BMC Genomics

RESEARCH

Open Access

A new chromosome-level genome assembly for western painted turtle *Chrysemys picta bellii*, a model for extreme physiological adaptations

Junhao Chen[1†], John R. Bermingham Jr[2*†], Patrick Minx[3], Milinn Kremitzky[4], Zhenguo Lin[1,2*] and Daniel E. Warren[1*]

Abstract

The western painted turtle, *Chrysemys picta bellii*, has the greatest tolerance to anoxia of any tetrapod studied to date. These turtles reside in the northern United States and southern Canada, and survive months of anoxia while submerged in ice-locked ponds and bogs. Reference genomes provide an important resource for elucidating the molecular bases for such unique physiological traits. The initial reference genome for this species, published in 2013, is highly fragmented, thereby limiting downstream analyses and biological interpretation. We created a new and improved assembly by combining PacBio HiFi, 10×Genomics Chromium, Hi-C sequence data and Bionano optical mapping derived from a single individual to generate a new haplotype-resolved chromosome-level reference assembly for *C. picta bellii*: "SLU_Cpb5.0". The genome size of the primary assembly is 2.372 Gb with a scaffold N50 of 133.6 Mb, a 6.5-fold improvement over the previous assembly. Annotation of SLU_Cpb5.0 revealed 12,242 novel genes compared to previous assemblies. PacBio Iso-Seq RNA sequencing using twelve tissues identified more than 100,000 novel transcript isoforms and 3,910 novel genes not previously annotated. This new genome assembly and annotation will support future comparative genomics studies, and the distinct patterns of tissue-specific isoform expression create a robust foundation for future characterization of the functions of these genes. Furthermore, to better understand the genetic basis of *C. picta bellii*'s extreme physiological adaptations and other aspects of its biology, we utilize existing RNA-seq data to identify dozens of novel, differentially expressed genes in the heart and brain of anoxic painted turtles.

Keywords Western painted turtle, Genome, *Chrysemys*, Anoxia tolerance, PacBio Iso-Seq, Transcript isoforms

†Junhao Chen and John R. Bermingham Jr contributed equally to this work.

*Correspondence:
John R. Bermingham Jr
jbermingham@wustl.edu
Zhenguo Lin
zhenguo.lin@slu.edu
Daniel E. Warren
daniel.warren@slu.edu
[1]Department of Biology, Saint Louis University, St. Louis, MO 63103, USA
[2]Department of Genetics, Washington University in St. Louis, St. Louis, MO 63110, USA
[3]Retired, St. Louis, USA
[4]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63110, USA

Background

Turtles are universally recognized because of their unique body plan, which results from the development of the shoulder girdle beneath fused ribs that form the dorsal carapace, formation of a ventral bony plastron, and replacement of teeth by a keratinous beak [1–4]. Additionally, they possess less conspicuous, but important specializations that make them fascinating and important research subjects, including long lifespans, temperature-dependent sex determination, and the ability to tolerate environmental extremes, especially low oxygen [5–7] and cold temperatures [8–11]. Notably, the western painted turtle (*Chrysemys picta bellii*, family Emydidae, Supplementary Figure S1A) has the greatest tolerance for anoxia (undetectable oxygen levels) of any tetrapod studied to date; as adults, they can survive anoxia for at least 30 h at 20 °C [12], and for more than 170 days at 3 °C [5, 7].

Chrysemys (painted) turtles are polytypic, with a very broad North American distribution (Supplementary Figure S1B). Until recently *Chrysemys* turtles were described as having diversified into four subspecies, but analysis of mitochondrial DNA supports division of *Chrysemys* into two clades, with Southern painted turtles sufficiently diverged to be considered a distinct species (*C. dorsalis*), and with *C. picta* consisting of three subspecies, Eastern (*Chrysemys picta picta*), Midland (*C. picta marginata*), and Western (*C. picta bellii*) [13, 14]. The *C. picta bellii* range exceeds those of the other *Chrysemys* taxa combined; currently it extends from northern Texas northward into southern Manitoba, and westward into Oregon, Washington, and southern British Columbia, with relict populations in New Mexico [15, 16]. *Chrysemys picta bellii*'s unique tolerance of anoxia, a critical adaptation to survive long winters, likely facilitated the species' range expansion.

The extreme hypoxia tolerance of painted turtles was first documented decades ago [5, 17]. Work since then has shown that survival relies on the avoidance of cellular ATP depletion during the anoxic bout [18–20], and preventing damage from reactive oxygen species (ROS) during reoxygenation (i.e., reperfusion injury; [21, 22]), the two principal causes of cellular injury in anoxia-sensitive animals, including humans [23, 24]. The first is achieved through a temperature-independent metabolic suppression of more than 90% [25], enabling ATP requirements to be met entirely through glycolysis, which is supported by the species' unusually large tissue glycogen stores [26, 27], and its highly carbonated skeletal mass, which prevents lethal decreases in body fluid pH caused by extreme lactic acidosis [6, 7]. The second is achieved primarily through the avoidance of mitochondrial ROS production [21, 22]. Many of the underlying mechanisms remain a mystery, and an improved western painted turtle genomic resource should facilitate their further characterization, potentially leading to the identification of new targets for therapeutic intervention in the treatment of ischemia-related conditions in humans associated with stroke, cardiac arrest, and organ transplantation. RNA-seq studies have examined anoxia tolerance [28], as well as sex determination [29] and endocrine-disrupting chemicals [30], but their interpretation depends on the quality of the genome to which they are aligned.

The genome of the western painted turtle was the first turtle assembly to be published [31], and despite attempts to improve it [32, 33], this original assembly remains highly fragmented. The data were generated using a combination of 454 technology, and Illumina short reads, compromising its utility for trait mapping and gene discovery. Further limitations of this original assembly include the geographic and individual origins of material used in the sequencing. Two turtles from the state of Washington, at the northwestern margin of the species' geographic range, were used, with one individual used for generating the genome sequencing data, and a second for assembling a Bacterial Artificial Chromosome (BAC) library. Subsequent Bionano optical maps were derived from a third turtle of uncertain origins [33]. These turtles may possess single nucleotide polymorphism (SNP) or structural differences with each other, and with turtles from the center of the range. To address these deficiencies, we generated a completely new reference genome for this species using tissue from an individual *C. picta bellii* turtle collected from McLeod County, MN near the center of the species' range (Supplementary Figure S1B). Datasets were generated using 10XGenomics, PacBio HiFi, and Hi-C sequencing, as well as Bionano optical mapping; all four datasets were used to generate new genome assemblies (Primary/merged, haplotype1, haplotype2). To annotate our assembly(ies), we generated PacBio Iso-Seq transcriptomes from twelve tissues, to identify as much transcript diversity as practical [34]. Additionally, we utilized previously generated Illumina short-read RNA-seq datasets from brain and heart of normoxic and anoxic adult and hatchling turtles acclimated to different temperatures [28, 35]. Collectively, these transcriptomes allow a more complete annotation of our new *C. picta bellii* reference genome than those currently available for other turtle genomes. The chromosomal level assembly and annotations for *C. picta bellii* generated in this study will facilitate future studies on the unique physiology and evolutionary biology of the western painted turtle, as well as their phylogeny.

Methods

Sample collection and DNA extraction

All genomic DNA used in this study was obtained from the same male western painted turtle, *Chrysemys picta bellii* (notch ID R12L10; Biosample SAMN13293682 (See

Supplementary Table S1 for NCBI accession numbers), which was purchased from Niles Biologicals (Sacramento, CA), who procured it from McLeod County, MN in 2017. RNA was obtained from different turtles: Two adults, and one hatchling turtle. All procedures involving turtles were performed in the Warren lab at Saint Louis University using protocols approved by the Saint Louis University IACUC. Turtles were euthanized by phenobarbital injection into the subcarapacial sinus. Fresh whole blood for 10× Genomics sequencing was collected from turtle R12L10 by cardiac puncture; other tissues were harvested, rinsed in PBS, and snap frozen in aluminum freeze-clamps cooled in liquid nitrogen, before being transferred to dry ice and stored at −80 °C. Frozen adult liver tissue was used for PacBio HiFi sequencing and for Bionano optical mapping. Frozen adult skeletal muscle was used for Hi-C sequencing.

10 × Genomics sequencing

DNA extraction, library preparation, and sequencing were performed at the McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO. DNA was extracted from fresh, whole blood collected from heart, using Qiagen MagAttract, then used to prepare a library with 10 × Chromium Genome v2, and sequenced to 74X coverage on the 10 × Genomics Chromium platform. This technology partitions high molecular weight genomic DNA fragments into droplets, where they are amplified with unique sequence identifiers (barcodes), to mark closely linked sequences prior to Illumina "short read" sequencing [36]. Approximately 150 Gb of sequence was obtained on a NovaSeq S4 300XP (Sequence Read Archive #SRX7705481; see Supplementary Table S1 for dataset accession numbers). The sequence was assembled using the de novo assembly program Supernova v2.1.1, and this genome assembly was submitted to NCBI, accession #GCA_011386835.1. Unassembled, barcoded 10 × Genomics reads were used in subsequent assemblies.

PacBio HiFi sequencing of genomic DNA

Pacific Biosystems (PacBio) HiFi genomic sequencing was performed by the DNA Sequencing Center at Brigham Young University, Provo, UT. Briefly, High molecular weight (HMW) DNA was isolated and cleaned from turtle liver using a Qiagen Genomic-tip kit and columns. Once isolated, the DNA was sheared on a Megaruptor (Diagenode/Hologic), AMPure cleaned, and made into a SMRTbell adapted library following PacBio manufacturer's protocol. Library size selection was performed on a Blue Pippin (Sage Science, Beverly MA), collecting DNA fragments 10 kb and larger. The size-selected library was prepared to run on a Sequel II using SMRT-Link recommendations. Three SMRTcells were used (Supplementary Table S2, Supplementary Figure S2), and their datasets were merged. The raw sequencing data has been submitted to NCBI SRA database (NCBI BioProject PRJNA589899; Sequence Read Archive #SRX22568648; Supplementary Table S1).

Hi-C sequencing

A Hi-C library was generated using DNA isolated from adult male liver tissue by Arima Genomics, Inc., Carlsbad, CA, using the Arima High Coverage Hi-C kit, part #A410110, following manufacturer's protocols. DNA was crosslinked using the standard input protocol, and digested with a cocktail of four restriction enzymes. The Hi-C library was prepared using the Swift Biosciences Accel-NGS 2S Plus DNA Library kit, and sequenced on an Illumina NovaSeq (Sequence Read Archive #SRX22568649; Supplementary Table S1).

Bionano optical mapping

Ultra-high molecular weight DNA was extracted from liver tissue, and fluorescently labeled using DLE-1. Labeling and imaging were performed by McDonnell Genome Institute at Washington University in St. Louis according to manufacturer's protocol (Bionano Genomics, San Diego CA). DNA molecules greater than 150 kb with a minimum of 9 labels were collected for Bionano labeling and loaded on the Saphyr chip for imaging. These data were assembled using Bionano Solve v.3.7 software (Bionano Genomics) with nonhaplotype_noES_DLE1_saphyr.xml parameters and placed to the assembled genome for validation.

Genome assembly

To achieve optimal assembly results, we tested different tools at each stage of the genome assembly process (Supplementary Figure S3). We evaluated the assemblies using various metrics, including N50, total length, total sequence number, and BUSCO scores, to determine the most effective assembly method [37]. Initial de novo assemblies of the turtle genome were generated by different tools that provide haplotype-resolved assemblies, including Hifiasm (version: 0.16.1-r375) [38], NextDenovo (version: 2.5.0) [39], and Wtdbg2 (version: 2.5) [40]. Hifiasm, which used both PacBio HiFi data and Hi-C data, yielded three assemblies (primary, haplotype1 & 2) that had the best metrics in most categories; thus, this assembly was selected for further scaffolding. A first-round scaffolding of the three assemblies was performed using 10 × Genomics data with tigmint (version: 1.2.6) [41, 42] and ARCS (version: 1.2.4) [41, 42]. A second round of scaffolding for each assembly was completed using Hi-C data by 3D-DNA (version: 180922) [43]. This process involved using Hi-C data to order and orient scaffolds and to fill gaps between the scaffolds.

Manual curation was performed using Bionano data with Bionano's official toolkit (Solve & Access [44]) to further refine the assembly. The final *Chrysemys picta bellii* assembly, SLU_Cpb5.0, was deposited to NCBI BioProject PRJNA589899, and was adopted by NCBI as the reference assembly for western painted turtle (Genome assembly ASM1138683v2 and RefSeq assembly accession GCF_011386835.1; Supplementary Table S1).

