UC Santa Cruz
UC Santa Cruz Previously Published Works

Title

Long-read isoform sequencing reveals tissue-specific isoform expression between active and hibernating brown bears (Ursus arctos)

Permalink

https://escholarship.org/uc/item/9wq4h73j

Journal

G3: Genes, Genomes, Genetics, 12(3)

ISSN

2160-1836

Authors

Tseng, Elizabeth
Underwood, Jason G
Hutzenbiler, Brandon D Evans
et al.

Publication Date

2022-03-04

DOI

10.1093/g3journal/jkab422

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

Long-read isoform sequencing reveals tissue-specific isoform expression between active and hibernating brown bears (*Ursus arctos*)

Elizabeth Tseng,[1] Jason G. Underwood,[1] Brandon D. Evans Hutzenbiler,[2,3] Shawn Trojahn,[4] Brewster Kingham (iD) ,[5]
Olga Shevchenko,[5] Erin Bernberg,[5] Michelle Vierra,[1] Charles T. Robbins,[3,4] Heiko T. Jansen,[2] and Joanna L. Kelley (iD) [4,*]

[1]Pacific Biosciences, Menlo Park, CA 94025, USA,
[2]Department of Integrative Physiology and Neuroscience, Washington State University, Pullman, WA 99164, USA,
[3]School of the Environment, Washington State University, Pullman, WA 99164, USA,
[4]School of Biological Sciences, Washington State University, Pullman, WA 99164, USA, and
[5]Sequencing & Genotyping Center, Delaware Biotechnology Institute, University of Delaware, Newark, DE 19711, USA

*Corresponding author: School of Biological Sciences, Washington State University, 100 Dairy Road, Pullman, WA, 99164, USA. Email: joanna.l.kelley@wsu.edu

Abstract

Understanding hibernation in brown bears (*Ursus arctos*) can provide insight into some human diseases. During hibernation, brown bears experience periods of insulin resistance, physical inactivity, extreme bradycardia, obesity, and the absence of urine production. These states closely mimic aspects of human diseases such as type 2 diabetes, muscle atrophy, as well as renal and heart failure. The reversibility of these states from hibernation to active season enables the identification of mediators with possible therapeutic value for humans. Recent studies have identified genes and pathways that are differentially expressed between active and hibernation seasons in bears. However, little is known about the role of differential expression of gene isoforms on hibernation physiology. To identify both distinct and novel mRNA isoforms, full-length RNA-sequencing (Iso-Seq) was performed on adipose, skeletal muscle, and liver from three individual bears sampled during both active and hibernation seasons. The existing reference genome annotation was improved by combining it with the Iso-Seq data. Short-read RNA-sequencing data from six individuals were mapped to the new reference annotation to quantify differential isoform usage (DIU) between tissues and seasons. We identified differentially expressed isoforms in all three tissues, to varying degrees. Adipose had a high level of DIU with isoform switching, regardless of whether the genes were differentially expressed. Our analyses revealed that DIU, even in the absence of differential gene expression, is an important mechanism for modulating genes during hibernation. These findings demonstrate the value of isoform expression studies and will serve as the basis for deeper exploration into hibernation biology.

Keywords: hibernation; brown bear; transcriptomics; Iso-Seq; alternative splicing; differential isoform expression; full-length transcript sequencing

Introduction

Hibernation in bears has long been championed as a promising system for understanding the extremes of mammalian physiology and for identifying novel therapeutic targets (Martin 2008; Stenvinkel *et al.* 2013). Annual hibernation in brown bears (*Ursus arctos*) involves massive physiological shifts to conserve energy during the food-scarce winter (Hellgren 1998), and every organ system in bears demonstrates a suite of adaptations driven by the needs of hibernation. Hibernating bears exhibit certain phenotypes present in human disease, but these phenotypes do not themselves negatively impact the bears' overall health (Fedorov *et al.* 2009; Berg von Linde *et al.* 2015; Welinder *et al.* 2016; Rigano *et al.* 2017). For example, heart rate slows to 10 to 15 beats per minute (Nelson and Robbins 2010), yet bears do not develop typical dysfunction that would be characteristic of severe bradycardia in humans. Bears prevent congestive heart failure or

ventricular dilation by decreasing atrial contractility and increasing atrial and ventricular stiffness (Nelson and Robbins 2010; Barrows *et al.* 2011). Hibernating bears are also well known for maintaining muscle strength, morphology, and composition during hibernation in the near complete absence of weight-bearing activity (Harlow *et al.* 2001; Hershey *et al.* 2008). Bears also exhibit insulin resistance during hibernation but are insulin sensitive during the active season (Rigano *et al.* 2017). Although humans do not hibernate, the unique metabolic adaptations that evolved in hibernators could provide clues to develop new treatments for human metabolic diseases, such as obesity and type 2 diabetes (Martin 2008). Recent studies have shown that these hibernation-induced physiological shifts are associated with massive changes in the regulation and expression of thousands of genes across hibernation-relevant processes (Fedorov *et al.* 2011, 2014; Jansen *et al.* 2019). Notably, many of these genes are involved in complex

metabolic and cellular signaling pathways (*e.g.*, insulin signaling and metabolism) that play critical roles in a variety of biological processes across vertebrates, including humans (Cheatham and Kahn 1995; Jansen *et al.* 2019).

Over 10,000 genes are differentially regulated in adipose, liver, and muscle tissues between active and hibernating states, which provides a set of candidate genes involved in the regulation of cellular and physiological processes that underlie the metabolic suppression observed in hibernation (Jansen *et al.* 2019). These genes represent key genes and genomic regions for testing hypotheses related to the evolution and regulation of hibernation. To date, studies in bears have focused on global gene expression levels only, even though mRNA isoforms vary between tissues, cell types, and developmental stages (Wang *et al.* 2008; Zhang *et al.* 2016) and play a role in cellular processes. The mRNA processing shifts that result in different isoforms and thus proteome output may differ between active and hibernating seasons.

For this study, we aimed to determine whether different transcript isoforms are differentially expressed between active and hibernation states. Isoform differences have been explored only in a few select cases in bears. For example, in brown bears and Himalayan black bears (*Ursus thibetanus ussuricus*), the amount of titin does not differ between active and hibernation seasons but the relative abundance of two prominent isoforms changes between seasons and has been proposed to explain increased ventricular stiffness during hibernation (Nelson *et al.* 2008; Salmov *et al.* 2015). Yet little is known about how different isoforms contribute to the hibernation phenotype on a large scale. Because of its capability in sequencing full-length RNA transcripts, SMRT sequencing is ideal for identifying isoforms that are differentially expressed between seasons. We compare full-length isoforms between hibernating and active bears in three metabolically active tissues—skeletal muscle, liver, and adipose. We also generate and use an improved full-length isoform reference together with existing short-read RNA-seq data from the same tissues and activity states to quantify gene expression profiles more accurately with isoform-level resolution.

