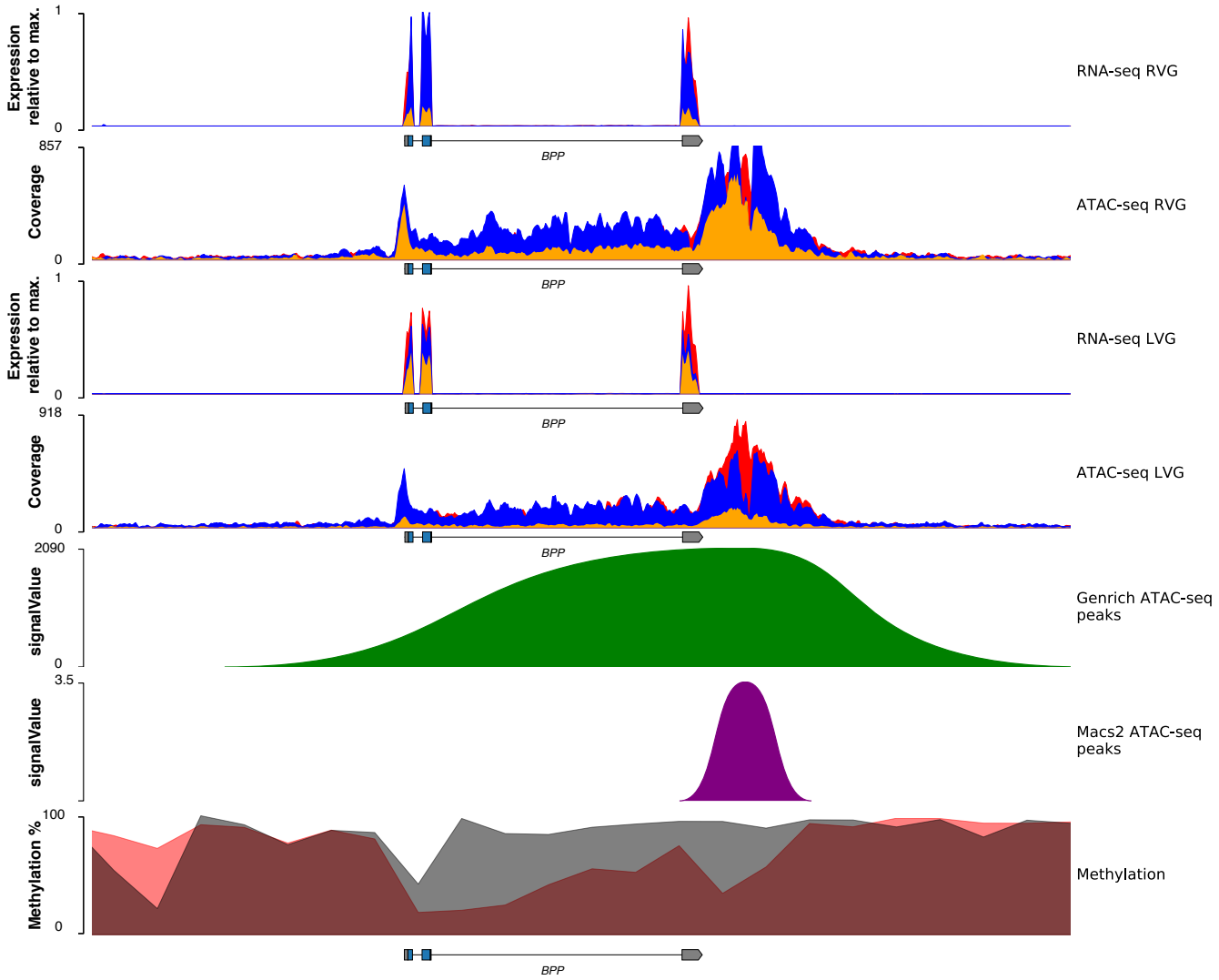

Fig. S15. Regulatory landscape for the BPP toxin-gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of BPP. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: BPP, bradykinin-potentiating peptide.

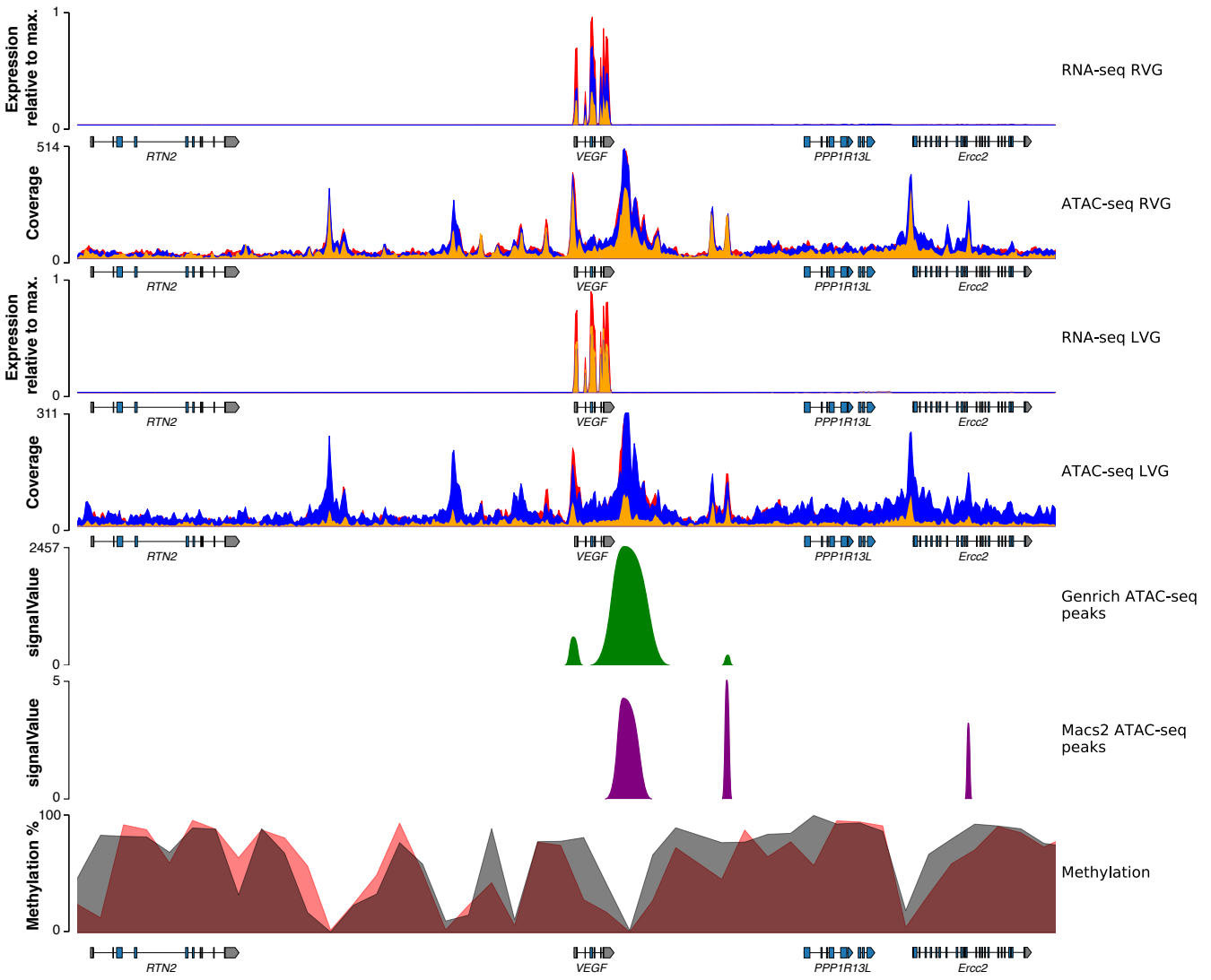

Fig. S16. Regulatory landscape for the VEGF toxin gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of VEGF. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: VEGF, vascular endothelial growth factor.

