bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

***Title*:**

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

Authors:

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

Author Affiliations:

[1]Pacific Biosciences, Menlo Park, CA, USA

[2]Department of Integrative Physiology and Neuroscience, Washington State University, Pullman, WA

[3]School of the Environment, Washington State University, Pullman, WA

[4]School of Biological Sciences, Washington State University, Pullman, WA

[5]University of Delaware, Delaware Biotechnology Institute, Sequencing & Genotyping Center, Suite 246, 15 Innovation Way, Newark, DE 19711

****Corresponding Author:***

Joanna L. Kelley, joanna.l.kelley@wsu.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Summary

Understanding hibernation in brown bears (*Ursus arctos*) can provide insight into many human diseases. During hibernation, brown bears experience states of insulin resistance, physical inactivity, extreme bradycardia, obesity, and the absence of urine production. These states closely mimic human diseases such as type 2 diabetes, muscle atrophy, renal and heart failure, cachexia, and obesity. The reversibility of these states from hibernation to active season allows for the identification of novel mediators with possible therapeutic value for humans. Recent studies have identified genes and pathways that are differentially expressed between active and hibernation seasons. However, little is known about the role of differential expression of gene isoforms on hibernation physiology. To identify both distinct and novel mRNA isoforms, we performed full-length RNA-sequencing (Iso-Seq) on three tissue types from three individuals sampled during both active and hibernation seasons. We combined the long-read data with the reference annotation for an improved transcriptome and mapped RNA-seq data from six individuals to the improved transcriptome to quantify differential isoform usage between tissues and seasons. We identified differentially expressed isoforms in all study tissues and showed that adipose has a high level of differential isoform usage with isoform switching, regardless of whether the genes were differentially expressed. Our analyses provide a comprehensive evaluation of isoform usage between active and hibernation states, revealing that differential isoform usage, 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

45 # Keywords

46 Hibernation, Brown bear, Transcriptomics, Iso-Seq, Alternative Splicing, Differential Isoform

47 Expression, Full-length transcript sequencing

48

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Background

Hibernation in bears has long been championed as a promising system for understanding the extremes of mammalian physiology and for identifying novel therapeutic targets [1, 2]. Annual hibernation in brown bears (*Ursus arctos*) involves massive physiological shifts to conserve energy during the food-scarce winter [3], 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 importantly, these phenotypes do not themselves negatively impact the bears' overall health [4-7]. For example, heart rate slows to 10 to 15 beats per minute [8], 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 [8, 9]. Hibernating bears are also well known for maintaining muscle strength, morphology, and composition during hibernation in the near complete absence of weight-bearing activity [10, 11]. Bears also exhibit insulin resistance during hibernation but are insulin sensitive during the active season [7]. 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 [2]. In fact, the need to accumulate tremendous amounts of fat and the development of insulin resistance evolved in bears as a survival strategy [12-14]. 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 [15-17]. Notably, many of these genes are involved in complex metabolic and cellular signaling pathways (e.g., insulin signaling, metabolism) that play critical roles in a variety of biological processes across vertebrates, including humans [15, 18].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Over ten thousand genes have been shown to be differentially regulated in adipose, liver, and muscle tissues between active and hibernating states, providing a set of candidate genes involved in the regulation of cellular and physiological processes that underly the metabolic suppression observed in hibernation [15]. The genes identified represent key genes and genomic regions for testing hypotheses related to the evolution and regulation of hibernation. While it is known that mRNA isoforms vary between tissues, cell types, and developmental stages [19, 20] and play a role in cellular processes, studies in bears have focused on global gene expression levels have not investigated the mRNA processing shifts that result in different isoforms and thus proteome output that occur with active and hibernating seasons.