Methods
Sample collection
For the PacBio Iso-Seq protocol, samples were collected from three bears (one female, two males) at the Washington State University Bear Research, Education, and Conservation Center in January 2019 and May 2019, during the winter hibernation and summer active periods, respectively. Muscle, liver, and adipose tissue samples were collected for a total of 18 samples. Samples for the Illumina Ribo-Zero RNA-seq data were collected in May 2015 and January 2016 and are described in detail in Jansen *et al.* (2019). Animal care details are described in Rigano *et al.* (2017). Bears were first anesthetized using the protocol described in Ware *et al.* (2012). Subcutaneous adipose samples were collected using a 6 mm punch biopsy (Miltex, York, PA) as described in Rigano *et al.* (2017), while skeletal muscle (gastrocnemius) and liver tissue samples were collected with a 14G tru-cut biopsy needle (Progressive Medical International, Vista, CA, USA). All samples were collected between 0800 and 1200 h. Samples were immediately flash frozen in liquid nitrogen and transferred to a – 80°C freezer, where they were stored until shipment to the University of Delaware Sequencing & Genotyping Center for RNA extraction. Procedures for all experiments were approved by the Institutional Animal Care and Use Committee at Washington State University (Protocol #04922).

PacBio Iso-Seq library preparation and sequencing
Total RNAs were extracted from tissue samples and isolated using the RNeasy Universal kit (Qiagen, Valencia, CA, USA) as per the manufacturer's standard protocol. Following total RNA isolation, the samples were concentrated using RNA Clean & Concentrator Kit (Zymo Research, Irvine CA, USA). The purity of RNA samples was measured using the DeNovix DS-11+ spectrophotometer (DeNovix Inc., Wilmington, DE, USA). RNA concentration was measured using Qubit High Sensitivity RNA Assay Kit and Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The integrity of total RNA was assessed on the Agilent Fragment Analyzer 5200 system (Agilent Technologies, Santa Clara, CA, USA) using the High Sensitivity RNA Kit. The RNA Quality Number (RQN) criteria for the RNA samples was RQN >7.0.

From 100 to 300 ng of total RNA was input for cDNA synthesis and amplification using NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (New England BioLabs Inc., Ipswich, MA, USA) as per the manufacturer's standard protocol. This is a poly-A selection library preparation method. A total of 10–15 PCR cycles were used to generate sufficient quantities of cDNA for PacBio Iso-Seq library preparations. Concentration and size profile of cDNA samples was assessed on the Agilent Fragment Analyzer 5200 system (Agilent Technologies, Santa Clara, CA, USA) using the High Sensitivity Large Fragment Kit.

Amplified cDNA samples were size selected using ProNex Size-Selective Purification System (Promega Corporation, Madison, WI, USA) as per the PacBio recommendation for standard length cDNA transcripts. Size selected cDNA was used to construct SMRTbell Iso-Seq libraries using Express Template Prep 2.0 (Pacific Biosciences, Menlo Park, CA, USA) as per the manufacturer's Iso-Seq Express Template Preparation protocol. The concentration of the Iso-Seq libraries was measured using the Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The fragment size profile of the Iso-Seq libraries was assessed on the (Agilent Technologies, Santa Clara, CA, USA). Each Iso-Seq library was run on a single Sequel system SMRT Cell using sequencing chemistry 3.0 with 4 h pre-extension and 20 h movie time. One SMRT Cell per tissue and state was used to provide deep coverage of the entire grizzly transcriptome. Raw reads were processed into circular consensus sequence (CCS) reads as per the manufacturer's standard pipeline (SMRT Link version 7.0).

PacBio Iso-Seq bioinformatic analysis
All 18 SMRT Cells of Iso-Seq data from different samples were pooled and run through the Iso-Seq Analysis in SMRTLink v8.1 which generated full-length, high-quality (HQ) isoform sequences. The HQ isoforms were mapped to the genome assembly (GCA_003584765.1) (Taylor *et al.* 2018) using minimap2, then filtered and collapsed into nonredundant isoforms using Cupcake following the analysis described at (https://github.com/Magdoll/cDNA_Cupcake/wiki/Cupcake:-supporting-scripts-for-Iso-Seq-after-clustering-step). The nonredundant isoforms were then classified against the GCA_003584765.1 reference transcriptome using SQANTI3 (https://github.com/ConesaLab/SQANTI3). After running SQANTI3 classification and filtering, we obtained a final set of PacBio isoforms that we used subsequently as the reference transcriptome for short read quantification. As part of the Cupcake processing pipeline, we obtained full-length read counts associated

with each isoform, which were then normalized into full-length counts per million (cpm) for cross-sample comparison.

Short-read RNA-seq quantification

Short-read Illumina data from Jansen *et al.* (2019) for the same individuals and tissues, sampled in a different year, were mapped to the final set of PacBio isoforms using the Kallisto quantification algorithm, with the rf-stranded flag (Bray *et al.* 2016). Read counts were compared to the Iso-Seq data using Pearson correlation on the $\log_2$ counts per million.

Merging annotations and mapping

The PacBio-only annotation set was merged with the reference-only annotations using gffcompare (Pertea and Pertea 2020). Short read Illumina data from Jansen *et al.* (2019) was mapped to each transcriptome and the reference genome [GenBank assembly accession: GCA_003584765.1 (Taylor *et al.* 2018)] using HISAT2 (version 2.2.1) with the –rf flag and otherwise default parameters (Kim *et al.* 2019).

Functional annotation and differential isoform expression analysis

Abundance count estimates for each individual were combined into a single input matrix for tappAS (version 1.0.7) (de la Fuente *et al.* 2020). The annotation file generated in SQANTI3 and the short-read count matrices for each tissue were input into tappAS (de la Fuente *et al.* 2020). Each tissue was analyzed separately to compare differential isoform expression in hibernation compared to active season. We excluded short-read data from sample 1AA because it was previously shown to contain a hair follicle (Jansen *et al.* 2019). Transcripts with cpm less than 1.0 or coefficient of variation cutoff of 100% were excluded from the analyses. In the differential isoform usage (DIU) analysis, minor isoforms with a proportional of expression difference less than 0.1 were excluded from the analysis and DIU was considered at an FDR < 0.05. Gene ontology (GO) enrichment of different gene sets was calculated using PANTHER (Mi *et al.* 2021) with the following parameters: Analysis Type: PANTHER Overrepresentation Test (Released 20210224), Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.5228828 Released 2021-08-18, Reference List: Homo sapiens (all genes in database), Annotation Data Set: GO biological process, Test Type: Fisher's Exact, Correction: Calculate False Discovery Rate.