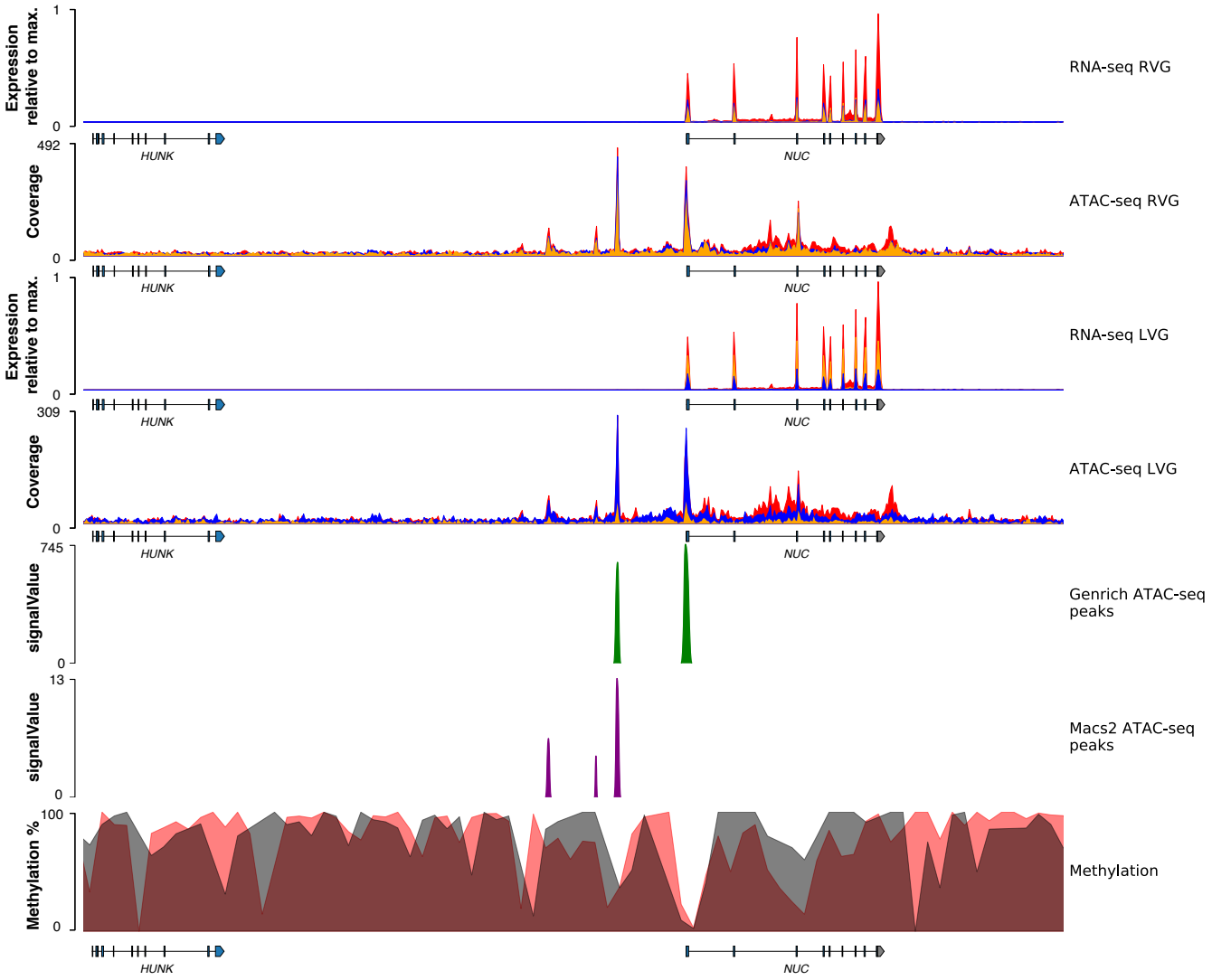

Fig. S17. Regulatory landscape for the NUC toxin gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of NUC. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: NUC, nucleotidase.

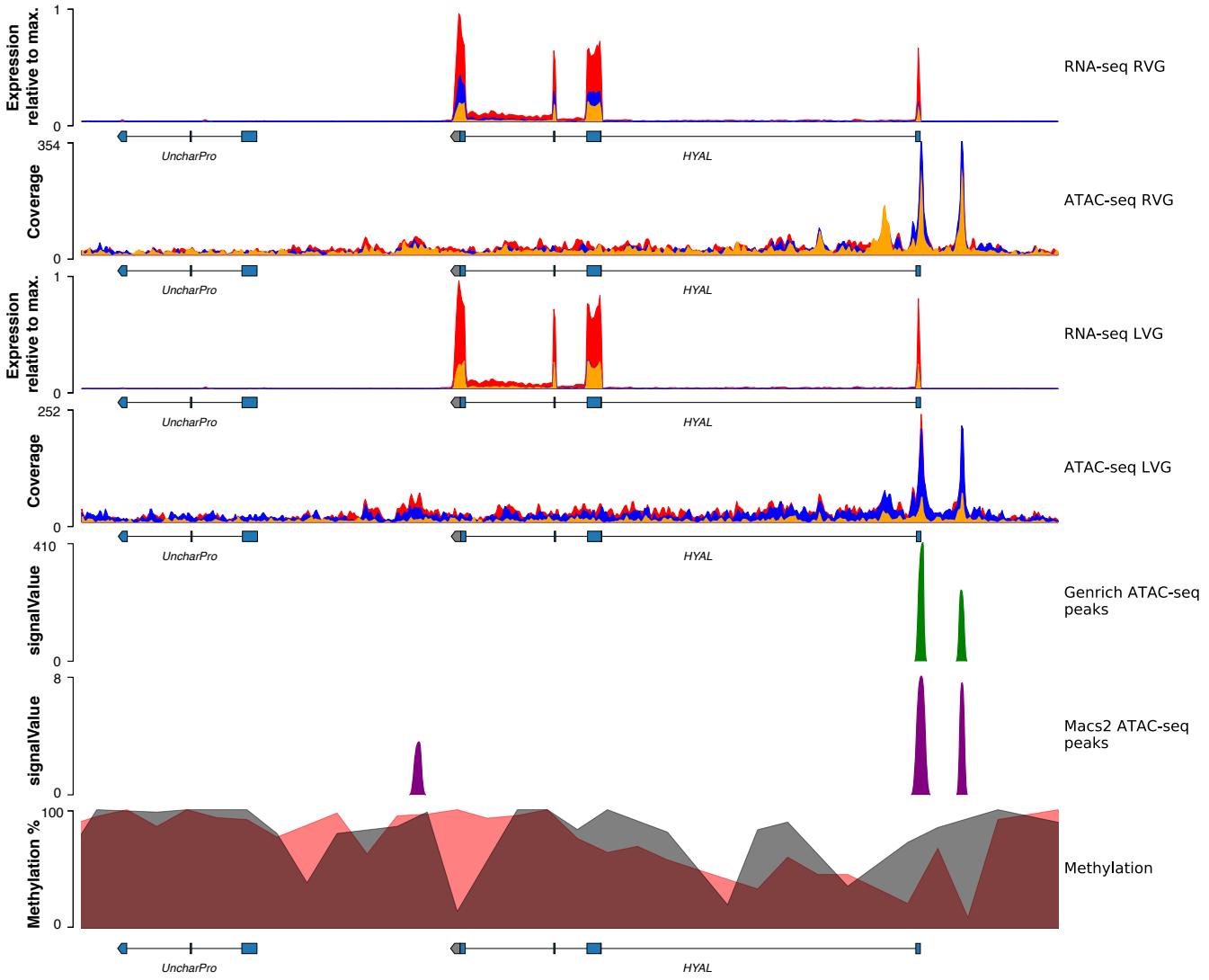

Fig. S18. Regulatory landscape for the HYAL toxin-gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of HYAL. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: HYAL, hyaluronidase.