We hypothesize that different transcript isoforms contribute to the reversible states achieved during hibernation. Indeed, in brown bears and Himalayan black bears (*Ursus thibetanus ussuricus*), it has been shown that the amount of titin does not differ between active and hibernation seasons but the relative abundance of two prominent isoforms may explain increased ventricular stiffness during hibernation [21, 22]. However, isoform differences have been explored only in a few select cases and little is known about the role different isoforms contribute to the hibernation phenotype on a large scale. Because of the capability of sequencing full-length RNA transcripts, SMRT Sequencing is ideal for identifying the 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

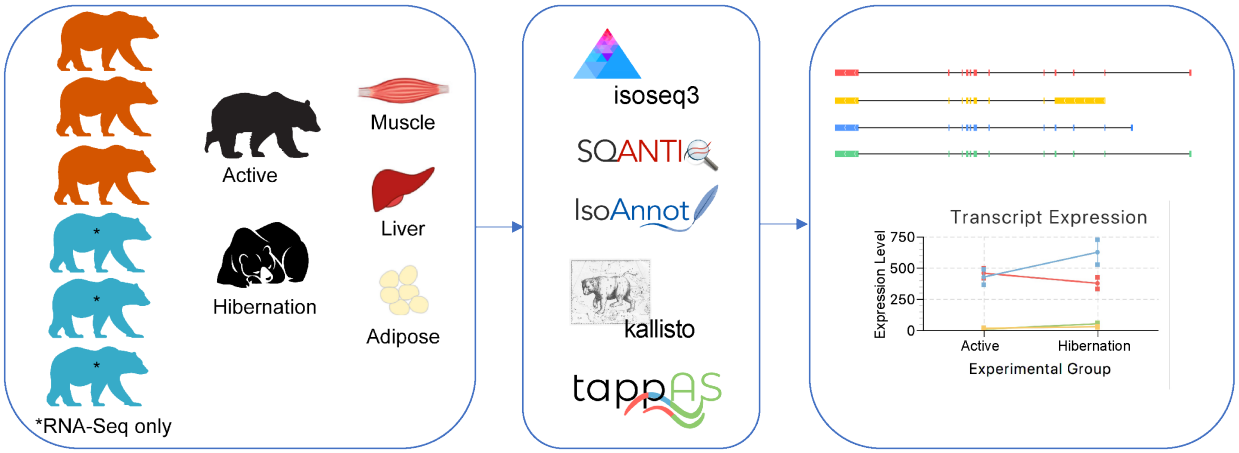


Figure 1. Bear transcriptome workflow. For each of the six bears, tissues (muscle, liver, adipose) were extracted during active or hibernation seasons. PacBio Iso-Seq and Illumina RNA-seq data were collected from three of the bears (orange) and Illumina RNA-seq data was collected from three additional bears (blue). Iso-Seq data was 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 usage was determined using tappAS.