Estimated genome size and degree of heterozygosity of western painted turtle R10L12 was inferred using HiFi reads and 10 × Genomics Illumina reads. K-mer frequency files of the two types of sequencing data were obtained by Jellyfish (version: v2.3.0) [45] based on a k-mer size of 21 (k = 21). These files were used as input for GenomeScope 2.0 (version: v1.0.0) [46] to estimate the genome size and heterozygosity level.

Comparisons of genome assembly and BAC clones

Only the BACs published in [32] were assigned NCBI Accession numbers, and therefore could be evaluated. These were queried on the NCBI BLAST site in January, 2024. These BAC sequences were used as queries for BLASTN searches against whole-genome shotgun contigs (wgs) of *Chrysemys picta bellii* (taxid:8478) on NCBI, optimized for highly similar sequences, typically with the following algorithm parameters: Max target sequences = 10; Expect threshold = 0.001; Word size = 256, Match/Mismatch scores = 4,−5; Gap costs = Existence:12,Extension:8; Filters: Low complexity regions; Species-specific repeats for *Chrysemys picta bellii*; Mask for lookup table only.

Assembly and comparative analysis of mitochondrial genome

We used MitoHiFi (version: v3.2.2) [47] to assemble the mitochondrial genome, with the primary assembly output from Hifiasm as input, the *C. picta bellii* mitochondrial genome NC_023890.1 in FASTA and GenBank formats as a template, and including the –mitos parameter in order to use MITOS2 (version: v2.0.2) [48] for annotation.

Genome annotation

The *Chrysemys picta bellii* genome assembly (GCF_011386835.1) was annotated using the NCBI Eukaryotic Genome Annotation Pipeline (EGAP, version: 10.2) [49]. In brief, repetitive sequences were masked using WindowMasker [50]. Gene prediction incorporated multiple sources of evidence. RNA-Seq data were obtained from the NCBI Sequence Read Archive (SRA) under BioProjects PRJNA589899, PRJNA323738, PRJNA371383, PRJNA438493, PRJNA526071, PRJNA594037, PRJNA919086, PRJNA931617, PRJNA1025287 and PRJNA1025288, comprising 8.6 billion reads from 48 samples representing diverse tissues (gonad, liver, ventricle, adrenal-kidney-gonad complex) and developmental stages (embryonic stages 9–22, hatchling, juvenile, and adult). RNA-Seq reads were aligned to the genome using STAR mapping to genome with default parameters [51]. Long-read sequences (SAMN13293682) were aligned to genome using minimap2 [52] with default parameters. Transcript evidence from 11 same-species RefSeq sequences was aligned using Splign [53]. Protein evidence was derived from RefSeq and GenBank sequences of related reptile species (*Malaclemys terrapin pileata*, *Emys orbicularis*), Xenopus, other Sauropsida taxa, and Homo sapiens, aligned using ProSplign [49]. Gene models, including alternative splicing isoforms, were predicted using the Gnomon algorithm, which integrates aligned evidence with ab initio predictions [49]. The unmasked RefSeq genome sequence file was utilized to identify small non-coding RNAs (sncRNAs). Transfer RNAs (tRNAs) were predicted using tRNAscan-SE (version: v2.0.12), while ribosomal RNAs (rRNAs), small nucleolar RNAs (snoRNAs), and small nuclear RNAs (snRNAs) were annotated through the alignment of eukaryotic RFAM Hidden Markov Models (HMMs) against the genome using Infernal's cmsearch tool (version: 1.1.5) [54–56]. The genome annotation files are available on the NCBI FTP server (see Annotation dataset, Supplementary Table S1). Annotation of transposable elements and repetitive sequences was conducted using the primary assembly by EarlGrey (version: v5.0.0) with default parameters [57]. The annotation report is available at: https://www.ncbi.nlm.nih.gov/refseq/annotation_euk/Chrysemys_picta_bellii/GCF_011386835.1-RS_2024_05/.

PacBio Iso-Seq RNA sequencing and characterization of tissue-specific transcript isoforms

To obtain more complete genome annotations and to identify tissue-specific transcript isoforms, we conducted PacBio Iso-Seq RNA sequencing for 12 tissues: 9 from an adult turtle and 3 from a hatchling turtle. The adult tissues used were: Testis, liver, kidney, lung, skeletal muscle, spinal cord, sciatic nerve, duodenum, and pooled thalamus, tectum, cerebellum and remaining hindbrain tissues, hereafter referred to as "hindbrain". Because of existing adult forebrain RNA-seq datasets, this tissue was excluded to maximize novel transcript discovery. Carapace, kidney, and (entire) brain were harvested from a hatchling turtle. Tissues were snap frozen in liquid N_2 and stored at −80 °C prior to RNA extraction. RNA extraction from these tissues, preparation of Iso-Seq libraries and sequencing using PacBio SMRT cell (Sequel II) were performed by the DNA Sequencing Center at Brigham Young University, Provo, UT. For PacBio Iso-Seq reads, histograms of subread and HiFi read distributions are

shown in Supplementary Figure S4, and summary statistics are presented in Supplementary Table S3.

Our Iso-Seq subreads data were first converted into 4,915,864 circular consensus sequences (CCS) using the ccs software (https://ccs.how/, version: 6.4.0) with a minimum read quality of 0.9 (–min-rq 0.9) to obtain more CCS reads. These CCS reads were then processed to extract full-length (FL) reads using lima (https://lima.how/) with the Iso-Seq processing mode (–Iso-Seq and –peek-guess). To remove chimeric reads, these FL reads were then used to generate full-length non-chimeric (FLNC) reads with the "refine" function of Iso-Seq3 https://Iso-Seq.how/, version: 3.8.0). This procedure yielded separate FLNC read files for each of the 12 tissues (a total of 3,822,332 FLNC reads, Supplementary Table S3).

Each tissue-specific FLNC read file was aligned to the reference genome (SLU_Cpb5.0) using pbmm2 (v 1.9.0, https://github.com/PacificBiosciences/pbmm2) generating 12 BAM files. Redundant transcript models were collapsed into isoforms using the 'collapse' function in TAMA (v b0.0.0) [58], producing a BED file that was then converted to GTF format using the TAMA conversion script (tama_convert_bed_gtf_ensembl_no_cds.py).

To assess the quality of transcript isoforms inferred using Iso-Seq, including completeness, splice junctions, and transcript ends, short-read RNA-seq data (Supplementary Table S4) were mapped to the reference genome (SLU_Cpb5.0) using HISAT2 [59], and used as input for "sqanti3_qc.py" in SQANTI3 [60]. Following QC, transcripts were filtered with SQANTI3's filtering module using a stringent structural category-specific criteria to minimize the impact of potential internal priming and incomplete reverse transcription. The filtered transcript isoforms, consisting primarily of incomplete splice match isoforms (Supplementary Figure S5), can result from degraded RNA; they varied from 5 to 18% of the reads (Supplementary Table S3); indicating that the input RNA was generally of good quality. Filtered transcript isoforms were then classified into three types by comparison to reference transcripts: 1) matched transcripts of annotated genes, 2) novel transcripts of annotated genes, and 3) novel transcripts of novel genes (Supplementary Figure S5). Matched transcripts showed a perfect match of splicing patterns to reference transcripts, with transcription start sites (TSS) and transcription termination sites (TTS) within 50 bp from annotated positions (GCF_011386835.1-RS_2024_05). Novel transcripts of annotated genes displayed alternative splicing patterns relative to the reference annotations, with all splice junctions canonical or supported by at least 10 RNA-seq reads, or TSS/TTS located > 50 bp from annotated positions. Novel transcripts from novel genes were identified if they contained at least two exons, and had all splice junctions canonical or supported by ≥ 10 RNA-seq reads. Novel genes were classified as protein-coding if they contained a predicted ORF ≥ 300 bp; otherwise, they were considered non-coding. SQANTI3 configuration file, tissue-specific GTF annotation files, and translated amino acid sequences are available on figShare (https://doi.org/10.6084/m9.figshare.29402339). Filtered annotation from all tissues were merged using the "Merge" function in TAMA [58] to generate single consensus annotation file. During the annotation merge process, some merged transcripts displayed different patterns of matching to the reference annotation or different coding potentials (classified as "coding" if harboring a predicted ORF ≥ 100 amino acids with a positive log-likelihood coding score, or "non_coding" otherwise). A transcript was classified as a known if it exhibited a full splice match to any reference-annotated transcript in at least one tissue. Similarly, a gene was designated as protein-coding if at least one of its isoform transcripts met the criteria for "coding". The merged annotation file is available on figshare (https://doi.org/10.6084/m9.figshare.29402339). Functional enrichment analysis was performed using clusterProfiler (version: 4.6.0) [61] in R (version: 4.2.0) [62]. Due to the lack of a species-specific Gene Ontology database for *Chrysemys picta bellii*, gene symbols were mapped to human orthologs, and enrichment analysis was conducted using org.Hs.eg.db (version: 3.22.0) [63] as the reference database. The multiple testing enrichGO function was used with Benjamini–Hochberg correction. The full set of annotated genes in org.Hs.eg.db was used as the background gene set.

Evaluations of the impact of new genome assembly on differential expression analysis using RNA-seq data

We retrieved raw sequencing reads of RNA-seq data obtained in turtles under anoxia exposure previously generated by our group [28, 35]. Briefly, western painted turtles were maintained at 19 °C and subjected to 24 h of anoxia by submergence in nitrogen-bubbled water. Telencephalon and ventricle tissues were collected for RNA extraction. Paired-end (2 × 100 bp) reads were generated on an Illumina HiSeq 2000 platform from poly-A selected RNA-Seq libraries, with four biological replicates per tissue under both anoxic and normoxic conditions. Detailed sample information can be found in Supplementary Table S4. Quality control and adapter trimming for aw reads were preprocessed using fastp (version: 0.23.4) [64] with default parameters. Clean reads were aligned to the older assembly (CPI3.0.4, GCF_000241765.5) to and SLU_Cpb5.0 version using HISAT2 (version: 2.2.1) [59] with default parameters. Gene level read counts were quantified using featureCounts (version: 2.0.2) [65] with -p flag for paired-end reads. Differential expression analysis was performed using DESeq2 (version: 1.48.1) [66].

Genes with an adjusted p-value < 0.05 and absolute log2 fold change > 1 were considered differentially expressed. Volcano plots were generated using the EnhancedVolcano package [67]. Scripts used for RNA-seq analysis are provided on figShare (https://doi.org/10.6084/m9.figshare.29402339) as "turtle_genome_scripts.zip".

Evaluations of the impact of new genome assembly on small RNA sequencing alignment

Small RNA sequencing data were obtained from the NCBI Sequence Read Archive under BioProject PRJNA375813. Raw reads were processed using fastp (version: 0.23.4) with default parameters to remove low-quality reads and adapter sequences. To evaluate the improvement of the updated genome assembly, clean reads were aligned to both the current reference genome (GCF_011386835.1) and the previous assembly (GCF_000241765.5) using HISAT2 (version: 2.2.1) with default parameters [59]. The resulting SAM files were converted to BAM format using SAMtools (version: 1.18) [68]. Alignment rates were compared between the two genome versions and visualized alongside RNA-seq data as violin plots were created using tidyverse [69] with ggsci [70] color palettes.

Results

De novo assembly of SLU_Cpb5.0, a new *Chrysemys picta bellii* genome

Initially we generated a new *C. picta bellii* genome assembly from 10 × Genomics linked reads [36] – 572.7 million pairs of Illumina reads (2 × 150 bp), with a total length of 171.8 Gb. This 10 × Genomics assembly (ASM1138683v1; referred as SLU_Cpb1.0 Supplementary Table S1; Supplementary Figure S3) consisted of 132,827 scaffolds with a Scaffold N50 of 15.9 Mb. This assembly had 99.4% sequence identity with the original *C. picta bellii* reference genome assembly [31, 32] and 98.2% sequence identity with the *T. scripta* assembly [71], but unexpectedly, more 10 × Genomics scaffolds could be aligned to the *T. scripta* genome than to the *C. picta* genome (L. Hillier, personal communication). Therefore, we sought to construct an improved, chromosome-level genome assembly for *C. picta bellii.* We generated additional PacBio HiFi long reads, Hi-C conformation capture sequence data, and Bionano optical genome maps from the same individual turtle (Supplementary Figure S3). We obtained a total of 3.48 million HiFi reads from three PacBio SMRT cells (Supplementary Table S2, Supplementary Figure S2). The average HiFi read lengths range from 13,876 to 14,024 bp, with median read Phred quality scores between 34 and 35 (Supplementary Table S2). The total length of HiFi reads was 48.5 Gb, providing an average coverage of ~ 20 × based on the previously assembled genome size of 2.4 Gb [31]. Our Hi-C sequencing produced 496.7 million pairs of Illumina reads (2 × 150 bp) (Supplementary Table S1).