Results

High correlation between long- and short-read data at gene level

We analyzed RNA-sequencing data from bears in active or hibernation season on three distinct tissue types (muscle, adipose, liver). Three of the bears were sequenced using the PacBio Iso-Seq protocol for full length RNA transcripts, and all six bears were sequenced using the short-read Illumina RNA-seq approach (Figure 1). The RNA-seq data was previously analyzed in Jansen *et al.* (2019). We first combined all the long-read Iso-Seq data across samples and replicates and obtained a total of 6.1 million full-length HiFi reads (Supplementary Table S1). After running the HiFi reads through analysis, mapping to the reference genome, and filtering for library artifacts, we obtained 76,071 unique, full-length isoforms ranging from 150 basepairs (bp) to 16.5 kilobases (kb) (mean: 3.2 kb). We then evaluated the gene-level correlation of long- *vs* short-read data. For this correlation, we used only transcripts that were present in both Iso-Seq and

RNA-seq datasets. When comparing data from the same individuals (albeit sampled in different years), gene-level correlations were high within sequencing platforms (mean Pearson correlation RNA-seq: 0.684, Iso-Seq: 0.546 and between platforms: 0.397) (Figure 2A). Correlations were high among tissues (mean Pearson correlations RNA-seq adipose = 0.934, liver = 0.954, muscle = 0.952; Iso-Seq adipose = 0.715, liver = 0.659, muscle = 0.712; between platforms adipose = 0.551, liver = 0.535, muscle = 0.576; Figure 2A). The lower concordance at the isoform-level within the Iso-Seq samples as compared to the within RNA-seq samples, is likely explained by the lower sequencing coverage of the long-read dataset (Figure 2B). Nevertheless, we see consistent gene-level correlation for the matching samples across data types, while samples from the same tissues or animals have higher correlation than different tissues, as expected.

An improved reference transcriptome and full-length isoform annotation using Iso-Seq

The existing reference transcriptome contained 30,263 genes encompassing 58,335 transcripts (Taylor *et al.* 2018). The Iso-Seq transcriptome was classified against the reference transcriptome and found to contain 12,018 known and 907 novel genes (Supplementary Table S2). Compared with the reference annotation, 27.8% of the Iso-Seq isoforms were categorized as full-splice matches (FSMs; perfect matches to a reference transcript), while over half of the isoforms were novel isoforms (NIC, novel in catalog or NNC, novel not in catalog). More than 30% of the genes had complex splicing events (greater than six isoforms). Novel isoforms had a higher proportion of having a predicted nonsense mediated decay (NMD) effect (see the left panel of Figure 3, A–D).

We merged the existing reference transcriptome with the new Iso-Seq transcriptome data, resulting in a total of 31,829 genes encompassing 107,649 transcripts. The merged data set had a reduced number of incomplete splice matches (ISM) and novel isoforms (NIC and NNC) while greatly increasing the number of known transcripts (Supplementary Table S3), suggesting a more comprehensive representation of the transcriptome. When analyzing transcripts expressed in each tissue, by mapping RNA-seq data from six bears, we found that each tissue (muscle, liver, and adipose) had different distributions of transcript structural changes, with adipose and liver having similar distributions and muscle having a much larger number of FSM and fewer NIC transcripts (Figure 3). Importantly, the improved reference transcriptome, with the full-length transcripts originating from samples of interest, provides a starting point for the discovery of DIU that would otherwise be missed. One example of such a finding is the isoform expression of the CA2 gene (Figure 3E), described in more detail in later sections.

To evaluate the degree to which the Iso-Seq dataset captures the expressed transcripts in the tissues of interest, we mapped the ribosomal RNA-depleted (Ribo-Zero) short-read RNA-seq data of the same tissues (adipose, liver, and muscle) and seasons (active and hibernation) to either a PacBio-only transcriptome, the reference-only transcriptome, a PacBio-reference merged transcriptome, or the reference genome (Figure 4). Given the lower sequencing depth of the PacBio data, we were not surprised to see a higher mappability of the short reads to the reference transcriptome than to the PacBio-only transcriptome. However, the merged PacBio-reference transcriptome showed the highest mappability of all transcriptomes. Of note, in the adipose and liver tissues, a higher proportion of intronic reads in the ribosomal-depleted RNA-seq data resulted in a much higher mappability using the reference genome. With the demonstrated
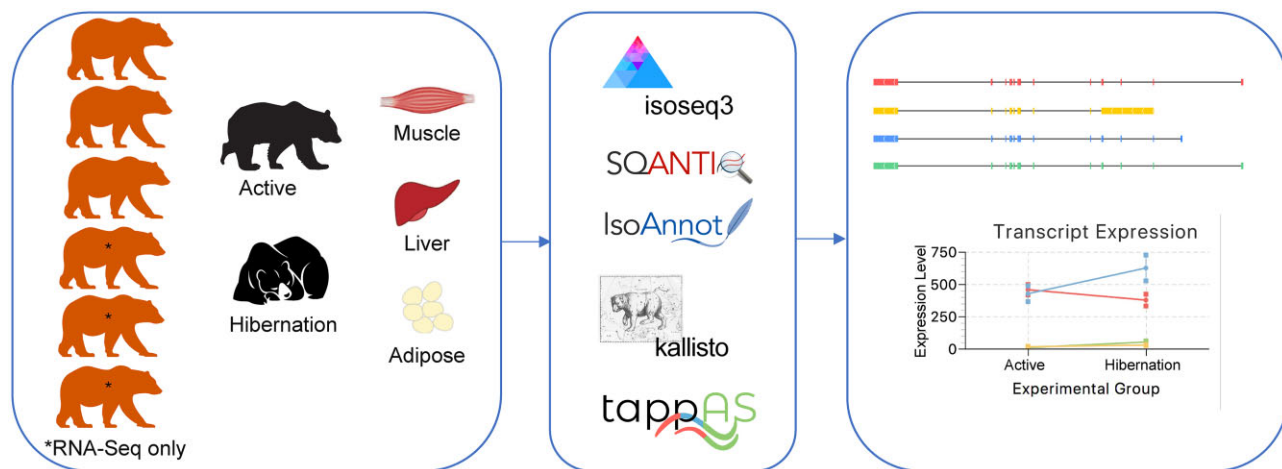
Figure 1 Bear transcriptome workflow. For each of the six bears, tissues (muscle, liver, and adipose) were extracted during active or hibernation seasons. PacBio Iso-Seq data were collected from three bears and Illumina RNA-seq data were collected from six bears. Iso-Seq data were processed through a pipeline of isoseq3, SQANTI3, and IsoAnnot, before merging with the reference transcriptome to create a merged annotation set to map RNA-seq data using Kallisto. Differential isoform expression and splice junction usage were determined using tappAS.
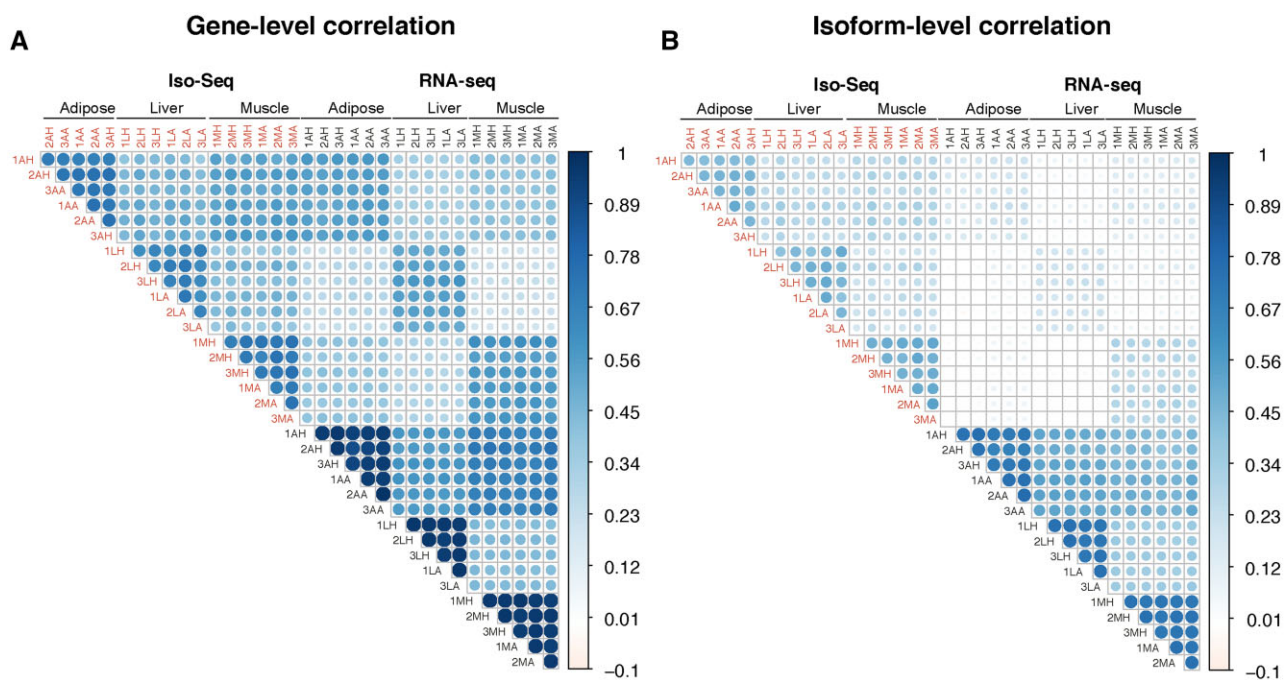