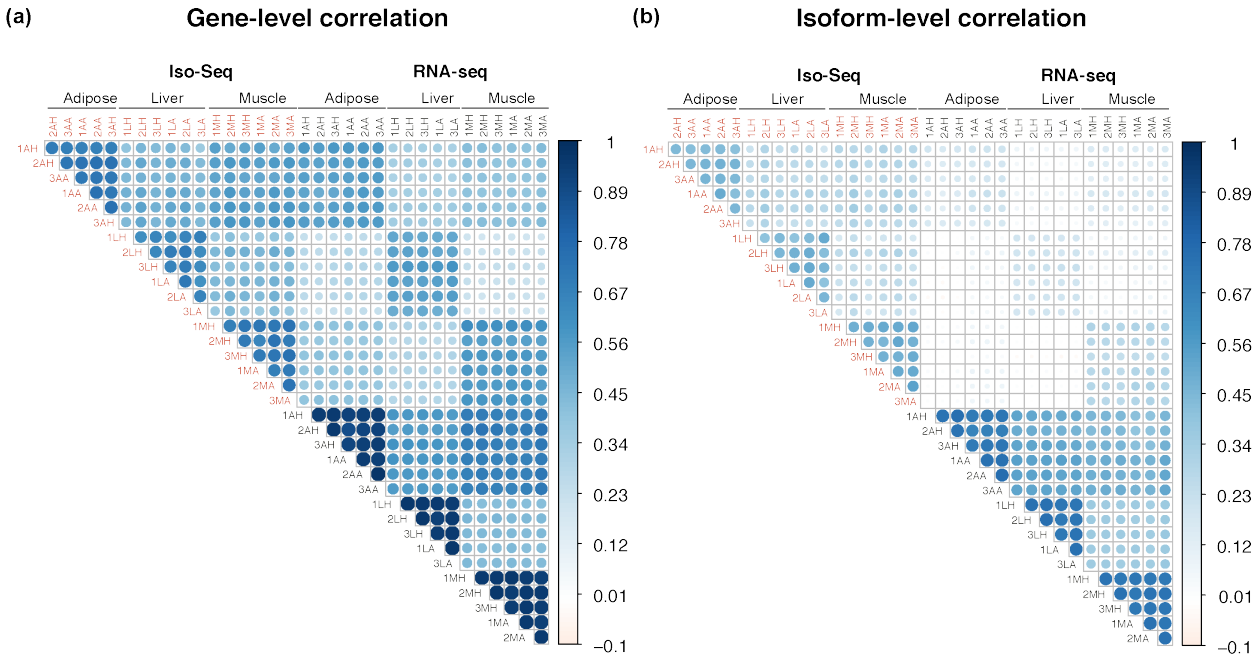
Results

High Correlation between Long- and Short-Read Data at Gene Level

We analyzed RNA-sequencing data from bears in active or hibernation seasons 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 [15]. 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 (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

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

111 then evaluated the gene-level correlation of long- versus short-read data. For this correlation, we
112 used only transcripts that were present in both Iso-Seq and RNA-seq datasets. When comparing
113 data from the same individuals (albeit sampled in different years), the highest correlations are
114 within data type, regardless of season (active or hibernation) (Figure 2). There is also a high
115 correlation between data types at the gene level, especially within the same tissues (Figure 2a).
116 The lower concordance at the isoform-level within the Iso-Seq samples as compared to the
117 within RNA-seq samples, is likely explained by the lower sequencing coverage of the long-read
118 dataset (Figure 2b). Nevertheless, we see consistent gene-level correlation for the matching
119 samples across data types, while samples from the same tissues or animals have higher
120 correlation than different tissues, as expected.



121

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

126

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

An Improved Reference Transcriptome and Full-Length Isoform Annotation using Iso-Seq

The existing reference transcriptome contained 30,263 genes encompassing 58,335 transcripts. The Iso-Seq transcriptome was classified against the reference transcriptome and found to contain 12,018 known and 907 novel genes (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 S1a-S1d).

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, NNC) while greatly increasing the number of known transcripts (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, adipose) had different distributions of transcript structural changes, with adipose and liver having similar distributions and muscle having a much larger number of full-splice match (FSM) and fewer novel in catalog (NIC) transcripts (Figure S2). Importantly, the improved reference transcriptome, with the full-length transcripts originating from samples of interest, provides a starting point for the discovery of differential isoform usage (DIU) that would otherwise be missed. One example of such a finding is the isoform expression of the *CA2* gene (Figure S1e), described in more detail in later sections.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