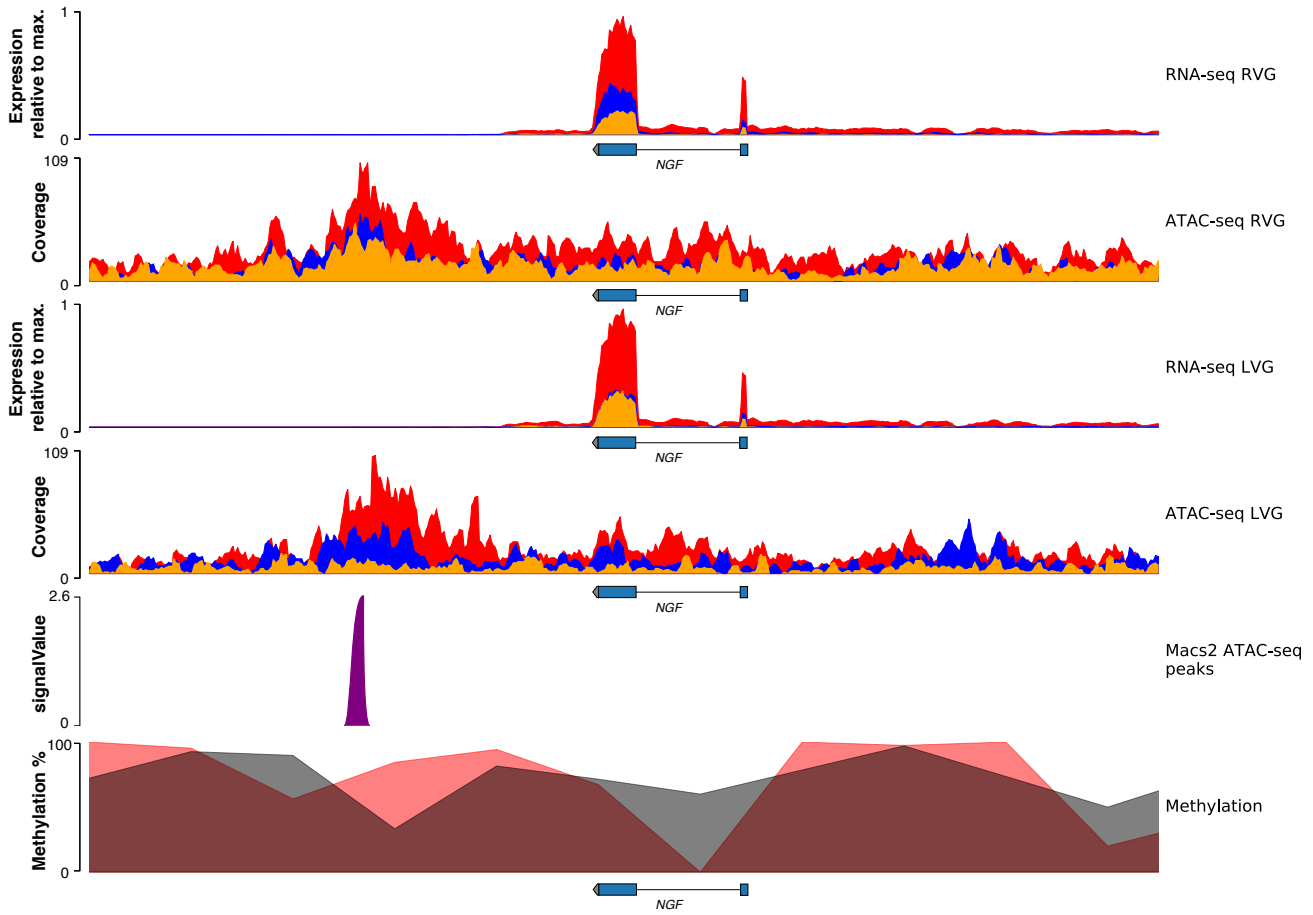

Fig. S19. Regulatory landscape for the NGF toxin gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of NGF. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage; no venom-gland specific Genrich peaks were called. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: NGF, nerve growth factor.

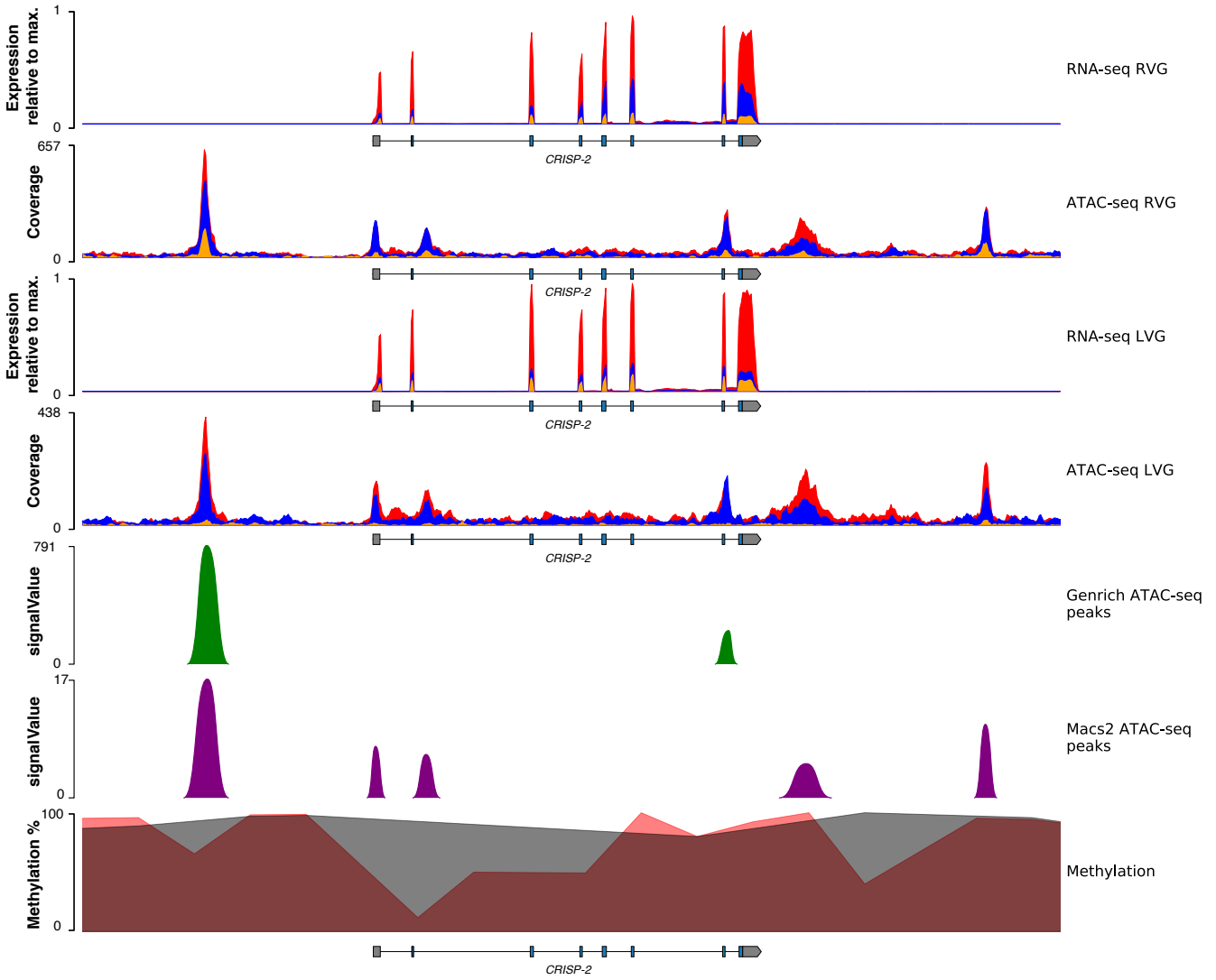

Fig. S20. Regulatory landscape for the CRISP-2 toxin-gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of CRISP-2. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: CRISP, cystine-rich secretory protein.