To achieve the best assembly quality, we tried different de novo genome assembly tools that provide haplotype-resolved assemblies, including Hifiasm [38], NextDenovo [39], and Wtdbg2 [40]. We noted that some assemblies with better contiguity statistics could possess incorrectly merged contigs [72]. We found that Hifiasm, with integration of Hi-C reads for haplotype phasing, yielded the best metrics in most categories for primary, haplotype1, and haplotype2 preliminary assemblies (assembly version SLU_Cpb2.0, Contig N50 size: primary assembly = 28.07 Mb, haplotype 1 = 3.54 Mb, and haplotype 2 = 3.57 Mb; Supplementary Table S5, Supplementary Figure S3). Thus, the assemblies that combined PacBio HiFi + Hi-C datasets by Hifiasm were used for subsequent scaffoldings.

The first round of scaffolding for the preliminary assemblies was performed with the 10 × Genomics linked reads using Tigmint (version: 1.2.6) [41]. The scaffold N50 size for the primary assembly was improved to 42.12 Mb, and 38.31 Mb and 43.67 Mb for haplotype 1 and 2, respectively (assembly version SLU_Cpb3.0, Supplementary Table S5). Hi-C sequencing reads were used for the second round of scaffolding by 3D-DNA [43], which linked scaffolds into 25 pseudochromosomes (assembled sequences that presumably correspond to actual chromosomes; assembly version SLU_Cpb4.0, Supplementary Figure S6). Hi-C scaffolding significantly improved the contiguity of all assemblies, as scaffold N50 sizes for the primary assembly, haplotype 1 and haplotype 2 were increased to 133.42 Mb, 132.97 Mb, and 132.08 Mb, respectively. All these N50 values are a significant improvement to the 20.68 Mb of the previous assembly (NCBI RefSeq assembly GCF_000241765.5 "Chrysemys_picta_Bionano-3.0.4", abbreviated as CPI3.0.4 hereafter) (Table 1). The primary assembly was aligned with Bionano optical maps, and major discrepancies were corrected manually (Supplementary Figures S3 and S7).

Our final primary assembly version "SLU_Cpb5.0", has a total size of 2.372 Gb, which closely matches the estimated genome size of 2.314 Gb based on k-mer analysis using10x Genomics Illumina reads (Supplementary Figure S8A), but is larger than 2.102 Gb inferred using PacBio HiFi reads (Supplementary Figure S8B). The difference in estimated genome size based on k-mer analysis between 10 × Genomics Illumina reads and PacBio HiFi long reads is likely because our 10 × Genomics data have a much higher sequencing depth (72.5 × vs 20.5 ×), and a higher sequencing depth generally provides a more accurate estimation of genome size in k-mer analyses [46]. We used our HiFi read data to assemble the mitochondrial genome of *C. picta bellii* with MitoHiFi [47], which

Table 1 Comparison of summary statistics for previous and new *Chrysemys picta bellii* primary and haplotype assemblies

	CPI3.0.4	SLU_Cpb5.0	SLU_Cpb5.0	SLU_Cpb5.0
		Primary	Hap1	Hap2
Assembly size (nts)	2,481,368,539	2,372,089,908	2,336,887,892	2,329,101,047
Total sequence #	76,388	9,850	5,376	9,371
Max. sequence length	163,127,350	360,060,246	360,398,192	362,207,599
Scaffold N50 length	20,675,201	133,555,881	132,972,522	132,077,892
# N50	25	6	6	6
Average sequence length	32,484	239,540	434,689	248,543
GC%	44.19%	44.88%	44.79%	44.76%
# Ns	309,640,510	541,420	576,246	427,110
Ns%	12.48%	0.02%	0.02%	0.02%

The table compares assembly statistics for the older CPI3.0.4 assembly [31, 32] and the SLU_Cpb5.0 assembly. PacBio HiFi and Hi-C sequencing generate both a primary assembly and haplotype-resolved assemblies. #N50 is the number of scaffolds, starting with the largest, that are required to cover 50% of the genome. #Ns refer to unknown bases in the assemblies

resulted in a circular sequence with a length of 16,931 bp. Annotation of the mitochondrial genome using MITOS2 [48] predicted 13 protein-coding genes, 22tRNA genes and 4 rRNA genes.

The primary and haplotype assemblies consist of 25 nuclear pseudochromosome sequences (placed scaffolds), corresponding to the species' 25 chromosomes. As chromosomes are conventionally numbered according to size, the pseudochromosomes from our assembly are numbered from largest to smallest accordingly (Fig. 1A). The total number of unplaced scaffold sequences has significantly improved, with SLU_Cpb5.0 having only 9,850, compared to 76,388 in the previous assembly CPI3.0.4. In addition, the average scaffold sequence length was increased by 7.4x, the percentage of Ns was reduced from 12.48% to 0.02%, which partially explains the smaller estimated size of the current assembly (2.372 Gb vs. 2.481 Gb) (Table 1).

To assess the structural and sequence-level consistency between the two haplotypes, we conducted collinearity analysis between the haplotype 1 and haplotype 2 assemblies. As shown in Fig. 1B, the two haplotypes align well with one another and are consistent in terms of large-scale genomic structure, with 99.85% and 99.44% reciprocal genome coverage and 99.48% sequence identity. The mean heterozygosity rate between the two haplotypes is 0.48%, which is consistent with 0.548% based on 10×Genomics Illumina reads using k-mer analysis by GenomeScope 2.0 [46], but is lower than 0.712% estimated using PacBio HiFi raw reads (See Material and Methods, Supplementary Table S6, and Supplementary Figure S8). The discrepancy in estimated heterozygosity between the two types of sequencing data is likely attributed to their different sequencing depths, or to poorer alignment of short reads at polymorphic sites. Nevertheless, all these numbers suggest high levels of genetic diversity in Minnesota western painted turtles when compared to individuals of different turtle species [75], as well as to the *C. picta bellii* individual from Washington state [76].

The new assembly SLU_Cpb5.0 is a significant improvement over the previous assembly CPI3.0.4 for the *C. picta bellii* genome

The completeness of the primary and haplotype assemblies was evaluated with the BUSCO v5 using the sauropsida_odb10 dataset [37]. As shown in Fig. 2A, the primary assembly achieved a BUSCO completeness score of 99.1%, indicating near-complete recovery of single-copy orthologs.

The mapping rate of RNA-seq reads can be used as a useful metric to evaluate the quality of a genome assembly, particularly in terms of completeness. To determine the extent to which our new assembly improves RNA-seq mapping rate, we downloaded 59 RNA-seq datasets of *C. picta bellii* from the NCBI SRA database (Supplementary Table S4). These RNA-seq datasets were mapped to both SLU_Cpb5.0 and the CPI3.0.4 assemblies using HISAT2 [59]. Our results show that the RNA-seq mapping rate using the SLU_Cpb5.0 assembly is significantly higher than that for the 3.0.4 assembly (median 91.32% vs. 80.25%, Fig. 2B), demonstrating that our assembly captures more coding regions in the genome and has a lower number of assembly errors and misassembled regions. Interestingly, the improvement in mapping rates for microRNA sequencing (miRNA-seq) was modest compared with those for regular RNA-seq data (Fig. 2B), suggesting that the smaller miRNA genes are less sensitive to fragmented genome assemblies than are the longer mRNAs.

Aligning assembled sequences to those of a closely related species also reveals insights into the accuracy, completeness, and structural integrity of the assembly. We aligned our assembly to the reference genome of a closely related deirochelid species *Trachemys scripta* (NCBI RefSeq assembly GCF_013100865.1 "CAS_Tse_1.0"), which also contains 25 chromosomes [71]. As shown in Fig. 2C, the lengths of 25 chromosomes are highly similar between SLU_Cpb5.0 and *T. scripta* "CAS_Tse_1.0", with 95.74% of the SLU_Cpb5.0 assembly aligning to the *T. scripta* genome. In contrast, the assembled

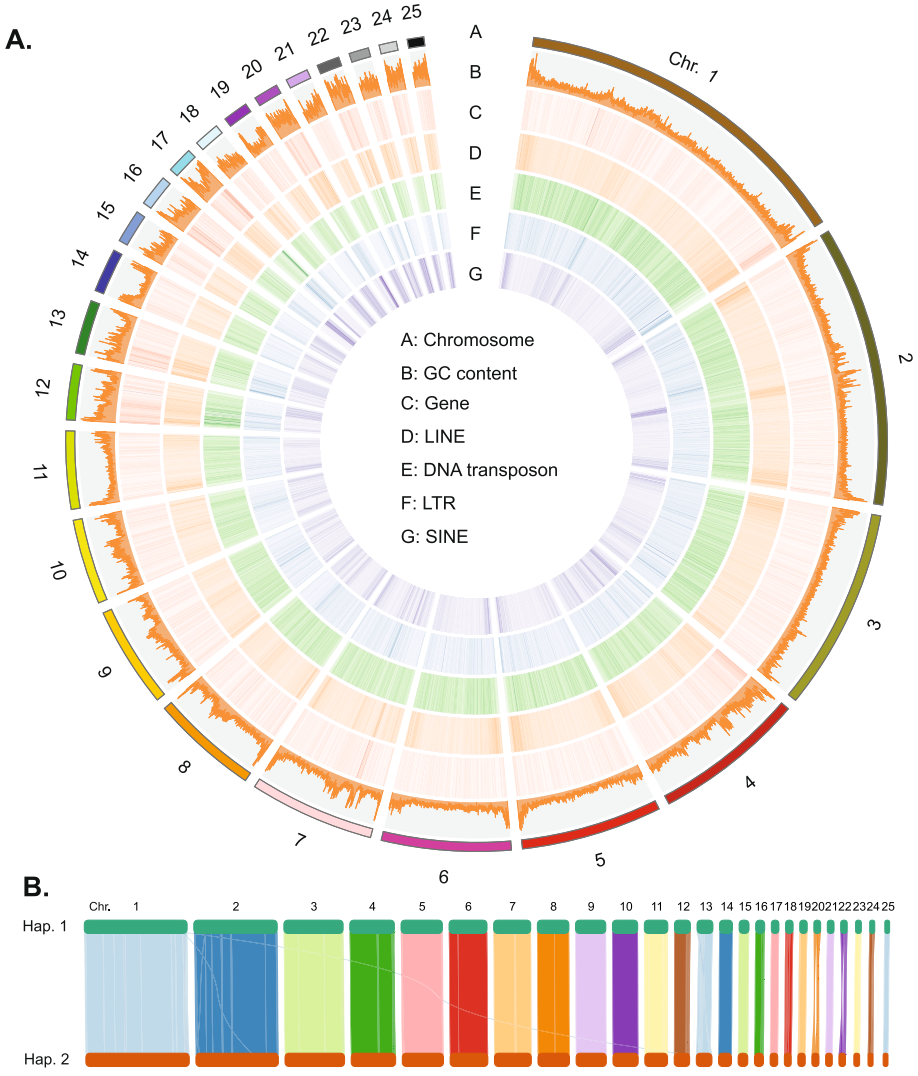

Fig. 1 Structure of primary assembly of the *Chrysemys picta bellii* genome SLU_Cpb5.0. **A** Circos plot of the 25 pseudochromosomes of the *C. picta bellii* genome. Different genomic features are depicted in seven concentric circles. The outer circle, labeled A, depicts the 25 assembled pseudochromosomes, which correspond to cytological chromosomes. Circles B-C: GC content and gene density, respectively. GC content increases toward telomeres and perhaps near some centromeres, and is increased in the smaller microchromosomes. Circles D-G represent densities of repetitive elements: LINEs, DNA transposons, LTRs and SINEs. Diagram of the assembled genome was generated using Circos [73], **B** Collinearity between the 25 chromosomal pairs of haplotypes 1 and 2. Each line between the haplotypes represents a colinear block of 100 kb. The two haplotypes are largely colinear; subtle differences may reflect heterozygous indels or could be assembly errors. The collinearity plot was created using NGenomeSyn [74]

chromosomes in the CPI3.0.4 assembly are significantly shorter than the other two assemblies, reflecting the many unmapped scaffolds in the CPI3.0.4 assembly. In addition, higher degrees of collinearity are observed in orthologous chromosomes between SLU_Cpb5.0 and *T. scripta*, than between SLU_Cpb5.0 and the CPI3.0.4 assembly, particularly from chromosome 4 to chromosome 10 (Fig. 2C). These results demonstrate that the CPI3.0.4 assembly is not only highly fragmented, but also misassembled across many chromosomes.