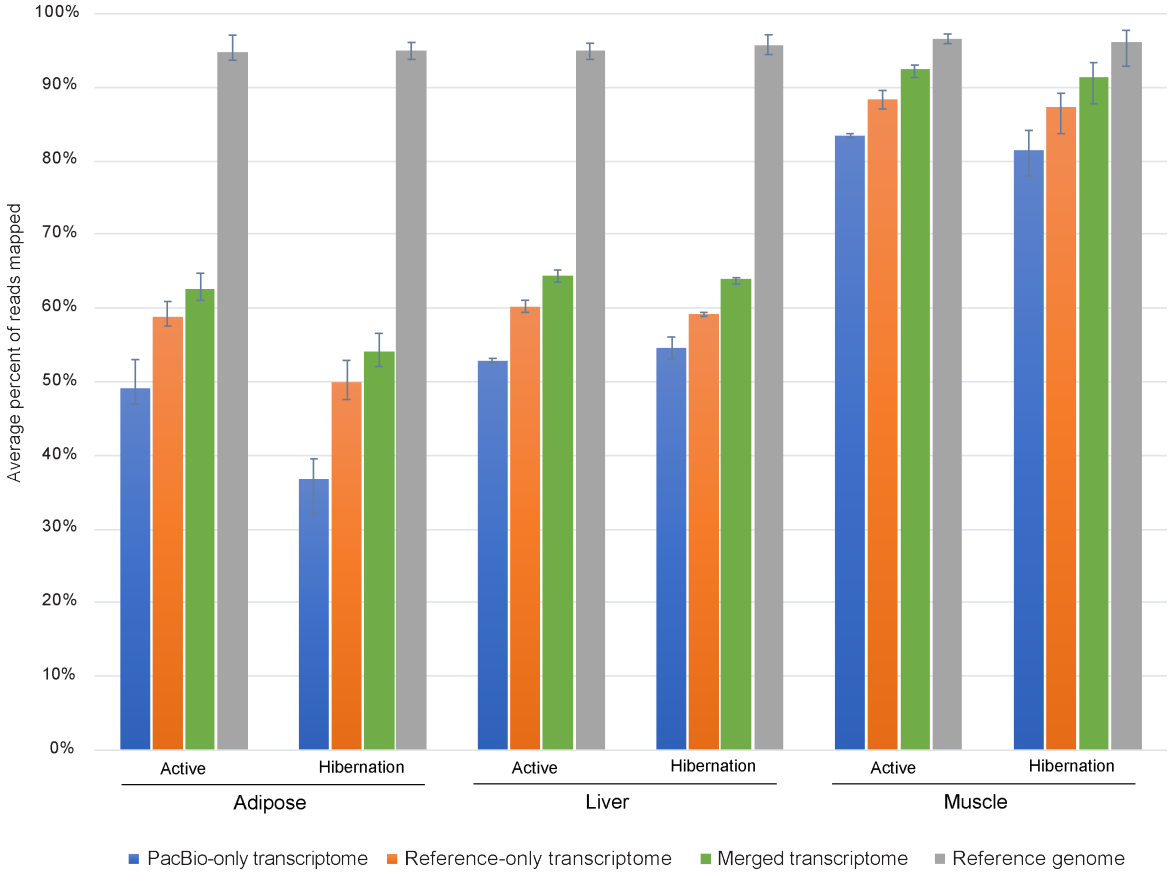


Figure 3. Ribosomal RNA depleted RNA-seq data mapped to transcriptomes and the reference genome. Average percent of reads mapped per season and tissue. Error bars indicate range.

Tissue-specific Differential Isoform Usage from Active to Hibernation Season

We assessed whether the relative abundance of different isoforms for each gene varied between the seasons (differential isoform usage, 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, Figure S3). Adipose had the highest incidence of DIU with regards to both

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

number and percent of all analyzed genes, with and without major isoform switching (27.5% of genes; Table 1). In contrast, major isoform switching between states without differential isoform usage was fairly consistent across tissues.

	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

Table 1: Number and percentage of genes showing differential isoform usage (DIU) and/or major isoform switching between hibernation and active seasons in each tissue.

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 most DIU + Major Isoform switches between seasons, with the encoded genes displaying enrichment for biosynthetic and metabolic processes (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. Testing this list of genes for enrichments in cellular function found a shift to more modest enrichments with the top categories focused on autophagy (Table S5). This indicates that one component of hibernation