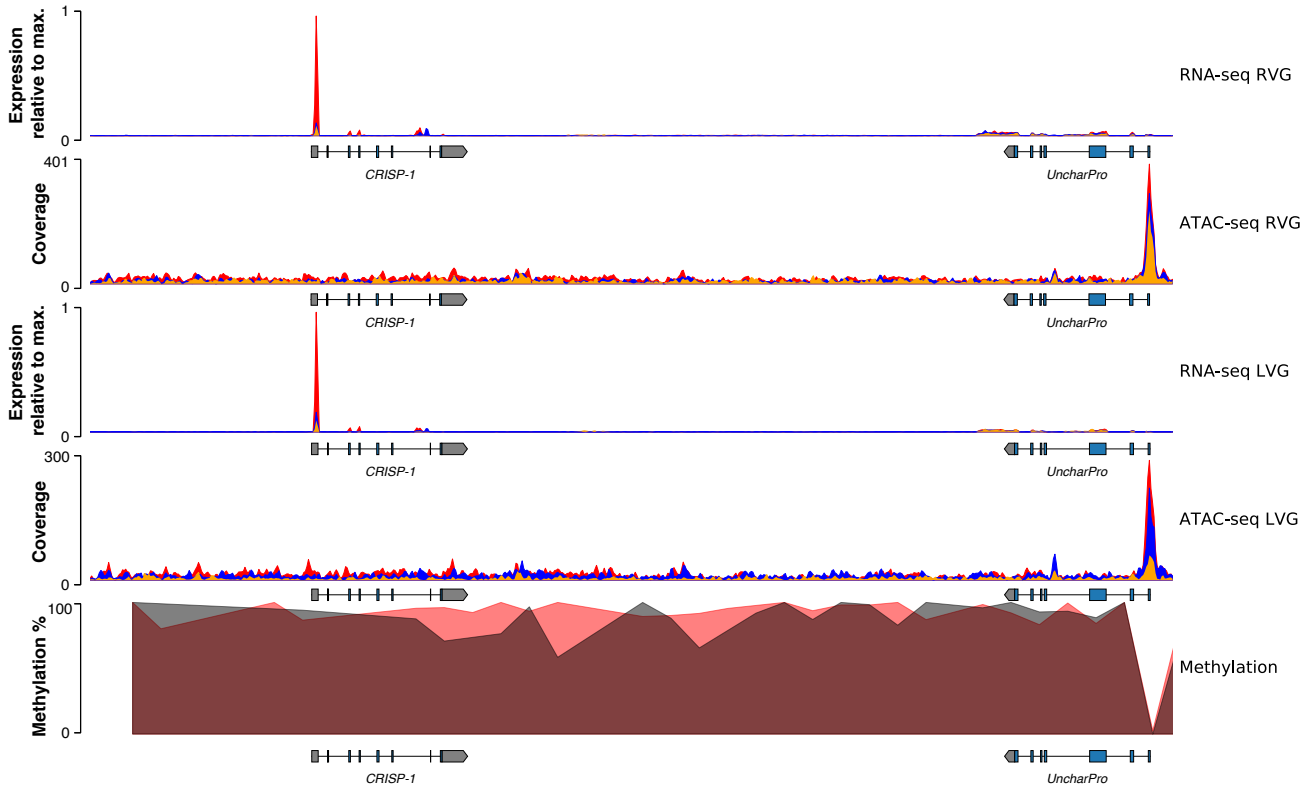

Fig. S21. Regulatory landscape for the CRISP-1 gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of CRISP-1. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: CRISP, cysteine-rich secretory protein; UncharPro, uncharacterized protein.

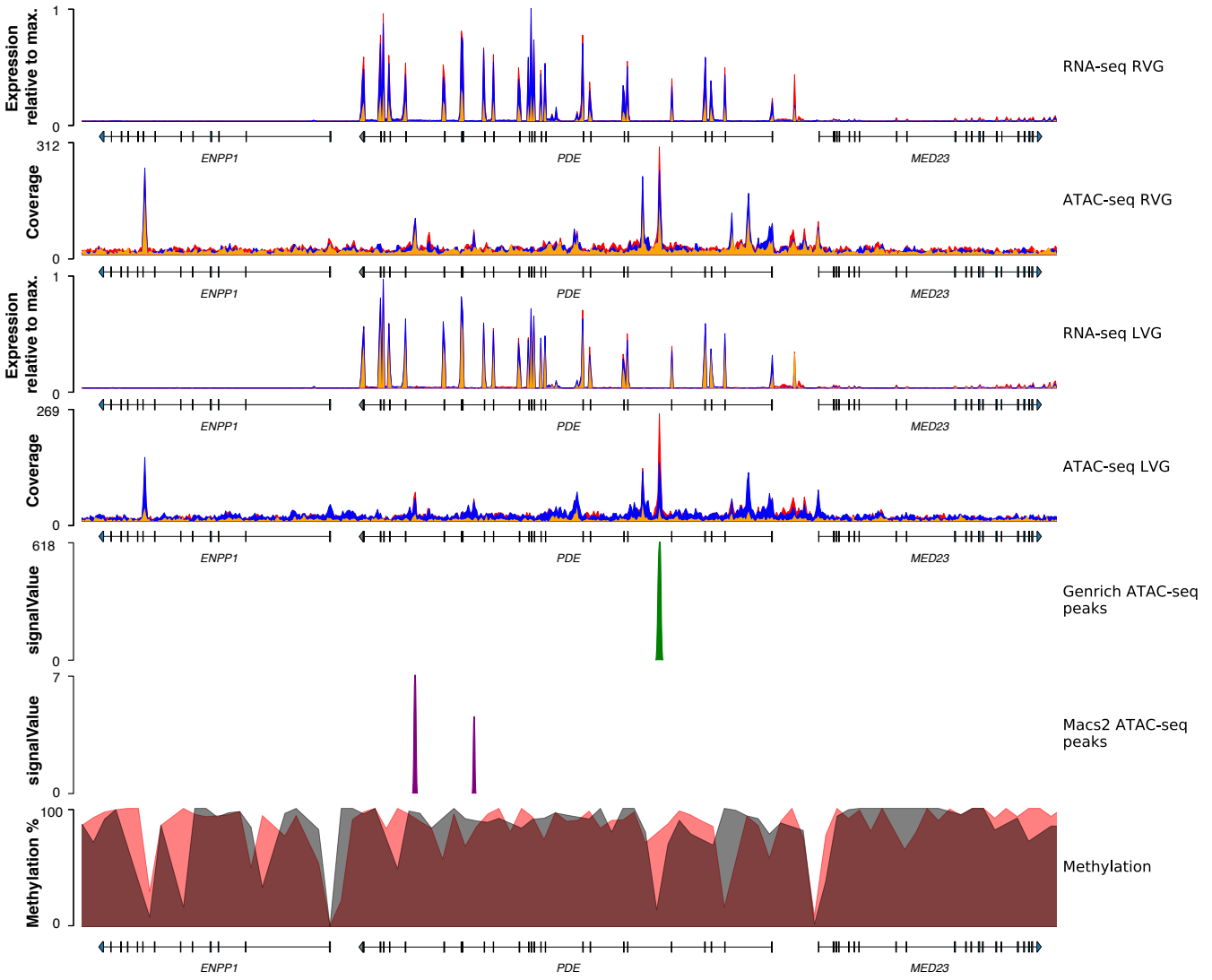

Fig. S22. Regulatory landscape for the PDE toxin gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of PDE. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: PDE, phosphodiesterase.