Bacterial Artificial Chromosomes (BACs) can serve as valuable tools to evaluate genome assembly quality, and many *C. picta bellii* BACs have been mapped to chromosomes by fluorescence in situ hybridization (FISH; [32, 33]). Therefore, we localized those BACs that have been assigned NCBI accession numbers to our SLU_Cpb5.0 assembly, to correlate the chromosomes that have been labelled by FISH to the pseudochromosomes that have been generated by our assembly. Cytogenetic chromosome and sequence-based pseudochromosome numbers are listed in Supplementary Table S7. Two BACs, CHY3-12H12 and CHY3-34H12, appear to have been mis-localized by FISH. [32, 33] list CHY3-12H12 as residing on chromosome 19 of *C. picta bellii* and *Trachemys scripta*, and additionally, [77] lists this BAC as residing on chromosome 20 of *Saurotypus triporcatus,*

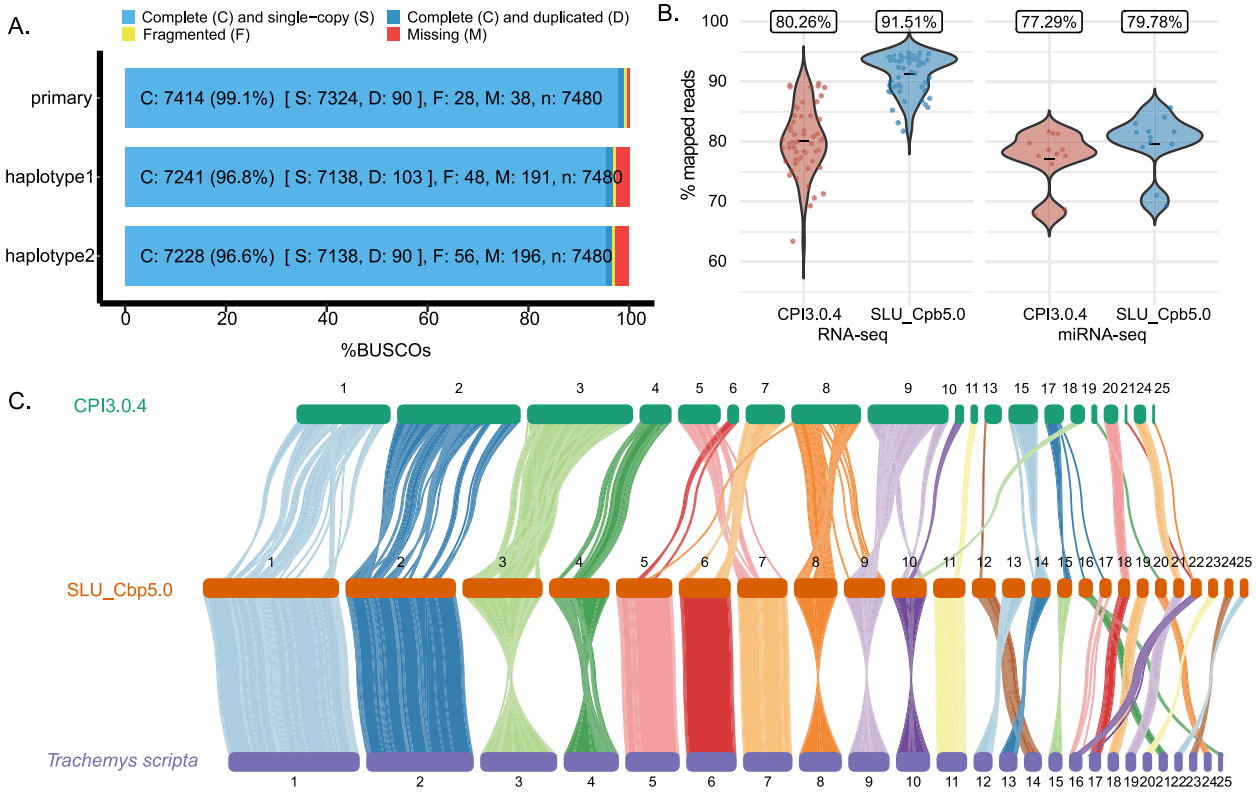

Fig. 2 Evaluation of the SLU_Cpb5.0 assembly(ies) of the *C. picta bellii* genome. **A** BUSCO plot for the primary (merged) genome, haplotype 1, and haplotype 2. The results are categorized into four groups: Complete (C), which includes single-copy (S) and duplicated (D) orthologs; Fragmented (F), representing partially recovered orthologs; and Missing (M), indicating orthologs absent in the dataset. The number of genes and percentages are shown for each category. These results are based on comparison with the sauropsida_odb10 lineage dataset. **B** Improved alignment rates of RNA-seq reads to new *C. picta bellii* genome assembly. Violin plots display the alignment rates for RNA-seq datasets. Each dot is a dataset, and it is plotted according to the percentage of its reads that align to each genome assembly. For each violin plot, the median alignment percentage is depicted by a horizontal bar, and is listed at the top. The percentage of reads aligned for mRNAs (left) and microRNAs (right) is better for the SLU_Cpb5.0 assembly than for the 3.0.4 assembly. Increasing assembly quality has a greater impact on mapping efficiency of mRNAs than it does for the smaller microRNAs. **C** Pseudochromosome homologies between *Chrysemys picta bellii* and *Trachemys scripta* assemblies. Alignment of the *C. picta bellii* assembly SLU_Cpb5.0 with assembly CPI3.0.4 and with the red-eared slider genome assembly CAS_Tse_1.0 [71], showing substantial contiguity between the two chromosome-level assemblies. The figure was generated using NGenomeSyn [74], with a 10 Kb cutoff for homologies

the Mexican musk turtle. However, its sequence (NCBI accession AC239503.2) resides on pseudochromosome 3 in the SLU_Cpb5.0 assembly. Similarly, BAC CHY3-34H12 (accession AC238643.2) was listed in [32] as residing on *C. picta bellii* chromosome 1p, whereas its sequence covers 195.9–196.1 M on pseudochromosome 3 in the SLU_Cpb5.0 assembly. BLAST searches of Whole Genome Sequence datasets for *Trachemys scripta*, *Sauroptypus triporcatus*, and *Mauremys reevesi* genomes confirmed that both AC239503.2 and AC238643.2 sequences reside chromosome3/linkage group 3. We believe that our placement of these BACs is correct for two reasons. First, it is extremely unlikely that independent assemblies of four different turtle genomes all misassembled this sequence. Second, our *C. picta bellii* Bionano optical map of this region shows no evidence of mis-assembly (data not shown). These observations demonstrate the utility of correlating whole genome sequences and FISH results, and will improve the reliability of future turtle cytological studies.

Annotation of the *C. picta bellii* SLU_Cpb5.0 genome assembly

Annotation of the primary assembly was completed by the NCBI eukaryotic genome annotation pipeline (see Materials and Methods) [49]. The final annotation yielded 35,699 genes, including 25,216 protein-coding genes and 1,155 pseudogenes. For non-coding RNA, the annotation identified 219 small nuclear RNA (snRNA) genes, 283 small nucleolar RNAs (snoRNAs) genes, 1,269 ribosomal RNA (rRNA) genes, 468 transfer RNA (tRNA) genes, and 6965 long non-coding RNA (lncRNA) genes. The genome coverage analysis showed that 5.27% of the assembly is covered by exons, while introns account for 43.51%. Among the protein-coding genes, 20,705 (82.1%) were assigned to at least one Gene Ontology (GO) term, suggesting potential function(s). At the transcript level, a

total of 114,986 transcripts were annotated, with an average of 3.22 transcripts per gene. Among them, 85,016 are mRNA transcripts, spanning lengths from 156 to 114,607 bp.

Transposable elements and repetitive sequences were inferred from the primary assembly using Earl Grey with default parameters [57]. The four largest groups of repetitive sequences are LINE, DNA transposon, LTR and SINE, which account for 19.07%, 9.58%, 5.2%, 2.01% of the turtle genome, respectively (Fig. 3A). The proportions of SINE and LTR sequences we observed are similar to those observed in several other turtle assemblies, including the original assembly of western painted turtle [31], Table S3 in [71], but we found a higher proportion of LINE elements, and somewhat lower proportion of DNA elements. We attribute this observation to improved genome continuity.

When we compared the annotations between the two assemblies, the new SLU_Cpb5.0 annotation includes 12,242 genes that are absent in the old CPI3.0.4 (Fig. 3B). Conversely, 2,991 genes previously annotated in the old CPI3.0.4 are now absent in the new SLU_Cpb5.0. These differences likely result from an increase in stringency of annotation parameters in the last decade, leading to the exclusion of some previously annotated candidate genes from our annotation. Among the newly predicted genes, 4,952, (40.5%) are annotated as long noncoding RNAs), and 4,730 (38.6%) as protein-coding genes (Fig. 3C). The third largest group of newly annotated genes are rRNA genes (1,262). This is notable because eukaryotic genomes typically contain numerous rRNA loci with high sequence identities, which poses significant challenges for assembly with short-read sequencing technologies. By utilizing PacBio long reads, we were able to achieve a substantial improvement in the assembly of these repetitive rRNA loci.

Identification of tissue-specific transcript isoforms in western painted turtle using PacBio Iso-Seq data

We generated PacBio Iso-Seq long-read transcript datasets using 12 tissues from adult and hatchling western painted turtles (Supplementary Tables S3 and Figure S4). These tissues were chosen either because they are expected to express a large fraction of the genome (testis), are important for anoxia/cold tolerance (hatchling shell, adult and hatchling kidney, adult liver, adult skeletal muscle, adult hindbrain, and hatchling brain), or because they are underrepresented in other painted turtle transcript datasets (adult lung, adult spinal cord, adult sciatic nerve, adult hindbrain, and adult duodenum) and would, therefore, expand the completeness of gene identification in our assembly. Alternative splicing is prevalent in vertebrates, and the expression of many transcript isoforms is very often restricted to specific tissues [78]. These Iso-Seq datasets also allow us to improve the completeness of genome annotation, to identify tissue-specific transcript isoforms, as well as to identify transcript isoforms that may be turtle-specific.

The full-length transcriptome identified by Iso-Seq data significantly improves the completeness of genome annotation. The updated annotation, including novel transcripts and genes identified from our Iso-Seq data, is available in GTF format at project's figshare repository (https://doi.org/10.6084/m9.figshare.29402339). To gener ate the Iso-Seq-based annotation, we merged transcript isoforms identified across all tissues and compared them to the NCBI reference annotation to distinguish annotated from novel isoforms (see Methods and Materials, Supplementary Figure S5). In total, a total of 164,663 consensus transcript isoforms were identified by our Iso-Seq data (Supplementary Table S8). The vast majority of these isoforms (94.7%) were assigned to NCBI-annotated genes (Fig. 4A). However, only 17,029 (10.3%) showed a splice junctions match to existing NCBI annotated transcripts (full splice match, allowing ≤ 50 bp differences at 5' and/or 3'ends), while 138,949 (84.3%) exhibited alternative splicing patterns to annotated transcripts, and were therefore classified as novel transcripts of known genes. In addition, we identified 8,685 transcript isoforms corresponding to 3,910 novel genes that are not present in the NCBI reference annotation (Supplementary Table S8). Of these novel gene isoforms, 8,056 (92.7%), which were assigned to 3,612 genes, contain a predicted ORF of at least 300 bp and were therefore classified as novel protein-coding genes. The discovery of these novel transcripts and genes expands the catalog of functional elements in the turtle genome and provides a valuable resource for future studies of their novel functions.