Figure 2 Gene- and isoform-level correlation between Iso-Seq and RNA-seq count data (log cpm). Correlation measured with Pearson correlation. Samples are coded using three letters representing [animal], [tissue], and [season]. The three bears were numbered. The tissues were adipose (A), liver (L), and muscle (M). The seasons were hibernation (H) or active (A).

improvement of the transcriptome by adding the Iso-Seq data, the merged (PacBio and reference) transcriptome was used for further analyses.

Tissue-specific DIU from active to hibernation season

We assessed whether the relative abundance of different isoforms for each gene varied between the seasons (DIU). We also determined whether the major (highest expressed) isoform switched between seasons. When analyzing for DIU and major isoform switching between the seasons (hibernation *vs* active), there were substantial differences among the tissues (Table 1, Supplementary Figure S2). Adipose had the highest incidence of DIU with regards to both number and % of all analyzed genes,

with and without major isoform switching (27.5% of genes; Table 1). In contrast, major isoform switching between states without DIU was fairly consistent across tissues.

Isoform Switching between active and hibernation seasons despite no gene-level expression change

In this study and in prior work, gene expression changes were detected in all three tissues between active and hibernating seasons. With the new isoform-level quantifications, we examined whether the genes classified as having both DIU and Major Isoform Switching between the two seasons (Table 1) were enriched in specific cellular functions. Of the three tissues, adipose displayed the highest number of DIU + Major Isoform
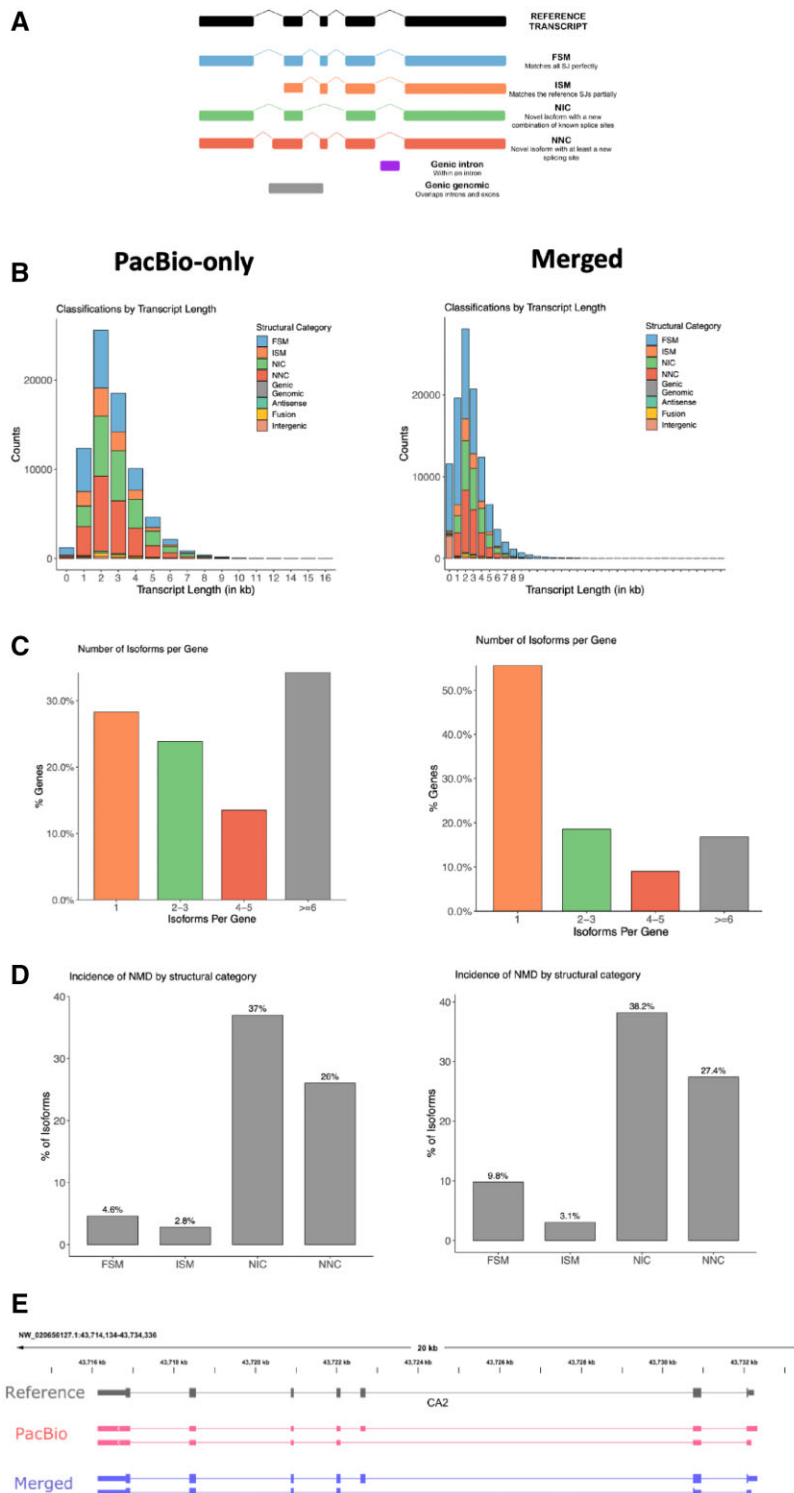
Figure 3 Isoform Classification (A) SQANTI3 isoform classification explanation, FSM, full-splice match; ISM, incomplete-splice match; NIC, novel in catalog; NNC, novel not in catalog. Genic Genomic = overlaps introns and exons. For (B) through (E) Left: PacBio-only. Right: PacBio+Reference Merged dataset. (B) Length distribution of transcripts, shaded by classification. (C) Isoforms per gene. (D) Percentage of predicted NMD by classification. (E) Example of the improved transcriptome by combining existing reference (top, gray) with the PacBio Iso-Seq transcriptome (middle, pink) into a merged (bottom, purple) transcriptome (see also Supplementary Figure S3).