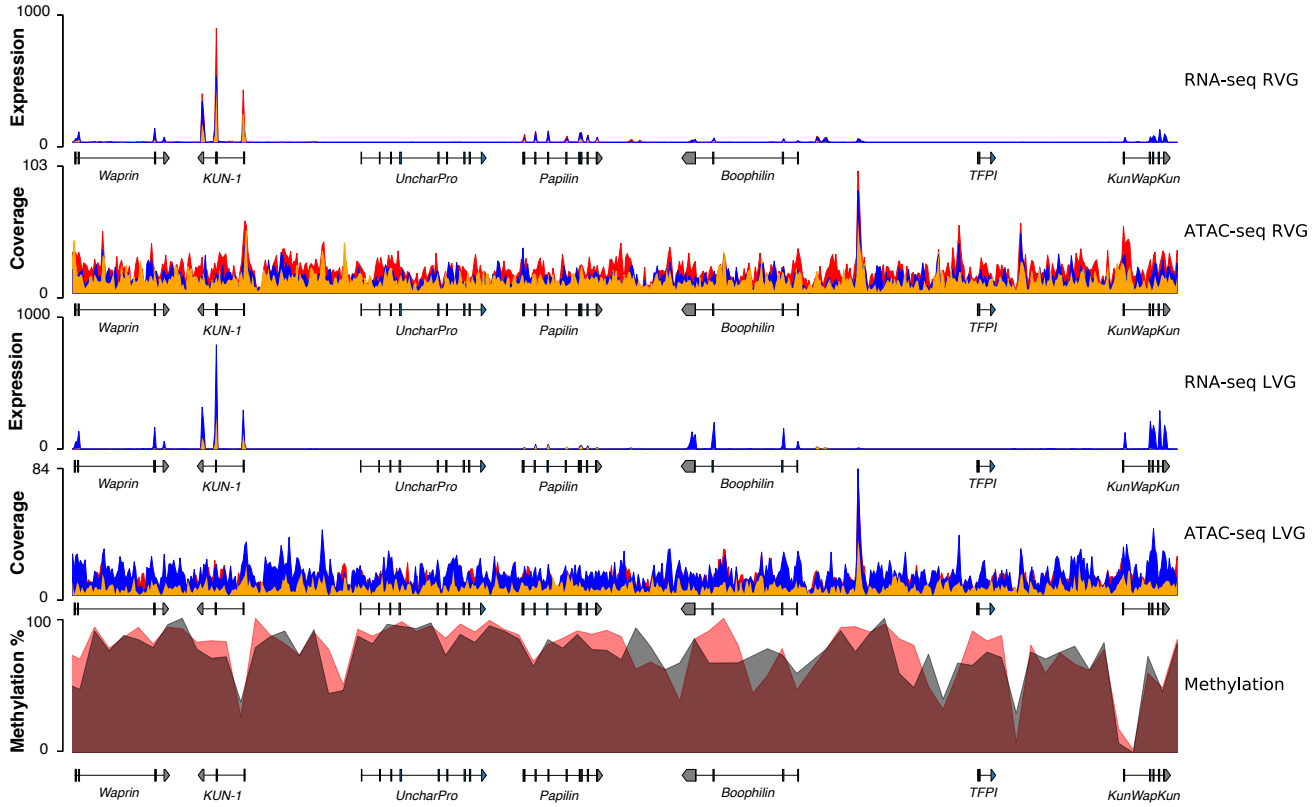

Fig. S23. Regulatory landscape for scf7180000017791 scaffold in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of several toxin genes on the scf7180000017791 scaffold. Putative toxin genes are as follows: Waprin, KUN-1, Papilin, Boophilin, and KunWapKun. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2. RNA-seq y-axes represent venom-gland expression levels (TPM). Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: KUN, Kunitz-type toxin; KunWapKun, Kunitz-Waprin-Kunitz fused toxin; UncharPro, uncharacterized protein.

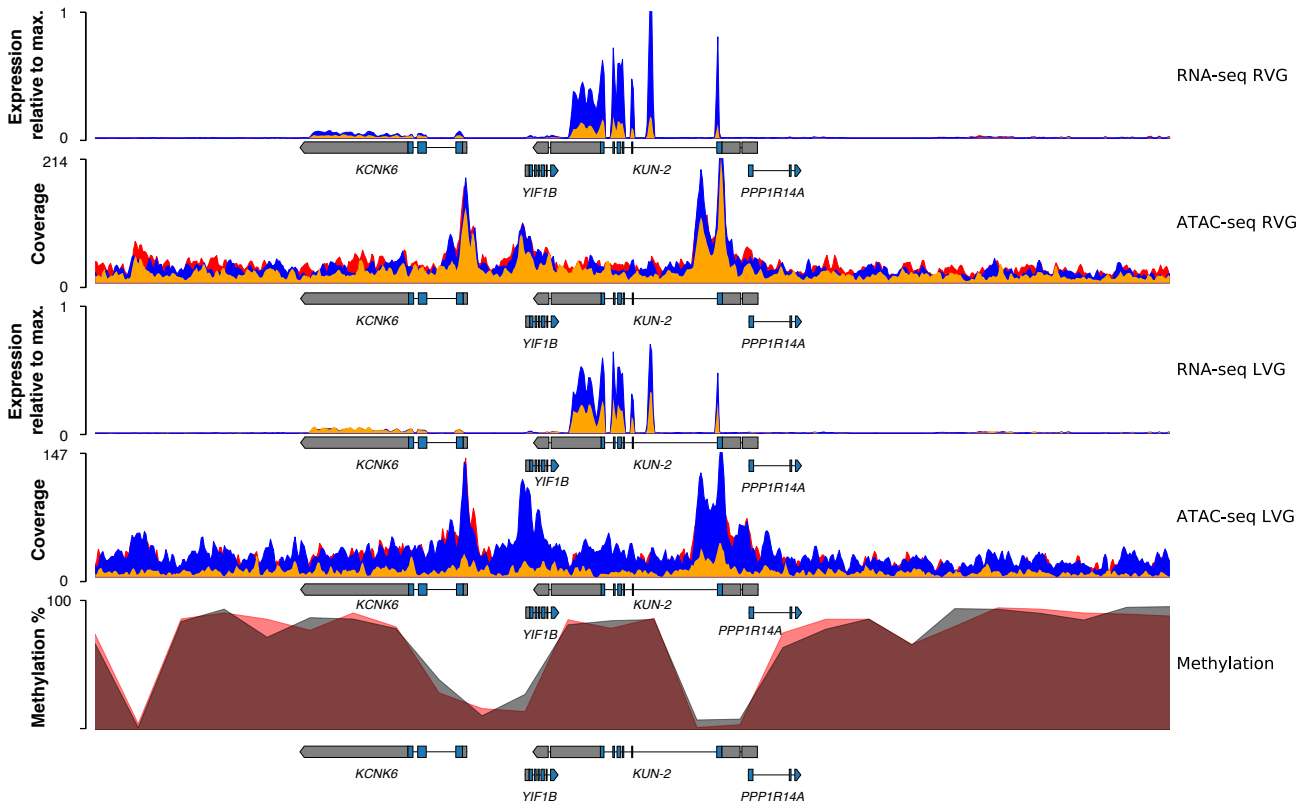

Fig. S24. Regulatory landscape for the KUN-2 gene in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of KUN-2. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2, suggesting the open chromatin regions shown here are accessible in several tissues. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: KUN, Kunitz-type toxin.