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

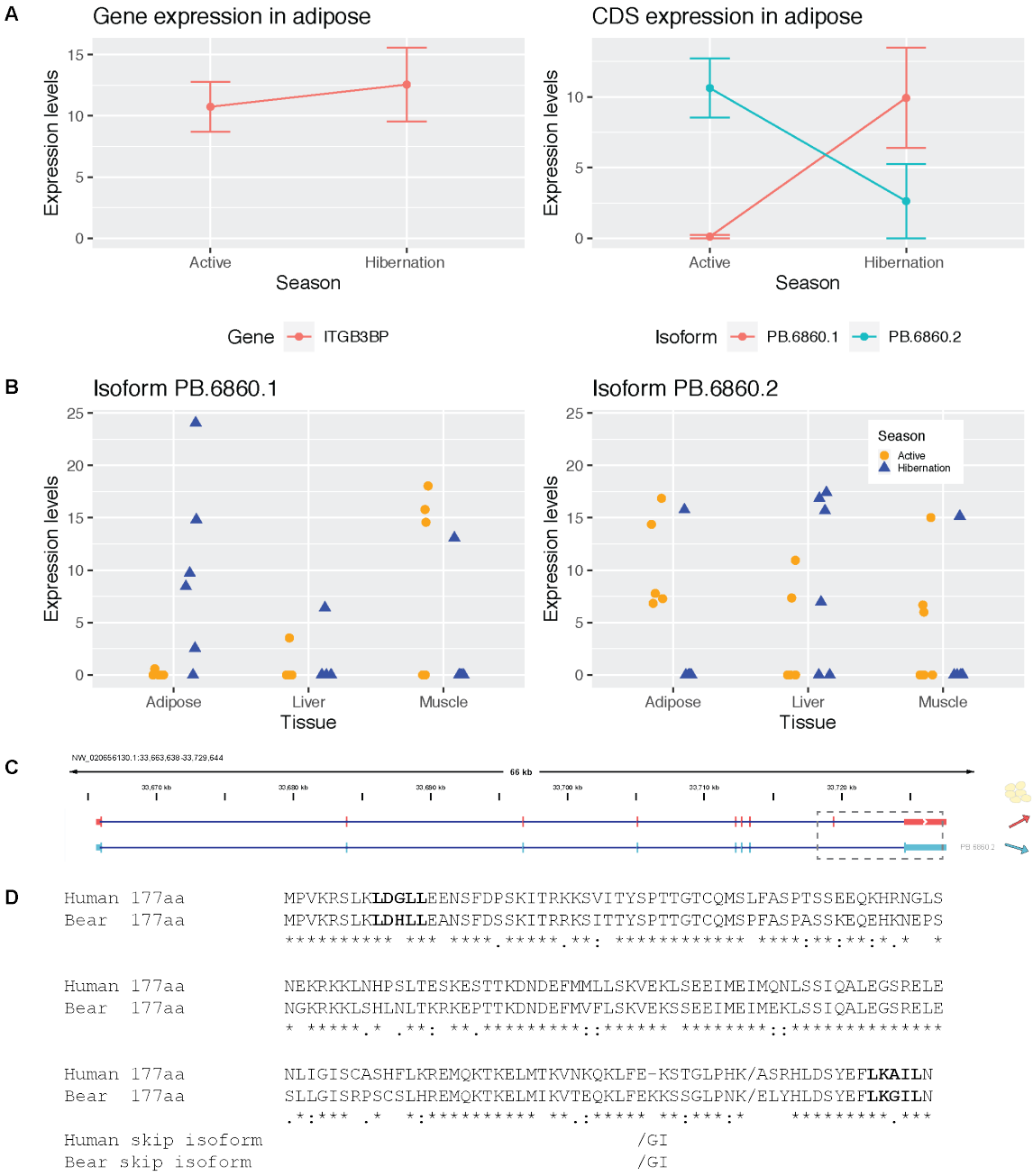
adaptations, the need to consume reserve fat stores, is partly re-programmed by a shift in mRNA isoform ratios rather than the overall expression level of the gene.

One gene that shows a dramatic isoform switch in adipose in concert with the seasons is Integrin Subunit Beta 3 Binding Protein (*ITGB3BP*), imparting a nearly binary switch in five of the six sampled bears (Figure 4). There is very little change in overall gene expression but the major isoform switches between active and hibernation (Figure 4a, 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. Since nuclear hormones play a key role in metabolic control and are implicated in homeostasis during hibernation, we further investigated the putative consequences of the isoform switch [23]. 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 4d). 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 GTeX project data [24]. 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. 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) [25]. This motif along with the human N-terminal LXXLL motif, also conserved in the bear, act in a cooperative

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



221

222 **Figure 4.** Differential isoform expression of *ITGB3BP* mRNA in adipose tissue. (a) expression changes at the