switches between seasons, with the encoded genes displaying enrichment for biosynthetic and metabolic processes (Supplementary Table S4). We then restricted the list of DIU and Major Isoform switching genes to those that displayed less than 20% change in overall gene-level expression in response to the

season to focus on genes that did not have substantial changes in gene expression levels between seasons. Testing this list of genes for enrichments in biological process found the top term as negative regulation of protein metabolic process (GO:0051248) (Supplementary Table S5). This indicates that one
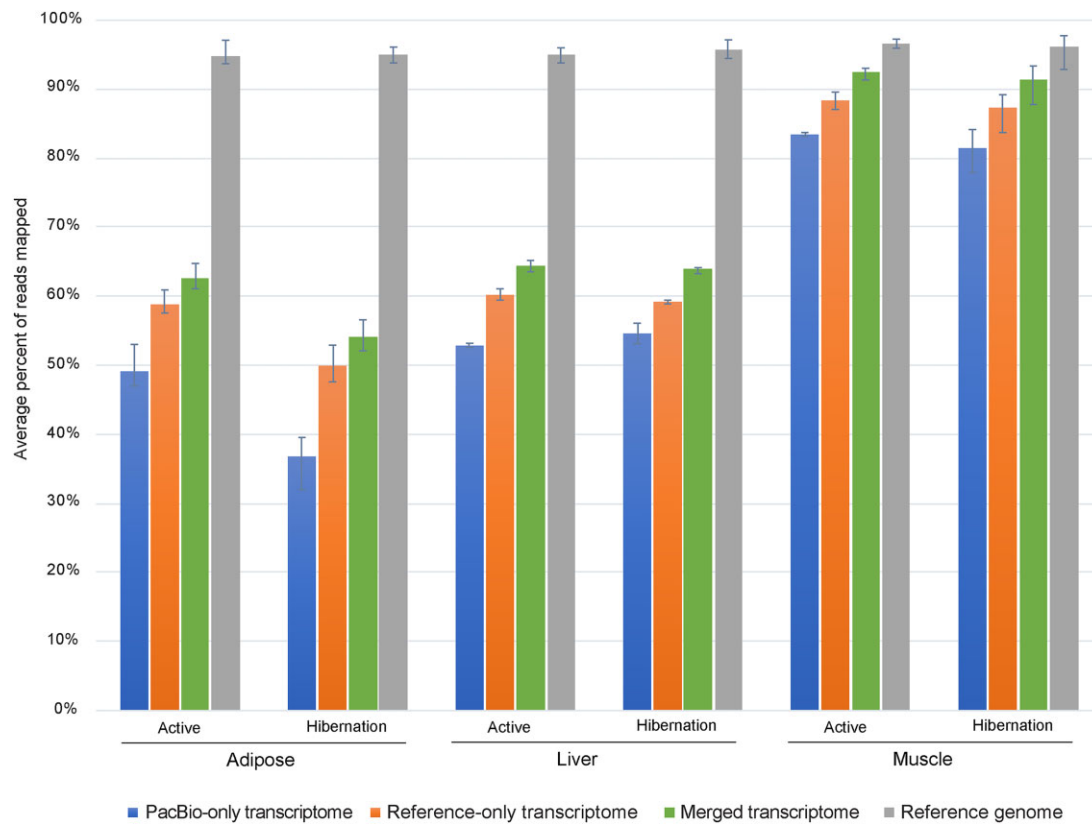
Figure 4 Ribosomal RNA depleted RNA-seq data mapped to transcriptomes and the reference genome. Average percentage of reads mapped per season and tissue. Error bars indicate range.

Table 1 Number and percentage of genes showing DIU and/or major isoform switching between hibernation and active seasons in each tissue

	Adipose	Liver	Muscle
DIU—major isoform switching	779 (8.7%)	201 (2.7%)	85 (1.6%)
DIU—no major isoform switching	1679 (18.8%)	379 (5.0%)	198 (3.7%)
Not DIU—major isoform switching	1196 (13.4%)	1228 (16.3%)	826 (15.4%)
Not DIU—no major isoform switching	5268 (59.0%)	5706 (75.9%)	4255 (79.3%)
Total genes analyzed	8922	7514	5364

component of hibernation adaptations, the need to consume reserve fat stores, is partly re-programmed by a shift in mRNA isoform ratios rather than the overall expression level of the gene.