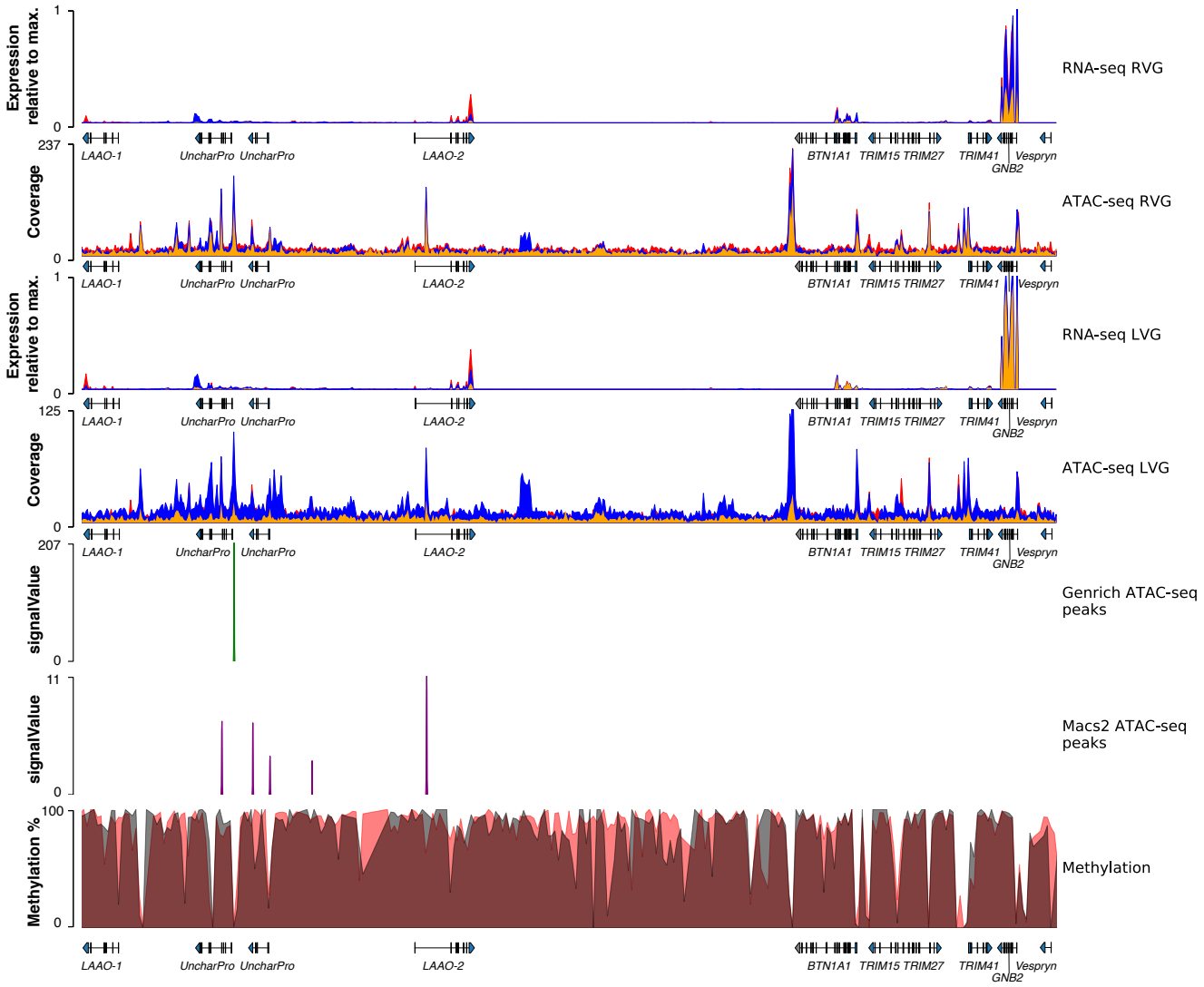

Fig. S25. Regulatory landscape for the LAAO toxin-gene family in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of LAAO and Vespryn. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, venom-gland specific ATAC-seq peaks identified by Genrich, venom-gland specific ATAC-seq peaks identified by MACS2, and methylation percentage. RNA-seq y-axes represent venom-gland expression levels scaled by the most highly expressed transcript in that genomic region. Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. Genrich-estimated signalValues represent the area under the peak; MACS2-estimated signalValues represent fold-change in coverage at the peak summit. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: LAAO, L-Amino acid oxidase; UncharPro, uncharacterized protein.

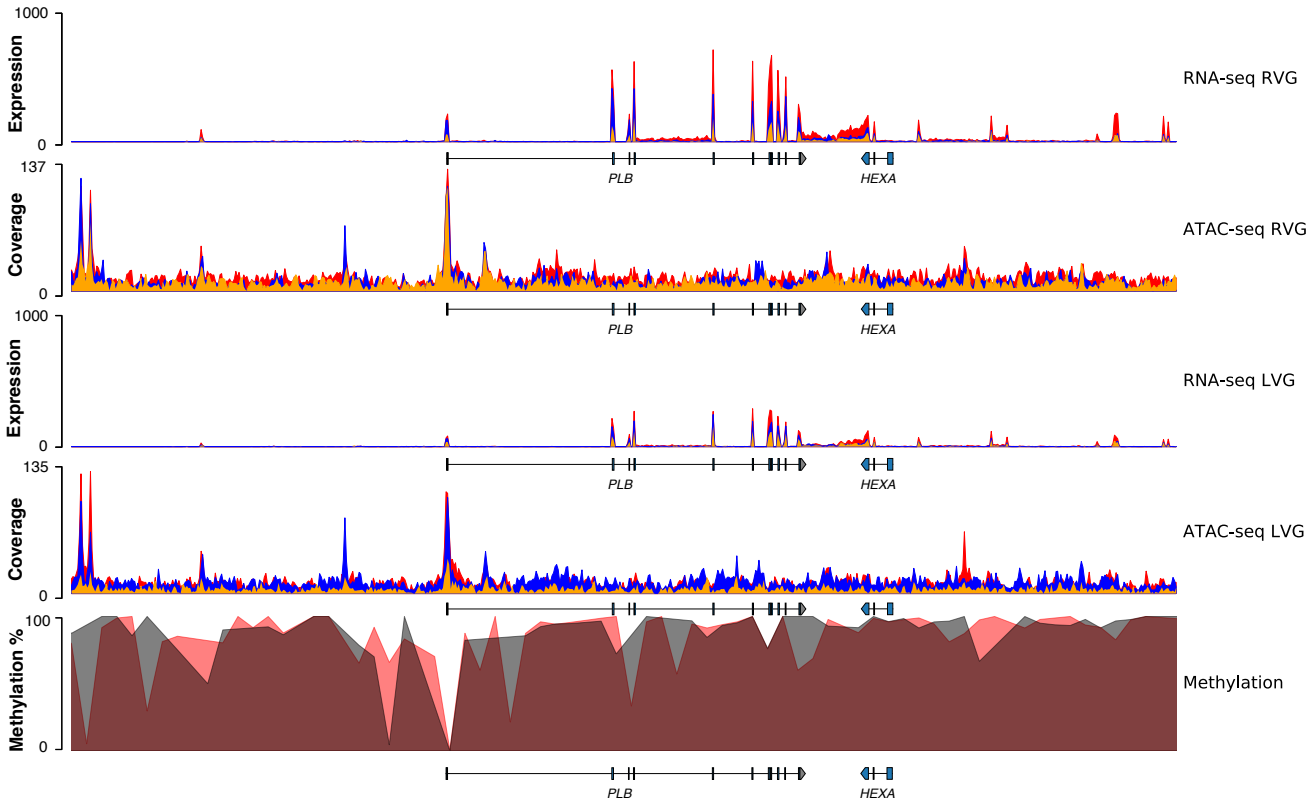

Fig. S26. Regulatory landscape for PLB in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of PLB. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2. RNA-seq y-axes represent venom-gland expression levels (TPM). Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: PLB, phospholipase B.