The number of expressed genes and transcript isoforms that we observed differ significantly between the 12 tissues that we used, likely due to extensive variation in read depth between them (Fig. 4B, Supplementary Table S8). In our Iso-Seq datasets, hatchling carapace has the largest number of expressed genes, followed by testis, whereas hatchling kidney and adult sciatic nerve have the least number of expressed genes and transcript isoforms (Supplementary Table S6). The numbers of expressed transcript isoforms per gene in our Iso-Seq data vary from 1.42 per gene in sciatic nerve to 4.51 per gene in hatchling brain (Fig. 4B). We also examined tissue-specific expression patterns of transcript isoforms (Fig. 4B). Only 258 transcript isoforms were common across all 12 tissues (Fig. 4C, Supplementary Table S8). By excluding the two tissues with extremely low sequencing depth, 1,949 transcript isoforms are expressed in the remaining 10 tissues, which likely represent housekeeping transcripts. Gene Ontology (GO) analysis suggests prominent roles for these shared transcripts in

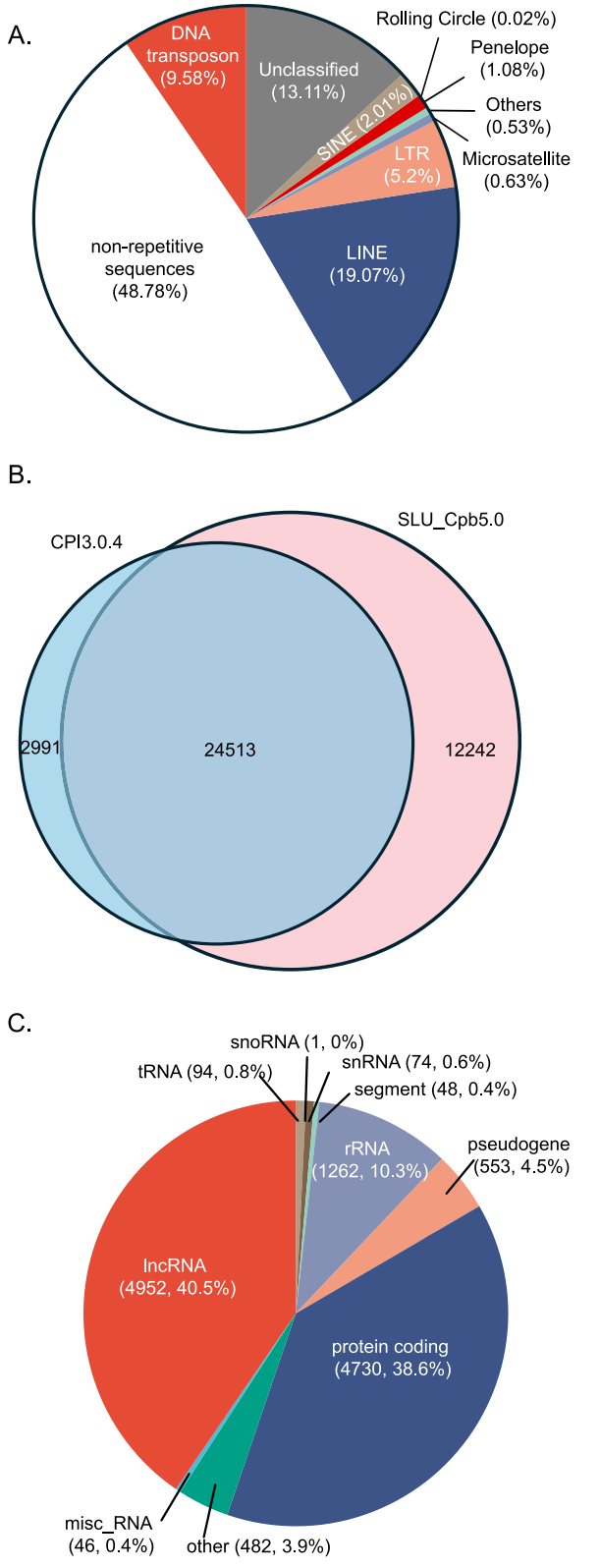

Fig. 3 Annotations of the *C. picta bellii* genome. **A** Pie chart showing the distribution of different types of repetitive elements in the *C. picta bellii* genome. **B** Venn diagram showing the differences in gene annotation between two different assemblies of the *C. picta bellii* genome (CPI3.0.4 vs. SLU_Cpb_5.0). **C** A pie chart showing the distributions of different types of genes that were annotated only in the new assembly SLU_Cpb_5.0

cell adhesion (cadherin binding), catalytic activity (ATP hydrolysis, GTPase), and translational regulation (Supplementary Figure S9A). Interestingly, several similar GO groups are enriched in tissue specific transcript isoforms in hatchling brain, hatchling carapace, and testis (Supplementary Figure S9B-D). These results suggest that genes of these functional categories tend to express multiple transcript isoforms, with some isoforms shared by all tissues, while other isoforms are expressed only in specific tissues. The shared- and tissue-specific isoforms that we have observed may reveal distinct cellular functions that are controlled by alternative splicing. These results reveal a distinct transcriptional landscape across tissues in the western painted turtle.

New hypoxia/anoxia-related transcripts

An important purpose for resequencing and reannotating the genome of the western painted turtle was to enable deeper insights into the transcriptomic basis for anoxia tolerance in this unique species. As described below, we have reanalyzed the RNA-seq dataset from our previous study of anoxic stress in telencephalon and ventricle of western painted turtle that were submerged in anoxic water for 24 h [28]. At the time of the original study, the differential expression analysis of these RNA-seq datasets used assembly CPI3.0.1 and outdated informatic tools, which enabled the identification of 19 differentially expressed genes from adult telencephalon, all of which increased and 27 differentially expressed genes in ventricle, 23 of which were upregulated and 4 downregulated [28]. Since that time, the *C. picta bellii* genome has undergone considerable improvements. A reanalysis of these original RNA-seq data using the CPI3.0.4 assembly should yield more differentially expressed genes, and our aim was to demonstrate how our new assembly, SLU_Cpb5.0, would reveal additional differentially expressed genes that are not observed with any previous assembly version.

To this end, we reanalyzed these 16 RNA-seq datasets (NCBI SRA: SRS385156-SRS385171) by using genome assemblies CPI3.0.4 and SLU_Cpb5.0 as reference. Differential expression analyses using DESeq2 [66] were performed for mapped RNA-seq reads in both assemblies. Because there was a large difference in the number of annotated gene features between the two assembles (26,230 for CPI3.0.4 vs 34,544 for SLU_Cpb5.0), we selected a raw p-value significance cut-off based on the adjusted p-values for the DE analysis conducted with the new assembly, and applied it to the p-values of the DE analysis data derived from the CPI3.0.4 assembly. This was the more conservative approach because it led to the inclusion of many genes from the old assembly analysis whose adjusted p-values were not significant. Supplementary Tables S9 and S10 provide DESeq2 results for

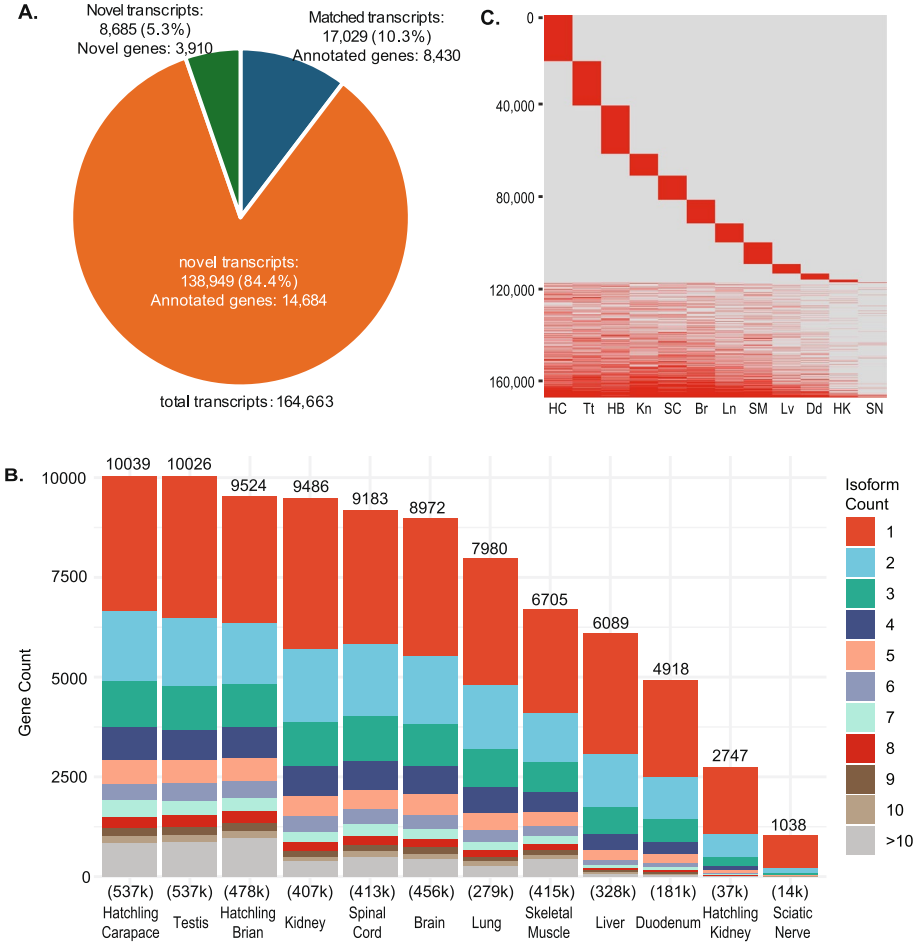

Fig. 4 Characterization of transcription isoforms across 12 tissues of western painted turtle using PacBio Iso-Seq data. **A** Distribution of transcript isoforms detected by Iso-Seq. Most transcript isoforms are not present in the NCBI genome annotation. A complete list of transcript isoforms in each category are provided in Supplementary Table S8, and a diagram of transcript types is given in Supplementary Figure S5. **B** Distribution of detected genes and the number of transcript isoforms per gene across the 12 tissues. Numbers in parentheses above tissue names indicate the total Iso-Seq read counts obtained for each tissue. **C** Distribution of transcript isoforms across the 12 tissues. Each row represents a transcript isoform; numbers of isoforms are listed on the Y-axis. Each column represents a tissue (HC: Hatchling Carapace; Tt: Testis; Kn: Kidney; HB: Hatchling Brain; SC: Spinal Cord; Br: Adult hindbrain; Ln: Lung; SM: Skeletal Muscle; Lv, Liver; Dd: Duodenum; HK: Hatchling Kidney; SN: Sciatic Nerve). Red indicates the presence of a transcript isoform in a tissue, and light grey indicates its absence

telencephalon and ventricle, using both the SLU_Cpb5.0 and CPI3.0.4 assemblies. These datasets were filtered to retain only those genes with a Padj (FDR) of < 0.05 with the SLU_Cpb5.0 assembly; the filtered results are given in Supplementary Tables S11 and S12, along with full gene names and gene types of the retained genes. A summary of differentially expressed genes that were identified using the SLU_Cpi5.0 assembly, and a comparison with the CPI3.0.4 assembly, are given in Supplementary Table 13. The 100 most abundantly expressed genes in telencephalon and ventricle that were identified using the new assembly are listed in Supplemental Tables 14 and 15.

For adult telencephalon, 165 genes were found to be significantly differentially expressed during anoxia with SLU_Cpb5.0, 69 (42%) of which were not significant with the analysis using the CPI3.0.4 (Supplementary Tables S11, S13). The expression of slightly more than half of these (33) was unquantifiable using the CPI3.0.4 assembly. The novel genes were mostly protein coding (44; 60%), 15 of which had not been quantified previously. The remaining were novel non-coding RNAs (24; 38%), most of which (18) had never been quantified before. Furthermore, a novel differentially expressed small non-coding RNA, the U2 spliceosome RNA, was discovered when these data were reanalyzed with SLU_Cpb5.0 (Fig. 5).