One gene that shows a dramatic, seasonal isoform switch in adipose tissue is Integrin Subunit Beta 3 (*ITGB3BP*), imparting a nearly binary switch in five of the six sampled bears (Figure 5). While at the gene-level there is very little change in overall mean expression of *ITGB3BP* (7.3%), a major isoform switch occurs between active and hibernation (Figure 5, A and B). The two isoforms differ by a cassette exon near the 3' end of the open reading frame, resulting in a new C-terminal peptide sequence for the protein. Prior characterization of the ITGB3BP protein in mammals demonstrated that it acts as transcriptional co-regulator amongst several nuclear hormone receptor circuits affecting both the retinoic acid (RXR) and thyroid hormone (TR) pathways (Li *et al.* 1999) Since nuclear hormones play a key role in metabolic control and are implicated in homeostasis during hibernation (Nelson *et al.* 2009), we further investigated the putative consequences of the isoform switch. The bear isoform PB.6860.1, which is predominant in bear adipose tissue during hibernation, includes a cassette exon near

the 3' end of the open reading frame and encodes a protein isoform of the same 177 amino acid length as the predominant human isoform with 76% identity (Figure 5D). Skipping of the cassette exon (PB.6860.2) is more common during the active season, and this splicing pattern is also observed in humans, according to Genotype-Tissue Expression (GTeX) project data (Aguet *et al.* 2017). This event results in a truncated C-terminus, which we predict would alter the downstream co-regulator activity of ITGB3BP based on studies of its domain structure (Li *et al.* 2001). Studies of the human protein showed that the C-terminal LXXIL motif, conserved in the bear and encoded by the exon-included isoform (PB.6860.1), acts as a receptor-interaction domain (RID) (Li *et al.* 2001). This motif along with the human N-terminal LXXLL motif, also conserved in the bear, act in a cooperative fashion during interaction with the nuclear hormone receptor dimer. As the isoform without the exon (PB.6860.2) lacks the C-terminal LXXIL motif and terminates instead with a dipeptide GI, this alternative splicing event may impart broad downstream consequences for hormone receptor signaling in adipose tissue while bears are in an active state.

Figure 5 Differential isoform expression of *ITGB3BP* mRNA in adipose tissue. (A) Expression changes at the gene and coding sequence (CDS) level. (B) Short read expression of the two *ITGB3BP* isoforms across all bears and tissues. (C) Isoform structure of two *ITGB3BP* isoforms, dashed box indicates the region of the isoforms that differs. (D) Protein alignment (MUSCLE) for the human isoform and the bear isoform observed in the active season. Putative functional motifs based on the human protein co-activator function are shown in bold text. The slash represents the beginning of C-terminal peptide sequence altered by the splicing event observed in hibernating bears. The skipped exon results in a shorter C-terminus shown below the full-length alignments.

In liver, an example of a gene with significant DIU and isoform switching was Carbonic Anhydrase 2 (*CA2*). Carbonic anhydrase is essential for bone resorption and acid-base balance and has been shown to be down-regulated in bones of hibernating black bears (Shah *et al.* 2004; Goropashnaya *et al.* 2021). Given the widespread tissue expression of *CA2*, it is reasonable to suspect it

plays an important role in many tissues besides bone in maintaining calcium homeostasis. Thus, isoform switching and DIU could provide novel means of regulation of this important gene product. The *CA2* gene expressed two isoforms (Supplementary Figure S3). The shorter isoform, which lacks a coding exon and initiates transcription downstream compared to the other

isoform, was upregulated in the hibernation season for all six bears, while the longer isoform showed decreased expression during hibernation. Interestingly, the shorter isoform, which is annotated in the human genome, was not annotated in the bear reference transcriptome and was only found in the PacBio Iso-Seq data (Figure 3).

In muscle, an example of a gene with significant DIU and isoform switching, was cryptochrome 2 (*CRY2*). There are four expressed isoforms of *CRY2*, two of which are novel and showed very little changes from active season to hibernation and the other two isoforms, which show major isoform switching (Supplementary Figure S4). The two isoforms that show major isoform switching differed by only 14 bp at the last donor site; both isoforms were annotated in the reference, and both were predicted to encode for the same protein. The longer isoform showed decreased isoform usage while the shorter isoform showed increased isoform usage during hibernation. Cryptochrome 2 belongs to a family of proteins integral to the operation of the circadian clock by acting as a transcriptional repressor (Griffin *et al.* 1999). Interestingly, a damaging mutation has been found in the polar bear genome suggesting that this circadian gene may be particularly prone to modification (Welch *et al.* 2014). Given the expression of circadian rhythms in hibernating brown bears (Jansen *et al.* 2016), DIU may represent a novel mechanism of regulation under these conditions. Studies examining DIU of *Cry2* in rodent hibernators that do not undergo circadian rhythms in hibernation could clarify this putative role.

In addition to identifying genes with DIU and major isoform switching, we also identified genes with evidence for DIU without major isoform switching (Table 1). For example, four transcript isoforms of the homeobox transcription factor *PRRX1* (Supplementary Figure S5), corresponding to human *PRRX1a* (PB.14470.3); *PRRX1b* (PB.14470.1), and an unnamed human isoform (PB.14470.4). The fourth transcript, PB.14470.2, is a noncoding isoform. Transcription factor *PRRX1* is highly expressed in adipose and muscle. In adipose, the PB.14470.1 isoform continues to predominate in both active and hibernation seasons compared to the other isoforms, but also shows significantly higher expression during hibernation. This isoform (PB.14470.1) also contains a 3280 bp 5′ UTR that is lacking in PB.14470.3, while the latter contains a fifth exon. This same transcription factor in muscle shows no change in isoform expression between the seasons (Supplementary Figure S5).

A gene that shows DIU and is positively regulated in adipose by *PRRX1* in humans is *COL6A3*. *COL6A3* expression is positively related to *PRRX1* in hibernation, but not in active season adipose (Supplementary Figure S6); overall the relationship between these two genes is opposite to that in humans. Two coding isoforms (out of six) of *COL6A3* predominate and both isoforms are reduced in hibernation compared to active season (Supplementary Figure S6).

Discussion

Our study provides an unprecedented view into hibernation biology through the lens of RNA processing by producing a dataset that improved the annotation of the brown bear genome and reinforced the important role adipose tissue plays in hibernation. This approach allowed us to characterize isoforms that are changing between active and hibernation states, even when the gene itself shows no significant change in expression levels. While studies of differentially expressed genes have provided much of our current understanding in hibernation biology (Srere

et al. 1992; Shimozuru *et al.* 2016; Chayama *et al.* 2018; Faherty *et al.* 2018; Gautier *et al.* 2018; Jansen *et al.* 2019; Srivastava *et al.* 2019), determining genes where functionally distinct isoforms change between seasons is the next essential biological mechanism to uncover.

As we have shown in this study, metabolically active tissues vary dramatically in their isoform usage. Adipose tissue has been recognized as a source of critical physiological mechanisms during hibernation (Rigano *et al.* 2017; Jansen *et al.* 2019); the inherent complexity and importance of adipose biology is further underscored in this study. The large percentage of genes with DIU compared to liver and muscle, and the large number of differentially expressed genes support the dynamic role of adipose in hibernation where both transcription and RNA processing play concerted roles. The difference in the percentage of reads mapping to the transcriptome as compared to the genome in adipose and liver suggests that the ribosomal-depleted RNA-seq data had more intronic and intergenic reads in those tissues. It is also important to note that due to the fibrous nature of muscle, slightly different extraction kits were used for the muscle as compared to adipose and liver tissue RNA extraction for the short-read data (Jansen *et al.* 2019). While extraction method may influence the percentage of intronic and intergenic reads, there may also be differences in splicing efficiency between tissues and states. Regardless, adipose showed differences between active and hibernation, suggesting that intron retention may be an especially important mechanism for gene regulation and function in adipose during hibernation. It is also possible that global rates and efficiency of RNA processing by the spliceosome may also vary between the tissues and states.