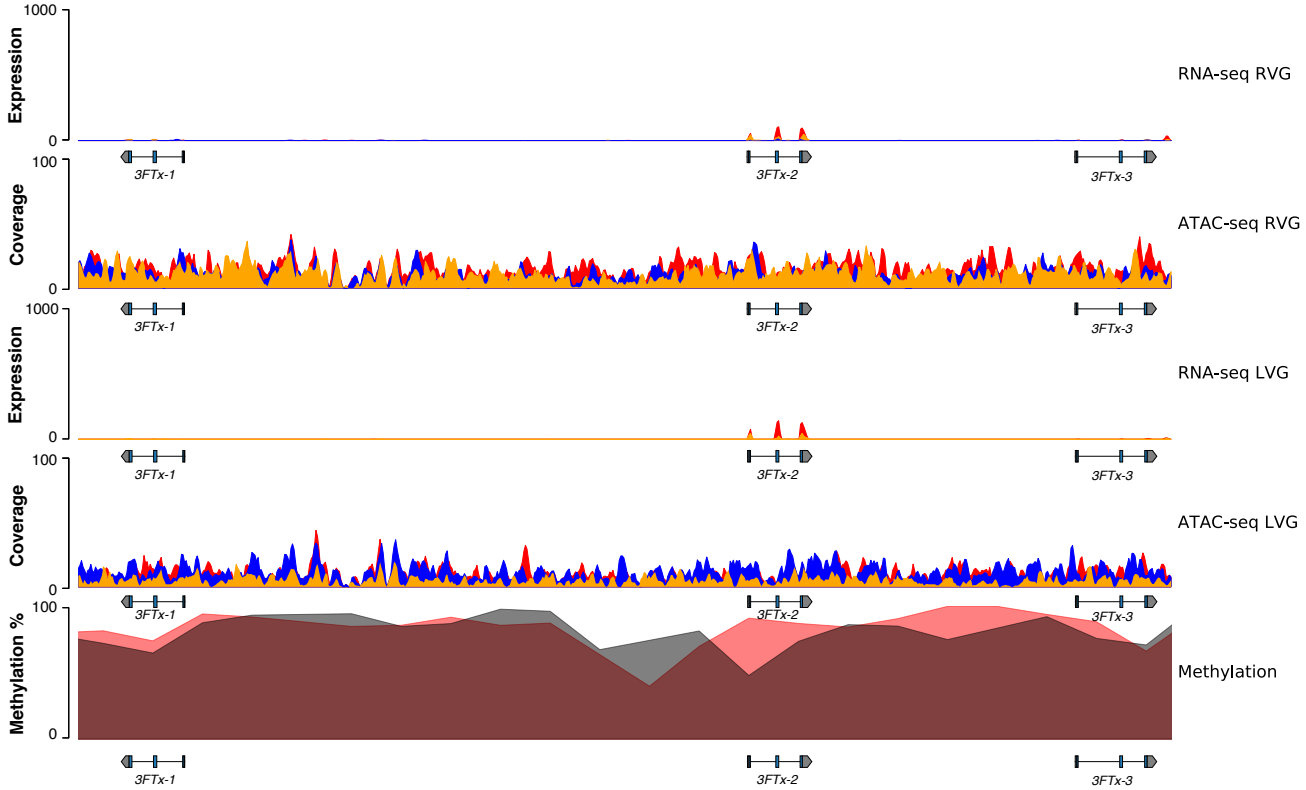

Fig. S27. Regulatory landscape for the 3FTx toxin-gene family in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of 3FTx toxin genes. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2. RNA-seq y-axes represent venom-gland expression levels (TPM). Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: 3FTx, three-finger toxin.

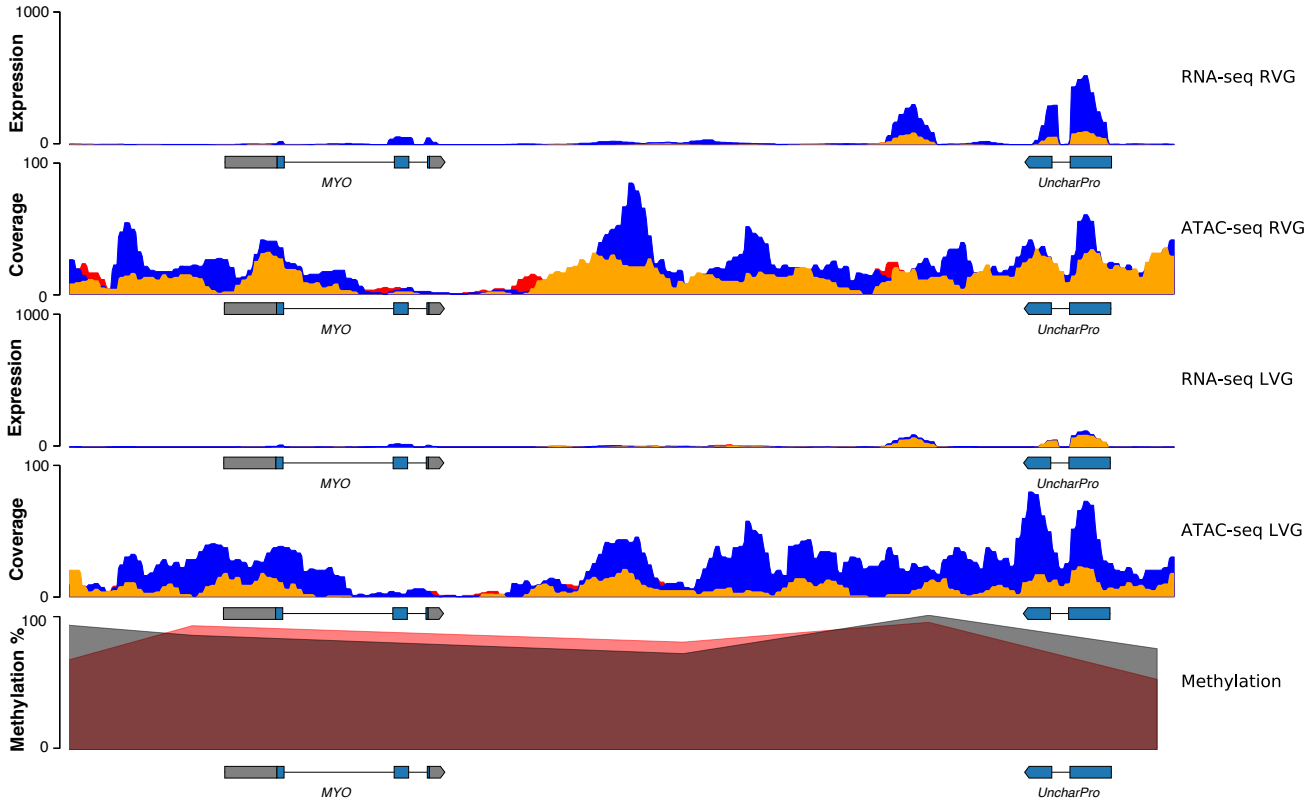

Fig. S28. Regulatory landscape for myotoxin in the Tiger Rattlesnake. Chromatin accessibility and methylation levels regulate the expression of myotoxin. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2. RNA-seq y-axes represent venom-gland expression levels (TPM). Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: MYO, myotoxin; UncharPro, uncharacterized protein.