For adult ventricle, 270 genes were significantly differentially expressed during anoxia with SLU_Cpb5.0, 110 (41%) of which were unalignable or not differentially expressed usingCPI3.0.4 (Supplementary Tables S12, S13). More than 60% of these novel genes were found

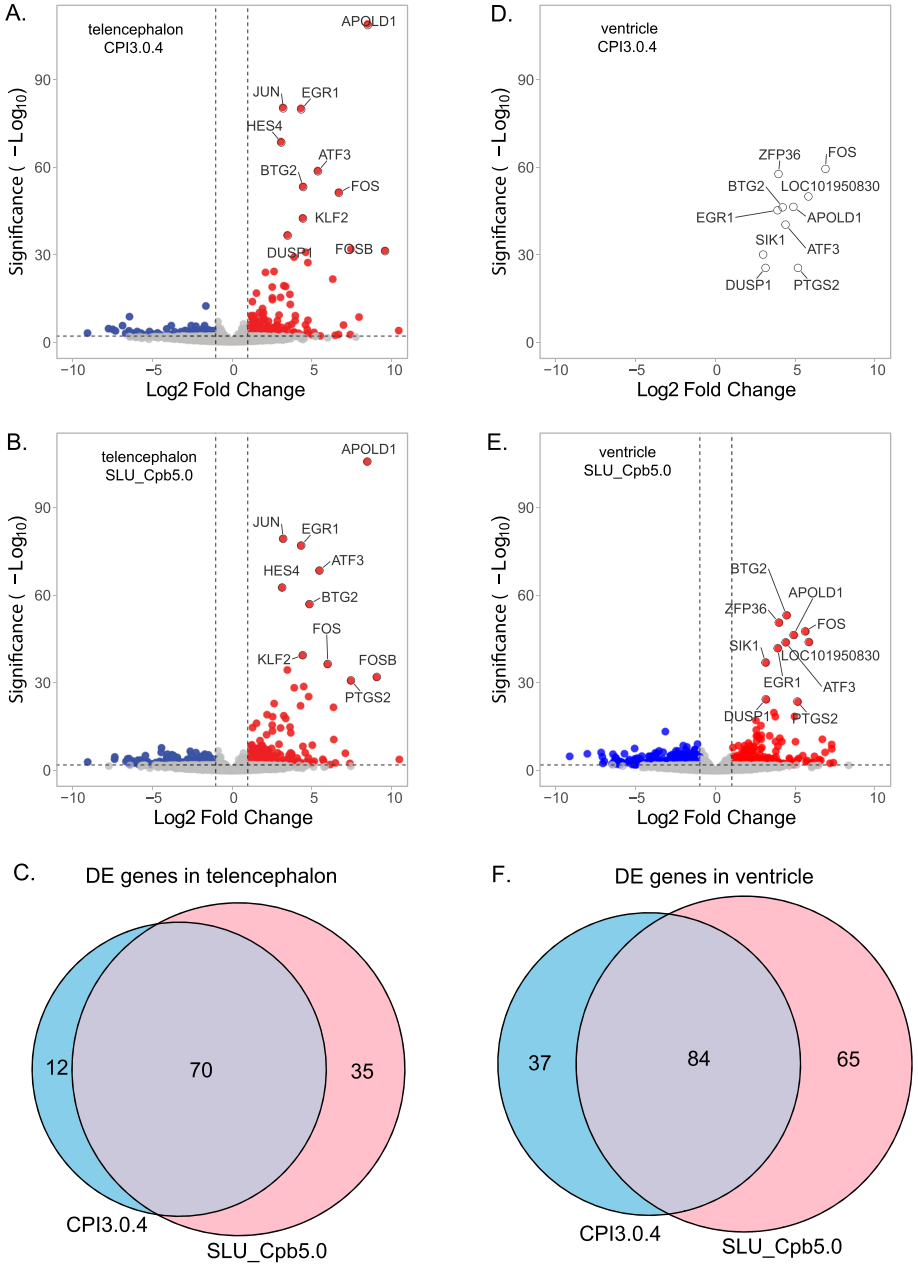

Fig. 5 Impact on new genome assembly and annotation on differential gene expression analysis using RNA-seq data. **A** Volcano plot showing differential gene expression analysis results from RNA-seq data in adult telencephalon using the CPI3.0.4 assembly. **B** Volcano plot showing differential gene expression analysis results from RNA-seq data in adult telencephalon using the SLU_Cpb5.0 assembly. **C** Venn diagram showing shared and assembly-specific DE genes in adult telencephalon. **D** Volcano plot showing differential gene expression analysis results from RNA-seq data in adult ventricle using the CPI3.0.4 assembly. **E** Volcano plot showing differential gene expression analysis results from RNA-seq data in adult ventricle using the SLU_Cpb5.0 assembly. **F** Venn diagram of differentially expressed genes in ventricle that were identified using the different assemblies. Names of the top 10 differentially expressed (DE) genes are provided in each plot of **A**-**B**, **D**-**E**

to be protein coding (66), and 28 of these had not been previously quantified. The remaining novel genes were mostly non-coding RNAs (42; 39%), most of which (38) had not been previously quantified. The U2 spliceosome RNA detected in telencephalon was also differentially expressed in ventricle. 60 DE genes are shared in the filtered telencephalon and ventricle datasets. Taken together, this analysis shows a significant improvement in the sensitivity of SLU_Cpb5.0 for detecting changes in differential gene expression in the tissues of anoxic turtles.

Discussion

We have generated a new *C. picta bellii* reference genome, SLU_Cpb5.0, that possesses several improvements over the previous reference, CPI3.0.4. First, our new assembly combined four datasets, 10 × Genomics pseudo-longreads, PacBio HiFi, Hi-C, and Bionano optical reads, to produce a chromosome-level assembly with far greater contiguity. A comparison of genome assemblies derived from short- and long-reads demonstrates that the short-read assemblies miss up to 11% of the genome [79]. Second, our new assembly was assembled from datasets that were obtained from the same individual, eliminating the possibility of errors introduced by sequence variations between individuals. Third, by anchoring BACs that have been used for FISH on our new assembly, we reconcile previous inconsistencies between painted turtle cytology and genome assemblies from multiple turtle species. Fourth, our new assembly was annotated using full-length or near full length transcripts from 12 tissues, as well as published RNA-seq datasets, providing extensive empirical evidence for annotation. The result of these efforts is one of the best assembled and annotated turtle genomes to date, and an undeniable improvement over the previous *C. picta bellii* reference assembly and annotation [80] compared a variety of genome assembly pipelines, and found no single assembly approach that lacked errors in scaffold or contig errors. Thus it is likely that some errors remain in our assembly, and no single reference genome truly reflects the variability inherent in natural populations [81]. Nonetheless, NCBI has designated our new assembly as the "reference" for *C. picta bellii* based on its improved contiguity and annotation.

Transcriptomics

A major feature of our painted turtle genome assembly is its annotation with full-length or near full-length transcript sequences obtained using PacBio Iso-Seq data from 12 tissues. In rainbow trout, another non-model species, long-read transcripts significantly improved genome annotation [82], and our Iso-Seq datasets will similarly advance the characterization of transcript diversity in turtles. Long-read sequencing reveals extensive variation in transcript start sites, splice junctions, and polyadenylation, similar to findings from the ENCODE4 RNA-seq datasets in mouse and human tissues [83]. Because splicing can be coupled to transcription [84], this process could be examined in painted turtle liver as well, for which both PacBio HiFi genomic methylation and Iso-Seq data are available. Our data provide an opportunity to assess the evolutionary conservation of transcript variation. In the painted turtle, we have identified over 138,000 novel transcript isoforms of known genes (Fig. 4A). The proteins that these isoforms encode may perform important yet-to-be discovered functions. Long-read RNA-seq data from turtle will help capture the full range of transcript structure diversity in vertebrate genes. As an example, a novel isoform of the sex determination gene DMRT1 appears to be unique to turtles [85]. Additionally, turtle isoform diversity may reveal genetic bases for painted turtles' unique physiology. Long-read isoform sequencing has uncovered tissue-specific isoform expression between active and hibernating brown bears (Ursus arctos; [86]. Moreover, alternative CDC20 isoforms regulate the duration of mitotic arrest [87], suggesting potential parallel mechanisms in regulating mitotic arrest during turtle hibernation. Our Iso-Seq datasets differ widely in the numbers of reads, and the numbers of expressed genes. Although we expected that testes would express the most genes [88], our observation of slightly more expressed genes in hatchling carapace was unexpected. Although this data is derived from only one animal, it strongly suggests that this tissue is one of the most transcriptionally active tissues in the turtle. The carapace consists of multiple different tissues, including blood, muscle, bone, connective tissue, and epidermis (scutes); additionally, its transcriptional activity could be related to the fact that the turtle shell is still mineralizing during the hatchling stage. Far fewer reads and genes were observed for sciatic nerve and hatchling kidney because the input total RNA from these tissues was much less than for the other tissues, resulting in a much lower Iso-Seq output (Supplementary Table S3). For hatchling kidney, lower expression could reflect the turtles' physiology: Hatchling painted turtles have little need for urine formation because they spend their first seven to eight months of life in their nest cavity before emerging to begin their semiaquatic lifestyle [89].

Anoxia tolerance

The Western painted turtle is the most anoxia-tolerant vertebrate known, capable of surviving for more than 30 h at room temperature [12] and 170 days at 3 °C [5, 7]. Because stroke, myocardial ischemia, and cardiac arrest produce tissue damage resulting from oxygen deprivation, understanding why some vertebrate species, such as turtles, are better than others at living without oxygen is medically important [90]. A well-annotated, chromosome-level western painted turtle genome should facilitate comparative genomics approaches to identify genes, genetic regulatory pathways, and physiological processes that enable extreme anoxia tolerance. To demonstrate the utility of this new genome (SLU_Cpb5.0) in elucidating the mechanisms of anoxia tolerance, we reanalyzed the short-read data from the first RNA-seq study of anoxia from our group [28], which involved Illumina Hi-seq paired-end sequencing of mRNAs from the telencephalon and cardiac ventricle of adult painted turtles

subjected to 24 h of anoxia at 19 °C. We have demonstrated that our further refinement of the painted turtle genome greatly improved our ability to detect differentially expressed transcripts in both tissues. Specifically, this amounted to increases in the number of detectable differentially expressed genes of 38% and 40% in telencephalon and cardiac ventricle, respectively. Many of the differentially expressed genes express multiple isoforms. We have not examined individual isoforms for differential expression in anoxia, but the observation in our datasets of genes that encode U2 spliceosome RNA and the splicing protein SFRS5, suggest the possibility of isoform-specific responses to anoxia. Notably, the differentially expressed genes that we have identified using the SLU_Cpb5.0 assembly are enriched for transcription factors and signalling proteins. These datasets are an important resource to identify the genetic mechanisms that control anoxia tolerance in painted turtles.

Because the purpose of the anoxia data reanalysis was to demonstrate the utility of the revised assembly for new discovery, a comprehensive analysis of the anoxia-responsive genes that we have identified is beyond the scope of this paper; however, several novel differentially expressed genes are noteworthy:

In telencephalon, several novel DE genes in telencephalon whose expression increases with anoxia could be related to regulation of the neuroepithelium and the blood–brain barrier (BBB). *Snai1*, which encodes Snail Family Transcriptional Repressor 1, a zinc-finger transcription factor that suppresses E-cadherin [91] and permeabilizes the BBB [92].

Snai1 transcription is increased by Hif1a [93], and we observed increased *Snai1* expression in anoxia both in telencephalon and in ventricle.

Cfap210 expression increases in anoxia; it encodes Cilia and Flagella-Associated Protein 210, a little-understood filamentous microtubule inner protein [94]. In the brain, neuroepithelial cilia circulate cerebrospinal fluid; if CFAP210 functions to stabilize microtubules within these cilia, it might preserve neuroepithelial function during anoxia.

A potentially important adaptive response in telencephalon is increased expression of the gene for Cold-Inducible RNA-binding Protein (*Cirbp*, sometimes *Cirp*). CIRBP regulates mRNA polyadenylation, stability and translation through binding on 3'-untranslated regions of target transcripts [95]. It is induced in cancer cells after exposure to cold shock, UV radiation, hypoxia, and other cellular stresses [96], and subsequently was found to be upregulated during hypothermia in mammalian organotypic hippocampal slice cultures [97]. CIRBP may be anti-apoptotic and potentially neuroprotective following ischemia due to its upregulation during mild therapeutic hypothermia [98–101]. Interestingly, CIRBP has been implicated in temperature-dependent sex determination in snapping turtles, and its allele frequencies change with latitude [102]. Therefore, we hypothesize that CIRBP may play a neuroprotective role during and following anoxia in turtles.

In ventricle, several newly discovered protein coding genes may play important roles in preserving cardiac function during and following anoxia:

Egln3 (egl-9 family hypoxia inducible factor 3), also known as prolyl hydroxylase domain-containing 3 (PHD3) is the Egln family member most abundantly expressed in heart. Egln proteins require oxygen for activity, and in normoxia, they negatively regulate HIF pathway(s) by hydroxylating HIF-1a and HIF-2b, which targets them for degradation [103, 104]. The increased expression of *Egln3* in anoxic turtle heart may compensate for its reduced activity and leads to the novel hypothesis that avoiding overactivation of HIF-dependent pathways may be a protective mechanism in the heart.

Tiparp (TCDD-inducible poly (ADP-ribose) polymerase also called Parp7, which is constitutively expressed at relatively high levels, is upregulated by anoxia. PARPs are poly-ADP-ribosyltransferases, which transfer the ADP-ribose moiety from NAD + onto target proteins [105]. PARP7 functions as an inhibitor of the interferon response to cytoplasmic nucleic acids. Recent work has characterized PARP7 as a cysteine-specific mono-ADP-ribosyltransferase that targets dozens of nuclear proteins. For example, PARP7 ADP-ribosylation inhibits TBK1 and NF-kB that induce proinflammatory cytokines, but it stabilizes FRA1(*Fosl1*), an AP-1 transcription factor that represses pro-apoptotic pathways [106, 107], and also is upregulated by anoxia (Supplementary Tables 11, 12). Thus, upregulation of PARP7 during anoxia may enhance survival by preventing inflammation and anoxia-induced apoptosis.