223 gene and coding sequence (CDS) level. (b) Short read expression of the two *ITGB3BP* isoforms across all

224 bears and tissues. (c) Isoform structure of two *ITGB3BP* isoforms, dashed box indicates the region of the

225 isoforms that differs. (d) Protein alignment (MUSCLE) for the human isoform and the bear isoform observed

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

226 in the active season. Putative functional motifs based on the human protein co-activator function are shown in
227 larger bold text. The bold slash represents the beginning of C-terminal peptide sequence altered by the splicing
228 event observed in hibernating bears. The skipped exon results in a shorter C-terminus shown below the full-
229 length alignments.
230
231 In liver, an example of a gene with significant DIU and isoform switching was Carbonic
232 Anhydrase 2 (*CA2*). The *CA2* gene expressed two isoforms (Figure S4). The shorter isoform,
233 which lacks a coding exon and initiates transcription downstream compared to the other isoform,
234 was upregulated in the hibernation season for all six bears, while the longer isoform showed
235 decreased expression during hibernation. Interestingly, the shorter isoform, which is annotated in
236 the human genome, was not annotated in the reference transcriptome and was only found in the
237 PacBio Iso-Seq data (Figure S1e).
238
239 In muscle, an example of a gene with significant DIU and isoform switching, was cryptochrome
240 2 (*CRY2*). There are four expressed isoforms of *CRY2*, two of which are novel and showed very
241 little changes from active season to hibernation and the other two isoforms, which show major
242 isoform switching (Figure S5). The two isoforms that show major isoform switching differed by
243 only 14 bp at the last donor site; both isoforms were annotated in the reference, and both were
244 predicted to encode for the same protein. The longer isoform showed decreased isoform usage
245 while the shorter isoform showed increased isoform usage during hibernation.
246
247 In addition to identifying genes with DIU and major isoform switching, we also identified genes
248 with evidence for DIU without major isoform switching (Table 1). For example, four transcript
249 isoforms of the homeobox transcription factor *PRRX1* (Figure S6), corresponding to human

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

PRRX1a (PB.14470.3); *PRRX1b* (PB.14470.1), and an unnamed human isoform (PB.14470.4). The fourth transcript, PB.14470.2, is a non-coding 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 (Figure S6).

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 (Figure S7); 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 (Figure S7).

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 [15,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

26-31], 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. The inherent complexity and importance of adipose, which has been recognized as a source of critical physiological mechanisms during hibernation [7, 15], continues to expand. 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 slightly different extraction kits were used for the muscle as compared to adipose and liver tissue RNA extraction for the short-read data [15]. 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 [32] is completely absent in

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

296 bears, providing a possible explanation for why bears exhibit little to no inflammatory signatures
297 even during periods of extreme adiposity and when consuming a high saturated fat diet
298 [33]. Even the expected (positive) relationship between *COL6A3* and TGFß is absent in bears.
299 However, elevated PRRX1 is associated with PPARG (Peroxisome Proliferator Activated
300 Receptor Gamma) reductions in hibernation bear adipose tissue as has been observed previously
301 [34]. Thus, certain aspects of adipose function appear dissociable in bears which could lead to a
302 more precise understanding of the contribution of adipose tissue to human disease states.
303 Furthermore, since various single nucleotide polymorphisms within the homeobox domain
304 superfamily associate with type 2 diabetes in humans, manipulation of PRRX1 isoforms in
305 hibernation bear adipocytes (insulin resistant) *in vitro* could be used to test these relationships
306 more precisely without the confounding interaction with COL6A3.
307
308 While PacBio Iso-Seq has been used widely for plant and animal genome annotation [35, 36],
309 and recently shown to shed light in cell-type specific isoform expressions in brain regions [37],
310 this study included multiple biological replicates per tissue and condition that incorporated both
311 RNA-seq and Iso-Seq data for identifying differential isoform usage. The high mappability of the
312 matching RNA-seq data to the Iso-Seq transcriptome shows promise for other less well-
313 annotated organisms, where Iso-Seq may serve as a sole reference transcriptome upon which
314 RNA-seq may be used for differential analysis. We saw a high correlation at the gene level
315 between data types from the same tissue despite the fact that the Iso-Seq data were lower
316 coverage. There was also a moderate correlation at the isoform level within data type, but less so
317 between data types, likely because of the low sequencing depth of the long-read data.
318