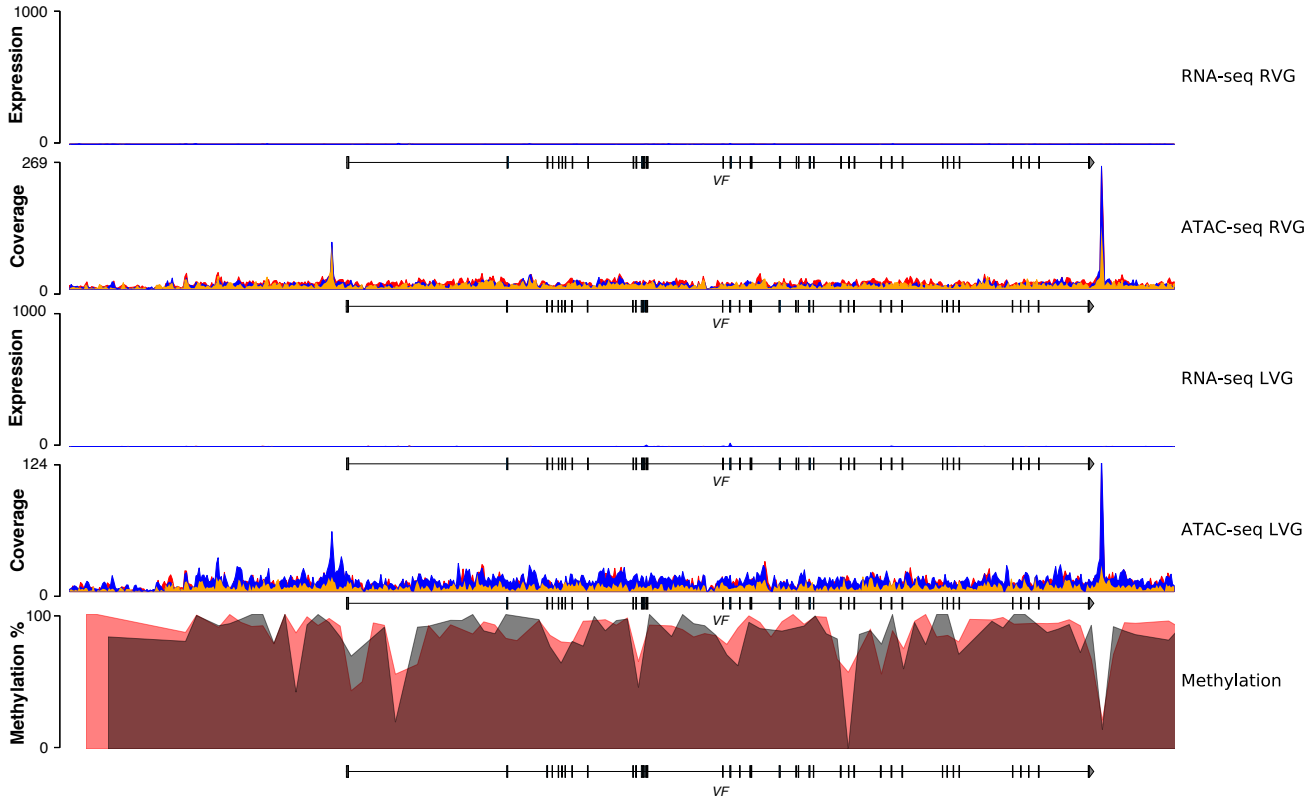

Fig. S29. **Regulatory landscape for VF in the Tiger Rattlesnake.** Chromatin accessibility and methylation levels regulate the expression of VF. Rows from top to bottom: right venom-gland (RVG) transcriptome, RVG ATAC-seq, left venom-gland (LVG) transcriptome, LVG ATAC-seq, and methylation percentage. No venom-gland specific ATAC-seq peaks were identified by Genrich or MACS2. RNA-seq y-axes represent venom-gland expression levels (TPM). Red, blue, and orange colors in the RNA-seq and ATAC-seq plots represent the three individuals sequenced, respectively; colors are overlaid and not all may be visible. For the methylation track, gray shading represents methylation levels in 2 kb bins in the pancreas (control) whereas red shading represents methylation levels in 2 kb bins in the venom gland. Gene annotations are shown under tracks where appropriate. Abbreviations: VF, venom factor.

Supplementary Table 1. Proteomic complexity for venom samples in Figure 1a.

Species	Shannon's H	Effective Peak #
C. tigris	1.639517	5.152681
C. adamanteus	2.900805	18.188782
C. horridus Type A	2.504663	12.239428
C. horridus Type B	2.735515	15.417683
C. scutulatus Type A	2.979810	19.684075
C. scutulatus Type B	2.684785	14.655046
C. atrox	2.935413	18.829271
C. viridis	2.448468	11.570602

Supplementary Table 2. RNA-seq raw data statistics. Total read pairs are post-trimming.

Sample ID	Tissue	Total Read Pairs	Alignment Rate (%)	NCBI Accession
CLP2741	Right Venom Gland	48,630,379	90.11	SRR11524051
	Left Venom Gland	18,202,650	90.63	SRR11524050
	Blood	30,979,002	92.15	SRR11524069
	Body Muscle	18,202,733	92.73	SRR11524058
	Brain	15,508,460	81.06	SRR11524057
	Cloaca	17,027,561	84.75	SRR11524056
	Eye	17,250,169	86.52	SRR11524055
	Gall Bladder	16,970,273	89.43	SRR11524054
	Gonad	17,604,769	75.40	SRR11524053
	Harderian Gland	22,387,479	77.92	SRR11524052
	Heart	7,312,245	89.74	SRR11524049
	Hemipenis	6,867,301	86.72	SRR11524048
	Kidney	6,940,121	84.28	SRR11524077
	Liver	19,005,589	87.85	SRR11524076
	Labial Gland	16,476,832	88.92	SRR11524075
	Lung	17,683,957	91.98	SRR11524074
	Pancreas	17,121,750	84.87	SRR11524073
	Pit	21,400,414	83.90	SRR11524072
	Small Intestine	20,856,222	90.14	SRR11524071
	Spleen	18,028,566	80.99	SRR11524070
	Stomach	18,449,110	64.62	SRR11524068
	Tail Muscle	20,675,086	65.47	SRR11524067
	Tongue	19,948,065	88.04	SRR11524066
	Venom Gland Muscle	19,309,760	87.63	SRR11524065
	Vomeronasal Organ	13,178,858	88.41	SRR11524064
CLP2742	Right Venom Gland	8,913,159	91.07	SRR11524063
	Left Venom Gland	22,092,017	89.44	SRR11524062
	Pancreas	17,302,239	86.40	SRR11524061
	Gonad	10,608,090	79.84	SRR11816479
	Liver	11,769,352	90.66	SRR11816478
	Vomeronasal Organ	44,161,215	89.81	SRR11816477
CLP2752	Right Venom Gland	33,028,059	89.92	SRR11524060
	Left Venom Gland	11,133,156	89.68	SRR11524059
	Liver	10,043,636	87.07	SRR11816476
	Vomeronasal Organ	21,689,070	82.31	SRR11816475
CLP1944	Combined Venom Gland	10,012,532	91.40	SRR11545023
CLP2130	Combined Venom Gland	12,374,178	92.12	SRR11545024
CLP2156	Combined Venom Gland	15,634,858	89.04	SRR11545022
KW1746	Combined Venom Gland	15,013,251	90.59	SRR5270853

Supplementary Table 3. Whole-genome bisulfite-sequencing raw data statistics.

Sample ID	Tissue	Uniquely Aligned Reads	Ambiguous Reads	Unaligned Reads	Total Reads	Total Coverage	CpG	CHG	CHH	NCBI Accession
CLP2741	Venom Gland (Active)	64,708,919	17,058,924	102,567,365	184,335,790	24×	77.06%	0.72%	0.98%	SRR11461883
CLP2742	Pancreas	42,527,349	10,658,072	67,886,989	121,072,822	17×	77.62%	0.62%	0.80%	SRR11461881
CLP2742	Venom Gland (Resting)	38,891,429	10,436,313	68,531,299	117,859,357	15×	77.03%	0.84%	1.02%	SRR11461882

Supplementary Table 4. ATAC-seq raw data statistics.

Sample ID	Tissue	Total Read Pairs	Total Coverage	Alignment Rate (%)	NCBI Accession
CLP2741	Right Venom Gland	85,446,006	14×	90.67	SRR11413282
CLP2741	Left Venom Gland	66,705,507	11×	87.36	SRR11413281
CLP2741	Harderian Gland	46,923,389	8×	95.81	SRR11413280
CLP2741	Labial Gland	47,920,246	8×	93.11	SRR11413279
CLP2742	Right Venom Gland	61,171,747	10×	81.68	SRR11413278
CLP2742	Left Venom Gland	26,111,571	5×	88.29	SRR11413277
CLP2742	Pancreas	83,165,940	15×	98.79	SRR11413276
CLP2752	Right Venom Gland	59,995,210	10×	89.19	SRR11413275
CLP2752	Left Venom Gland	64,830,921	11×	90.27	SRR11413274