In addition to hibernation biology, our results also highlight several transcriptional events that may have important implications for humans. For example, the very strong positive relationship between *COL6A3* and *PRXX1* expression in human adipose tissue (Dankel *et al.* 2020) is completely absent in bears, providing a possible explanation for why bears exhibit little to no inflammatory signatures even during periods of extreme adiposity and when consuming a high saturated fat diet (Rivet *et al.* 2017). Even the expected (positive) relationship between *COL6A3* and TGFß is absent in bears. However, elevated PRRX1 is associated with *PPARG* (Peroxisome Proliferator Activated Receptor Gamma) reductions in hibernating bear adipose tissue as has been observed previously (Claussnitzer *et al.* 2014). Thus, certain aspects of adipose function appear dissociable in bears which could lead to a more precise understanding of the contribution of adipose tissue to human disease states. Furthermore, since various single nucleotide polymorphisms within the homeobox domain superfamily associate with type 2 diabetes in humans, manipulation of PRRX1 isoforms in hibernating bear adipocytes (insulin resistant) in vitro could be used to test these relationships more precisely without the confounding interaction with COL6A3.

While PacBio Iso-Seq has been used widely for plant and animal genome annotation (Wang *et al.* 2020; Ramberg *et al.* 2021), and recently shown to shed light in cell-type specific isoform expressions in brain regions (Gupta *et al.* 2018), this study included multiple biological replicates per tissue and condition that incorporated both RNA-seq and Iso-Seq data for identifying DIU. The high mappability of the matching RNA-seq data to the Iso-Seq transcriptome shows promise for other less well-annotated organisms, where Iso-Seq may serve as a sole reference transcriptome upon which RNA-seq may be used for analyses. We saw a high correlation at the gene level between data types from the same tissue despite the fact that the Iso-Seq data

were lower coverage. There was also a moderate correlation at the isoform level within data type, but less so between data types, likely because of the low sequencing depth of the long-read data.

Moving forward, we aim to create an improved reference transcriptome at higher coverage, as well as improve the existing reference genome, to create a comprehensive reference that may better serve for studying differential isoform expressions in hibernation. Comparative studies across bears will also provide a much-needed framework for comparing hibernating and nonhibernating species. Indeed, long-read sequencing was used to improve the polar bear (*Ursus maritimus*) reference genome annotation set (Byrne *et al.* 2019), which provides a valuable additional resource for comparative studies. Although the brown bear (*U. arctos*) genome (Taylor *et al.* 2018) is not a chromosome-level assembly, we were able to improve the annotations and these data can be used in future improvements of the genome assembly.

In summary, our study demonstrates the utility of PacBio Iso-Seq for determining isoform differences between hibernating and active brown bears. Importantly, we found that adipose is the most dynamic tissue during hibernation with the highest number of genes with DIU and isoform switching as compared to liver and muscle. These findings and datasets provide a rich new resource for studying hibernation biology and understanding metabolic function. Additionally, this resource provides a basis for incorporating isoform changes in studying hibernation and translating findings to solving human diseases.

Data availability

The datasets generated for the current study are available in the NCBI SRA repository under BioProject PRJNA727613. The datasets reanalyzed in this study are available in the NCBI SRA repository under BioProject PRJNA413091. The code for this project is available at: https://github.com/jokelley/brownbear-isoseq-act-hib

Supplementary material is available at G3 online.

Ethics approval

The bears used in this study were housed at the Washington State University Bear Research, Education, and Conservation Center. All procedures were approved by the Washington State University Institutional Animal Care and Use Committee (IACUC) under protocol number ASAF 6546.

Conflicts of interest

E.T., J.G.U., and M.V. are employees of Pacific Biosciences.

Literature cited

Aguet F, Brown AA, Castel SE, Davis JR, He Y, *et al.* 2017. Genetic effects on gene expression across human tissues. Nature. 550: 204–213.

Barrows ND, Nelson OL, Robbins CT, Rourke BC. 2011. Increased cardiac alpha-myosin heavy chain in left atria and decreased myocardial insulin-like growth factor (IGF-I) expression accompany low heart rate in hibernating Grizzly bears. Physiol Biochem Zool. 84:1–17.

Berg von Linde M, Arevstrom L, Frobert O. 2015. Insights from the Den: how hibernating bears may help us understand and treat human disease. Clin Transl Sci. 8:601–605.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 34:525–527.

Byrne A, Supple MA, Volden R, Laidre KL, Shapiro B, *et al.* 2019. Depletion of hemoglobin transcripts and long-read sequencing improves the transcriptome annotation of the polar bear (*Ursus maritimus*). Front Genet. 10:643.

Chayama Y, Ando L, Sato Y, Shigenobu S, Anegawa D, *et al.* 2018. Molecular basis of white adipose tissue remodeling that precedes and coincides with hibernation in the syrian hamster, a food-storing hibernator. Front Physiol. 91973:1973.

Cheatham B, Kahn CR. 1995. Insulin action and the insulin signaling network. Endocr Rev. 16:117–142.

Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, *et al.*; DIAGRAM+Consortium. 2014. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 156:343–358.

Dankel SN, Grytten E, Bjune JI, Nielsen HJ, Dietrich A, *et al.* 2020. COL6A3 expression in adipose tissue cells is associated with levels of the homeobox transcription factor PRRX1. Sci Rep. 10: 20164.

de la Fuente L, Arzalluz-Luque A, Tardaguila M, Del Risco H, Marti C, *et al.* 2020. tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. Genome Biol. 21:119.

Faherty SL, Villanueva-Cañas JL, Blanco MB, Albà MM, Yoder AD, *et al.* 2018. Transcriptomics in the wild: hibernation physiology in free-ranging dwarf lemurs. Mol Ecol. 27:709–722.

Fedorov VB, Goropashnaya AV, Stewart NC, Tøien Ø, Chang C, *et al.* 2014. Comparative functional genomics of adaptation to muscular disuse in hibernating mammals. Mol Ecol. 23:5524–5537.

Fedorov VB, Goropashnaya AV, Tøien Ø, Stewart NC, Chang C, *et al.* 2011. Modulation of gene expression in heart and liver of hibernating black bears (Ursus americanus). BMC Genomics. 12:171.

Fedorov VB, Goropashnaya AV, Toien O, Stewart NC, Gracey AY, *et al.* 2009. Elevated expression of protein biosynthesis genes in liver and muscle of hibernating black bears (*Ursus americanus*). Physiol Genomics. 37:108–118.