Urotensin-II receptor (LOC101936734, *Uts2r*, GPR14) was found to be significantly elevated during anoxia in the heart. Urotensin II is a highly conserved peptide that has strong vasoactive effects when it binds GPR14 [108]. In mammalian heart, GPR14 is localized to cardiac myocytes with little expression in the vasculature [109]. Urotensin II increases contractile force of rat ventricular papillary muscles [109] and activates sarcolemmal Na + /H + exchange in ventricular myocytes [110]. Both responses would be adaptive during anoxia by defending contractility against the negative ionotropic effects of anoxia-induced lactic acidosis.

Txnip which is constitutively expressed at high levels, was also further increased during anoxia. This gene encodes Thioredoxin Interacting Protein, which binds and inhibits the activity of thioredoxin [111]. Theoretically, the inhibition of thioredoxin should increase cellular ROS production, but because anoxic turtle heart

avoids ROS production altogether [22, 112], the role of TXNIP may serve functions other than redox balance. Targeted deletion of *Txnip* in mammalian heart showed no effect on thioredoxin or on ROS but, instead, stimulated rates of glucose uptake without changes in GLUT1 and GLUT4 expression, which implies a higher rate of glucose utilization [113]. Therefore, the role of TXNIP may be more related to adjusting the balance of glycolytic and oxidative pathways of energy production toward anaerobic glycolysis.

LOC135976089 (TRPM8 channel-associated factor 2-like) No prior RNA-seq studies of anoxic turtles showed any changes in the expression level of genes directly associated with cellular excitability or excitation contraction coupling [28, 35]. However, our reanalysis of the earlier of these studies with SLU_Cpb5.0 showed a significant decrease in the expression of a gene orthologous to mammalian TCAF2, which is a protein that promotes trafficking of TRPM8 to the membrane and negatively regulates it once it is there. TRPM8 is a cold- and menthol-sensitive calcium and sodium channel [114]. Inactivation of this channel during anoxia should decrease the excitability of cardiomyocytes, which would support cardiac functional suppression, preserve ATP, and also maintain intracellular calcium homeostasis.

Conclusions

Reliably identifying the genes that underlie traits of interest, a major focus of modern biology, requires highly contiguous and well-annotated genome assemblies. Western painted turtles are highly tolerant of anoxia, a trait with important ecological and biomedical significance. The new reference genome assembly and associated long-read transcriptome datasets are among the most complete for any turtle. The majority of painted turtle transcript isoforms are novel, potentially encoding novel functions of known genes. Using the new genome assembly, we identify candidate anoxia-related genes that were missed using previous assemblies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12683-1.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
Bob Fulton, Chad Tomlinson for 10x Genomics sequencing and assembly.
LaDeana Hillier for initial analysis of the 10x Genomics assembly.
Kaitlyn Bagwill for assistance in initial PacBio HiFi assembly.
Jim Henderson for initial BUSCO analysis.
Natasha Schneider, Nelson Membreno, and Michael/Mickey Arias for tissue dissections.
Paul Hime for advice on BAC-genome alignments.

Authors' contributions
DEW, PM, and JRBJr conceived the study. DEW provided turtles, laboratory space and supplies. JC assembled genomes, performed annotations, analyzed Iso-Seq and RNA-seq datasets, with input from JRBJr, DEW, and ZL. MK generated and analyzed Bionano optical maps. JRBJr wrote the initial manuscript draft; JRBJr, DEW, ZL and JC wrote sections, and revised the manuscript. JRBJr, DEW, and ZL provided funding. JC and JRBJr are equal contributors to this work and can reference this work as a first authorship paper in their curriculum vitae.

Funding
This work was supported by NSF grants 1253939 and 2421029 to DE Warren, NSF grant 1951332 to Zhenguo Lin, and by the John R. Bermingham Jr. donor advised fund at the Denver Foundation.

Data availability
All data is publicly available. Sequence data has been uploaded to NCBI, with accessions listed in Supplementary Table S2. Genome repetitive annotation and Iso-Seq annotation files can be find in figshare (http://doi.org/10.6084/m9.figshare.29402339).

Declarations

Ethics approval and consent to participate
All procedures involving turtles were performed in the Warren lab at Saint Louis University using protocols approved by the Saint Louis University IACUC.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 1 December 2025 / Accepted: 19 February 2026
Published online: 06 March 2026

References

1. Valenzuela N. The painted turtle, Chrysemys picta: a model system for vertebrate evolution, ecology, and human health. Cold Spring Harb Protoc. 2009;2009(7):pdb emo124.
2. Cordero GA. Re-emergence of the painted turtle (*Chrysemys picta*) as a reference species for evo-devo. Evol Dev. 2014;16(4):184–8.
3. Moustakas-Verho JE, Cebra-Thomas J, Gilbert SF. Patterning of the turtle shell. Curr Opin Genet Dev. 2017;45:124–31.
4. Alibardi L. Development, structure, and protein composition of the corneous beak in turtles. Anat Rec (Hoboken). 2021;304(12):2703–25.
5. Ultsch GR, Jackson DC. Long-term submergence at 3 degrees C of the turtle Chrysemys picta bellii in normoxic and severely hypoxic water. III. Effects of changes in ambient PO2 and subsequent air breathing. J Exp Biol. 1982;97:87–99.
6. Jackson DC. Living without oxygen: lessons from the freshwater turtle. Comp Biochem Physiol A Mol Integr Physiol. 2000;125(3):299–315.
7. Odegard DT, Sonnenfelt MA, Bledsoe JG, Keenan SW, Hill CA, Warren DE: Changes in the material properties of the shell during simulated aquatic hibernation in the anoxia-tolerant painted turtle. J Exp Biol. 2018;221(Pt 18).
8. Storey KB, Storey JM, Brooks SP, Churchill TA, Brooks RJ. Hatchling turtles survive freezing during winter hibernation. Proc Natl Acad Sci U S A. 1988;85(21):8350–4.
9. Paukstis GL, Shuman RD, Janzen FJ. Supercooling and freeze tolerance in hatchling painted turtles (*Chrysemys picta*). Can J Zool. 1989;67(4):1082–4.
10. Packard GC, Packard MJ. Hatchling painted turtles (*Chrysemys picta*) survive exposure to subzero temperatures during hibernation by avoiding freezing. J Comp Physiol B. 1993;163(2):147–52.
11. Costanzo JP, Litzgus JD, Iverson JB, Lee RE. Seasonal changes in physiology and development of cold hardiness in the hatchling painted turtle *Chrysemys picta*. J Exp Biol. 2000;203(22):3459–70.
12. Johlin JM, Moreland FB. Studies of the blood picture of the turtle after complete anoxia. J Biol Chem. 1933;103(1):107–14.

13. Starkey DE, Shaffer HB, Burke RL, Forstner MR, Iverson JB, Janzen FJ, et al. Molecular systematics, phylogeography, and the effects of Pleistocene glaciation in the painted turtle (*Chrysemys picta*) complex. Evolution. 2003;57(1):119–28.
14. Seidel ME, Ernst CH. A systematic review of the turtle family Emydidae. Vertebr Zool. 2017;67(1):1–122.
15. Ernst CH, Lovich JE. Turtles of the United States and Canada. Second edn. The Johns Hopkins University Press; 2009.
16. Barela KL, Olson DH. Mapping the Western Pond Turtle (Actinemys marmorata) and Painted Turtle (Chrysemys picta) in western North America. Northwest Nat. 2014;95(Spring):1–12.
17. Musacchia X. The viability of *Chrysemys picta* submerged at various temperatures. Physiol Zool. 1959;32(1):47–50.
18. Wasser JS, Inman KC, Arendt EA, Lawler RG, Jackson DC. ^{31}P-NMR measurements of pHi and high-energy phosphates in isolated turtle hearts during anoxia and acidosis. Am J Physiol. 1990;259(3 Pt 2):R521-530.
19. Jackson DC, Warburton SJ, Meinertz EA, Lawler RG, Wasser JS. The effect of prolonged anoxia at 3 degrees C on tissue high energy phosphates and phosphodiesters in turtles: a ^{31}P-NMR study. J Comp Physiol B. 1995;165(1):77–84.
20. Bundgaard A, James AM, Gruszczyk AV, Martin J, Murphy MP, Fago A. Metabolic adaptations during extreme anoxia in the turtle heart and their implications for ischemia-reperfusion injury. Sci Rep. 2019;9(1):2850–2850.
21. Pamenter ME, Richards MD, Buck LT. Anoxia-induced changes in reactive oxygen species and cyclic nucleotides in the painted turtle. J Comp Physiol B. 2007;177(4):473–81.
22. Bundgaard A, Gruszczyk AV, Prag HA, Williams C, McIntyre A, Ruhr IM, et al. Low production of mitochondrial reactive oxygen species after anoxia and reoxygenation in turtle hearts. J Exp Biol. 2023;226(9):jeb245516.
23. Lipton P. Ischemic cell death in brain neurons. Physiol Rev. 1999;79(4):1431–568.
24. Hochachka PW, Lutz PL. Mechanism, origin, and evolution of anoxia tolerance in animals. Comp Biochem Physiol B Biochem Mol Biol. 2001;130(4):435–59.
25. Herbert CV, Jackson DC. Temperature effects on the responses to prolonged submergence in the turtle Chrysemys picta bellii. II. Metabolic rate, blood acid-base and ionic changes, and cardiovascular function in aerated and anoxic water. Physiol Zool. 1985;56(6):670–81.
26. Daw JC, Wenger DP, Berne RM. Relationship between cardiac glycogen and tolerance to anoxia in the western painted turtle, Chrysemys picta bellii. Comp Biochem Physiol. 1967;22(1):69–73.
27. Warren DE, Jackson DC. The metabolic consequences of repeated anoxic stress in the western painted turtle, Chrysemys picta bellii. Comp Biochem Physiol A Mol Integr Physiol. 2017;203:1–8.
28. Keenan SW, Hill CA, Kandoth C, Buck LT, Warren DE. Transcriptomic responses of the heart and brain to anoxia in the western painted turtle. PLoS ONE. 2015;10(7):e0131669.
29. Radhakrishnan S, Literman R, Neuwald J, Severin A, Valenzuela N. Transcriptomic responses to environmental temperature by turtles with temperature-dependent and genotypic sex determination assessed by RNAseq inform the genetic architecture of embryonic gonadal development. PLoS ONE. 2017;12(3):e0172044.
30. Manshack LK, Conard CM, Bryan SJ, Deem SL, Holliday DK, Bivens NJ, et al. Transcriptomic alterations in the brain of painted turtles (*Chrysemys picta*) developmentally exposed to bisphenol A or ethinyl estradiol. Physiol Genomics. 2017;49(4):201–15.
31. Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, Valenzuela N, et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome Biol. 2013;14(3):R28.
32. Badenhorst D, Hillier LW, Literman R, Montiel EE, Radhakrishnan S, Shen Y, et al. Physical mapping and refinement of the painted turtle genome (*Chrysemys picta*) inform amniote genome evolution and challenge turtle-bird chromosomal conservation. Genome Biol Evol. 2015;7(7):2038–50.
33. Lee LS, Navarro-Dominguez BM, Wu Z, Montiel EE, Badenhorst D, Bista B, et al. Karyotypic evolution of sauropsid vertebrates illuminated by optical and physical mapping of the painted turtle and slider turtle genomes. Genes (Basel). 2020;11(8):928.
34. Leung SK, Jeffries AR, Castanho I, Jordan BT, Moore K, Davies JP, et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. Cell Rep. 2021;37(7):110022.
35. Fanter CE, Lin Z, Keenan SW, Janzen FJ, Mitchell TS, Warren DE: Development-specific transcriptomic profiling suggests new mechanisms for anoxic survival in the ventricle of overwintering turtles. J Exp Biol. 2020;223(Pt 4).
36. Zheng GX, Lau BT, Schnall-Levin M, Jarosz M, Bell JM, Hindson CM, Kyriazopoulou-Panagiotopoulou S, Masquelier DA, Merrill L, Terry JM et al: Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat Biotechnol. 2016;34(3):303–311.
37. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38(10):4647–54.
38. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5.
39. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. 2024;25(1):107.
40. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17(2):155–8.
41. Jackman SD, Coombe L, Chu J, Warren RL, Vandervalk BP, Yeo S, et al. Tigmint: correcting assembly errors using linked reads from large molecules. BMC Bioinformatics. 2018;19(1):393.
42. Yeo S, Coombe L, Warren RL, Chu J, Birol I. ARCS: scaffolding genome drafts with linked reads. Bioinformatics. 2018;34(5):725–31.
43. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356(6333):92–5.
44. Shelton JM, Coleman MC, Herndon N, Lu N, Lam ET, Anantharaman T, et al. Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. BMC Genomics. 2015;16:734.
45. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.
46. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11(1):1432.
47. Uliano-Silva M, Ferreira J, Krasheninnikova K, Darwin Tree of Life C, Formenti G, Abueg L, Torrance J, Myers EW, Durbin R, Blaxter M *et al*: MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics* 2023, 24(1):288.
48. Donath A, Juhling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47(20):10543–52.
49. Thibaud-Nissen F, DiCuccio M, Hlavina W, Kimchi A, Kitts PA, Murphy TD, et al. P8008 the NCBI eukaryotic genome annotation pipeline. J Animal Sci. 2016;94(suppl_4):184–184.
50. Morgulis A, Gertz EM, Schaffer AA, Agarwala R. WindowMasker: window-based masker for sequenced genomes. Bioinformatics. 2006;22(2):134–41.
51. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
52. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.
53. Kapustin Y, Souvorov A, Tatusova T, Lipman D. Splign: algorithms for computing spliced alignments with identification of paralogs. Biol Direct. 2008;3:20.
54. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933–5.
55. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 2018;46(D1):D335–42.
56. Chan PP, Lin BY, Mak AJ, Lowe TM. TRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96.
57. Baril T, Galbraith J, Hayward A. Earl Grey: a fully automated user-friendly transposable element annotation and analysis pipeline. Mol Biol Evol. 2024;41(4):msae068.
58. Kuo RI, Cheng Y, Zhang R, Brown JWS, Smith J, Archibald AL, et al. Illuminating the dark side of the human transcriptome with long read transcript sequencing. BMC Genomics. 2020;21(1):751.
59. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.