Gautier C, Bothorel B, Ciocca D, Valour D, Gaudeau A, *et al.* 2018. Gene expression profiling during hibernation in the European hamster. Sci Rep. 8:13167.

Goropashnaya AV, Toien O, Ramaraj T, Sundararajan A, Schilkey FD, *et al.* 2021. Transcriptional changes and preservation of bone mass in hibernating black bears. Sci Rep. 11:8281.

Griffin EA, Jr, Staknis D, Weitz CJ. 1999. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. Science. 286: 768–771.

Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, *et al.* 2018. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol. 36:1197–1202.

Harlow HJ, Lohuis T, Beck TD, Iaizzo PA. 2001. Muscle strength in overwintering bears. Nature. 409:997.

Hellgren EC. 1998. Physiology of hibernation in bears.. Ursus. 10: 467–477.

Hershey JD, Robbins CT, Nelson OL, Lin DC. 2008. Minimal seasonal alterations in the skeletal muscle of captive brown bears. Physiol Biochem Zool. 81:138–147.

Jansen HT, Leise T, Stenhouse G, Pigeon K, Kasworm W, *et al.* 2016. The bear circadian clock doesn't 'sleep' during winter dormancy. Front Zool. 13:42.

Jansen HT, Trojahn S, Saxton MW, Quackenbush CR, Evans Hutzenbiler BD, *et al.* 2019. Hibernation induces widespread transcriptional remodeling in metabolic tissues of the grizzly bear. Commun Biol. 2:336.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 37:907–915.

Li D, Desai-Yajnik V, Lo E, Schapira M, Abagyan R, Samuels HH.NRIF3 is a novel coactivator mediating functional specificity of nuclear hormone receptors. Mol Cell Biol. 1999;19(10): 7191–7202.

Li D, Wang F, Samuels HH. 2001. Domain structure of the NRIF3 family of coregulators suggests potential dual roles in transcriptional regulation. Mol Cell Biol. 21:8371–8384.

Martin SL. 2008. Mammalian hibernation: a naturally reversible model for insulin resistance in man? Diab Vasc Dis Res. 5:76–81.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, *et al.* 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 49:D394–D403.

Nelson CJ, Otis JP, Carey HV. 2009. A role for nuclear receptors in mammalian hibernation. J Physiol. 587:1863–1870.

Nelson OL, Robbins CT. 2010. Cardiac function adaptations in hibernating grizzly bears (*Ursus arctos horribilis*). J Comp Physiol B. 180: 465–473.

Nelson OL, Robbins CT, Wu YM, Granzier H. 2008. Titin isoform switching is a major cardiac adaptive response in hibernating grizzly bears. Am J Physiol Heart Circ Physiol. 295:H366–H371.

Pertea G, Pertea M. 2020. GFF Utilities: GffRead and GffCompare [version 2; peer review: 3 approved]. F1000Research. 9:304.

Ramberg S, Hoyheim B, Ostbye TK, Andreassen R. 2021. A de novo full-length mRNA transcriptome generated from hybrid-corrected PacBio long-reads improves the transcript annotation and identifies thousands of novel splice variants in Atlantic Salmon. Front Genet. 12:656334.

Rigano KS, Gehring JL, Evans Hutzenbiler BD, Chen AV, Nelson OL, *et al.* 2017. Life in the fat lane: seasonal regulation of insulin sensitivity, food intake, and adipose biology in brown bears. J Comp Physiol B. 187:649–676.

Rivet DR, Nelson OL, Vella CA, Jansen HT, Robbins CT. 2017. Systemic effects of a high saturated fat diet in grizzly bears (*Ursus arctos horribilis*).. Can J Zool. 95:797–807.

Salmov N, Vikhlyantsev I, Ulanova A, Gritsyna Y, Bobylev A, *et al.* 2015. Seasonal changes in isoform composition of giant proteins of thick and thin filaments and titin (connectin) phosphorylation level in striated muscles of bears (Ursidae, Mammalia). Biochemistry (Mosc). 80:343–355.

Shah GN, Bonapace G, Hu PY, Strisciuglio P, Sly WS. 2004. Carbonic anhydrase II deficiency syndrome (osteopetrosis with renal tubular acidosis and brain calcification): novel mutations in CA2 identified by direct sequencing expand the opportunity for genotype-phenotype correlation. Hum Mutat. 24:272.

Shimozuru M, Nagashima A, Tanaka J, Tsubota T. 2016. Seasonal changes in the expression of energy metabolism-related genes in white adipose tissue and skeletal muscle in female Japanese black bears. Comp Biochem Physiol B-Biochem Mol Biol. 196–197: 38–47.

Srere HK, Wang LCH, Martin SL. 1992. Central role for differential gene-expression in mammalian hibernation. Proc Natl Acad Sci U S A. 89:7119–7123.

Srivastava A, Kumar Sarsani V, Fiddes I, Sheehan SM, Seger RL, *et al.* 2019. Genome assembly and gene expression in the American black bear provides new insights into the renal response to hibernation. DNA Res. 26:37–44.

Stenvinkel P, Jani AH, Johnson RJ. 2013. Hibernating bears (Ursidae): metabolic magicians of definite interest for the nephrologist. Kidney Int. 83:207–212.

Taylor GA, Kirk H, Coombe L, Jackman SD, Chu J, *et al.* 2018. The genome of the North American brown bear or grizzly: *Ursus arctos ssp. horribilis.* The Genome of the North American Brown Bear or Grizzly: Ursus arctos ssp. horribilis. Genes (Basel). 9:598.

Wang B, Tseng E, Baybayan P, Eng K, Regulski M, *et al.* 2020. Variant phasing and haplotypic expression from long-read sequencing in maize. Commun Biol. 3:78.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, *et al.* 2008. Alternative isoform regulation in human tissue transcriptomes. Nature. 456:470–476.

Ware JV, Nelson OL, Robbins CT, Jansen HT. 2012. Temporal organization of activity in the brown bear (*Ursus arctos*): roles of circadian rhythms, light, and food entrainment. Am J Physiol Regul Integr Comp Physiol. 303: R890–R902.

Welch AJ, Bedoya-Reina OC, Carretero-Paulet L, Miller W, Rode KD, *et al.* 2014. Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. Genome Biol Evol. 6:433–450.

Welinder KG, Hansen R, Overgaard MT, Brohus M, Sønderkær M, *et al.* 2016. Biochemical foundations of health and energy conservation in hibernating free-ranging subadult brown bear *Ursus arctos.* J Biol Chem. 291:22509–22523.

Zhang X, Chen MH, Wu X, Kodani A, Fan J, *et al.* 2016. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. Cell. 166:1147–1162e1115.

Communicating editor: A. Sethuraman