60. Pardo-Palacios FJ, Arzalluz-Luque A, Kondratova L, Salguero P, Mestre-Tomas J, Amorin R, et al. SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. Nat Methods. 2024;21(5):793–7.
61. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. Clusterprofiler 4.0: a universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141.
62. Team RC. R: a language and environment for statistical computing. 4.2.0 edn. Vienna, Austria: R Foundation for Statistical Computing; 2024.
63. Carlson M: org.Hs.eg.db: Genome wide annotation for Human. In., R package version 3.22.0 edn: Bioconductor; 2024.
64. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
65. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
66. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
67. Blighe K, Rana S, Lewis M: EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. In.: Bioconductor; 2025.
68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. Genome project data processing S: The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
69. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019;4(43):1686.
70. Xiao N: ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'. In.; 2025.
71. Simison WB, Parham JF, Papenfuss TJ, Lam AW, Henderson JB. An annotated chromosome-level reference genome of the red-eared slider turtle (*Trachemys scripta elegans*). Genome Biol Evol. 2020;12(4):456–62.
72. Salzberg SL, Yorke JA. Beware of mis-assembled genomes. Bioinformatics. 2005;21(24):4320–1.
73. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
74. He W, Yang J, Jing Y, Xu L, Yu K, Fang X. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics. 2023;39(3):btad121.
75. Bentley BP, Carrasco-Valenzuela T, Ramos EKS, Pawar H, Souza Arantes L, Alexander A, et al. Divergent sensory and immune gene evolution in sea turtles with contrasting demographic and life histories. Proc Natl Acad Sci U S A. 2023;120(7):e2201076120.
76. Ren Y, Zhang Q, Yan X, Hou D, Huang H, Li C, et al. Genomic insights into the evolution of the critically endangered soft-shelled turtle *Rafetus swinhoei*. Mol Ecol Resour. 2022;22(5):1972–85.
77. Bista B, Gonzalez-Rodelas L, Alvarez-Gonzalez L, Wu ZQ, Montiel EE, Lee LS, et al. De novo genome assemblies of two cryptodiran turtles with ZZ/ZW and XX/XY sex chromosomes provide insights into patterns of genome reshuffling and uncover novel 3D genome folding in amniotes. Genome Res. 2024;34(10):1553–69.
78. Lee Y, Rio DC. Mechanisms and regulation of alternative pre-mRNA splicing. Annu Rev Biochem. 2015;84:291–323.
79. Kim J, Lee C, Ko BJ, Yoo DA, Won S, Phillippy AM, et al. False gene and chromosome losses in genome assemblies caused by GC content variation and repeats. Genome Biol. 2022;23(1):204.
80. Jarvis ED, Formenti G, Rhie A, Guarracino A, Yang C, Wood J, et al. Semi-automated assembly of high-quality diploid human reference genomes. Nature. 2022;611(7936):519–31.
81. Eisenstein M. Every base everywhere all at once: pangenomics comes of age. Nature. 2023;616(7957):618–20.
82. Ali A, Thorgaard GH, Salem M. PacBio iso-seq improves the rainbow trout genome annotation and identifies alternative splicing associated with economically important phenotypes. Front Genet. 2021;12:683408.
83. Reese F, Williams B, Balderrama-Gutierrez G, Wyman D, Celik MH, Rebboah E, Rezaie N, Trout D, Razavi-Mohseni M, Jiang Y et al: The ENCODE4 long-read RNA-seq collection reveals distinct classes of transcript structure diversity. bioRxiv 2023.
84. Shenasa H, Bentley DL. Pre-mRNA splicing and its cotranscriptional connections. Trends Genet. 2023;39(9):672–85.
85. Mizoguchi B, Valenzuela N. Alternative splicing and thermosensitive expression of Dmrt1 during urogenital development in the painted turtle, *Chrysemys picta*. PeerJ. 2020;8:e8639.
86. Tseng E, Underwood JG, Evans Hutzenbiler BD, Trojahn S, Kingham B, Shevchenko O, Bernberg E, Vierra M, Robbins CT, Jansen HT et al: Long-read isoform sequencing reveals tissue-specific isoform expression between active and hibernating brown bears (Ursus arctos). G3 (Bethesda). 2022;12(3).
87. Tsang MJ, Cheeseman IM. Alternative CDC20 translational isoforms tune mitotic arrest duration. Nature. 2023;617(7959):154–61.
88. Lin S, Lin Y, Nery JR, Urich MA, Breschi A, Davis CA, et al. Comparison of the transcriptional landscapes between human and mouse tissues. Proc Natl Acad Sci U S A. 2014;111(48):17224–9.
89. Ultsch GR. The ecology of overwintering among turtles: where turtles overwinter and its consequences. Biol Rev Camb Philos Soc. 2006;81(3):339–67.
90. Bickler PE. Clinical perspectives: neuroprotection lessons from hypoxia-tolerant organisms. J Exp Biol. 2004;207(Pt 18):3243–9.
91. Batlle E, Sancho E, Franci C, Dominguez D, Monfar M, Baulida J, et al. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol. 2000;2(2):84–9.
92. Kim BJ, Hancock BM, Bermudez A, Del Cid N, Reyes E, van Sorge NM, et al. Bacterial induction of Snail1 contributes to blood-brain barrier disruption. J Clin Invest. 2015;125(6):2473–83.
93. Zhang L, Huang G, Li X, Zhang Y, Jiang Y, Shen J, et al. Hypoxia induces epithelial-mesenchymal transition via activation of SNAI1 by hypoxia-inducible factor -1alpha in hepatocellular carcinoma. BMC Cancer. 2013;13:108.
94. Gui M, Croft JT, Zabeo D, Acharya V, Kollman JM, Burgoyne T, et al. SPACA9 is a lumenal protein of human ciliary singlet and doublet microtubules. Proc Natl Acad Sci U S A. 2022;119(41):e2207605119.
95. Corre M, Lebreton A. Regulation of cold-inducible RNA-binding protein (CIRBP) in response to cellular stresses. Biochimie. 2024;217:3–9.
96. Rana S, Jogi MK, Choudhary S, Thakur R, Sahoo GC, Joshi V. Unraveling the intricacies of cold-inducible RNA-binding protein: a comprehensive review. Cell Stress Chaperones. 2024;29(4):615–25.
97. Tong G, Endersfelder S, Rosenthal LM, Wollersheim S, Sauer IM, Buhrer C, et al. Effects of moderate and deep hypothermia on RNA-binding proteins RBM3 and CIRP expressions in murine hippocampal brain slices. Brain Res. 2013;1504:74–84.
98. Wu L, Sun HL, Gao Y, Hui KL, Xu MM, Zhong H, et al. Therapeutic hypothermia enhances cold-inducible RNA-binding protein expression and inhibits mitochondrial apoptosis in a rat model of cardiac arrest. Mol Neurobiol. 2017;54(4):2697–705.
99. Saito K, Fukuda N, Matsumoto T, Iribe Y, Tsunemi A, Kazama T, et al. Moderate low temperature preserves the stemness of neural stem cells and suppresses apoptosis of the cells via activation of the cold-inducible RNA binding protein. Brain Res. 2010;1358:20–9.
100. Li S, Zhang Z, Xue J, Liu A, Zhang H. Cold-inducible RNA binding protein inhibits H(2)O(2)-induced apoptosis in rat cortical neurons. Brain Res. 2012;1441:47–52.
101. Zhang HT, Xue JH, Zhang ZW, Kong HB, Liu AJ, Li SC, et al. Cold-inducible RNA-binding protein inhibits neuron apoptosis through the suppression of mitochondrial apoptosis. Brain Res. 2015;1622:474–83.
102. Schroeder AL, Metzger KJ, Miller A, Rhen T. A novel candidate gene for temperature-dependent sex determination in the common snapping turtle. Genetics. 2016;203(1):557–71.
103. Henze AT, Riedel J, Diem T, Wenner J, Flamme I, Pouyseggur J, et al. Prolyl hydroxylases 2 and 3 act in gliomas as protective negative feedback regulators of hypoxia-inducible factors. Cancer Res. 2010;70(1):357–66.
104. Strocchi S, Reggiani F, Gobbi G, Ciarrocchi A, Sancisi V. The multifaceted role of EGLN family prolyl hydroxylases in cancer: going beyond HIF regulation. Oncogene. 2022;41(29):3665–79.
105. Bock FJ, Todorova TT, Chang P. RNA regulation by poly(ADP-ribose) polymerases. Mol Cell. 2015;58(6):959–69.
106. Manetsch P, Bohi F, Nowak K, Leslie Pedrioli DM, Hottiger MO. PARP7-mediated ADP-ribosylation of FRA1 promotes cancer cell growth by repressing IRF1- and IRF3-dependent apoptosis. Proc Natl Acad Sci U S A. 2023;120(49):e2309047120.
107. Manetsch P, Hottiger MO. Unleashing viral mimicry: a combinatorial strategy to enhance the efficacy of PARP7 inhibitors. BioEssays. 2025;47(2):e2400087.
108. Ross B, McKendy K, Giaid A. Role of urotensin II in health and disease. Am J Physiol Regul Integr Comp Physiol. 2010;298(5):R1156-1172.
109. Gong H, Wang YX, Zhu YZ, Wang WW, Wang MJ, Yao T, et al. Cellular distribution of GPR14 and the positive inotropic role of urotensin II in the myocardium in adult rat. J Appl Physiol (1985). 2004;97(6):2228–35.

110. Kato K, Yasutake M, Jia D, Snabaitis AK, Avkiran M, Kusama Y, et al. Urotensin II activates sarcolemmal Na+/H+ exchanger in adult rat ventricular myocytes. J Cardiovasc Pharmacol. 2010;55(2):191–7.
111. Wang BF, Yoshioka J. The emerging role of thioredoxin-interacting protein in myocardial ischemia/reperfusion injury. J Cardiovasc Pharmacol Ther. 2017;22(3):219–29.
112. Ruhr IM, McCourty H, Bajjig A, Crossley DA, Shiels HA, Galli GLJ. Developmental plasticity of cardiac anoxia-tolerance in juvenile common snapping turtles (*Chelydra serpentina*). Proc Biol Sci. 2019;286(1905):20191072.
113. Yoshioka J, Imahashi K, Gabel SA, Chutkow WA, Burds AA, Gannon J, et al. Targeted deletion of thioredoxin-interacting protein regulates cardiac dysfunction in response to pressure overload. Circ Res. 2007;101(12):1328–38.
114. Gkika D, Lemonnier L, Shapovalov G, Gordienko D, Poux C, Bernardini M, et al. TRP channel-associated factors are a novel protein family that regulates TRPM8 trafficking and activity. J Cell Biol. 2015;208(1):89–107.

Publisher's note