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 non-hibernating species. Indeed, long-read sequencing was used to improve the polar bear reference genome annotation set [38], which provides a valuable additional resource for comparative studies. Although the *U. arctos* genome [39] 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 differential isoform usage 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Methods

Sample Collection

For the PacBio Iso-Seq protocol, samples were collected from three bears (1 female, 2 males) at the Washington State University Bear Research, Education, and Conservation Center in January 2019 and May 2019, to represent 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 [15]. Animal care details are described in [7]. Bears were first anesthetized using the protocol described in Ware JV, Nelson OL, Robbins CT and Jansen HT [40]. Subcutaneous adipose samples were collected using a 6mm punch biopsy (Miltex, York, PA) as described in [7], 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 1200hr. 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+

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

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 100ng to 300ng 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

system SMRT Cell using sequencing chemistry 3.0 with 4 hour pre-extension and 20 hour 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 IsoSeq 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) using minimap2, then filtered and collapsed into non-redundant 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 non-redundant 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 [15] 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,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

405 with the rf-stranded flag [43]. Read counts were compared to the IsoSeq data using Pearson
406 correlation on the $\log_2$ count per million.
407
408 **Merging annotations and mapping**
409 The PacBio-only annotation set was merged with the reference-only annotations using
410 gffcompare [41]. Short read Illumina data from [15] was mapped to each transcriptome and the
411 reference genome (GenBank assembly accession: GCA_003584765.1 [39]) using HISAT2
412 (version 2.2.1) with the --rf flag and otherwise default parameters [42].
413
414 **Functional Annotation and Differential Isoform Expression Analysis**
415 Abundance count estimates for each individual were combined into a single input matrix for
416 tappAS [44]. The annotation file generated in SQANTI3 and the short-read count matrices for
417 each tissue were input into tappAS [44]. Each tissue was analyzed separately to compare
418 differential isoform expression in hibernation compared to active season. We excluded short-read
419 data from sample 1AA because it was previously shown to contain a hair follicle [15].
420 Transcripts with counts per million (cpm) less than 1.0 or coefficient of variation cutoff of 100%
421 were excluded from the analyses. In the differential isoform usage (DIU) analysis, minor
422 isoforms with a proportional of expression difference less than 0.1 were excluded from the
423 analysis and DIU was considered at an FDR < 0.05. Gene ontology (GO) enrichment of different
424 gene sets was calculated using PANTHER [45] with the following parameters: Analysis
425 Type: PANTHER Overrepresentation Test (Released 20210224), Annotation Version and
426 Release Date: GO Ontology database DOI: 10.5281/zenodo.4495804 Released 2021-02-01,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

427 Reference List: Homo sapiens (all genes in database), Annotation Data Set: GO biological

428 process, Test Type: Fisher's Exact, Correction: Calculate False Discovery Rate.

429

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Declarations

Ethics Approval and Consent to Participate

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.

Availability of Data and Materials

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

Competing Interests

Elizabeth Tseng, Jason G. Underwood and Michelle Vierra are employees of Pacific Biosciences.

Funding

The sequencing for this project was funded by a Pacific Biosciences (PacBio) SMRT grant. We would also like to acknowledge the Interagency Grizzly Bear Committee, USDA National Institute of Food and Agriculture (McIntire-Stennis project 1018967), International Association for Bear Research and Management, T. N. Tollefson and Mazuri Exotic Animal Nutrition, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Raili Korkka Brown Bear Endowment, Nutritional Ecology Endowment, and Bear Research and Conservation Endowment at Washington State University for funding and support.

Authors' Contributions

BDEH, ST, BK, MV, CTR, HTJ, JLK contributed to sampling. BK, OS, and EB performed the RNA extraction, library preparation, and sequencing. JGU, MV, and HTJ contributed to the manuscript. ET and JLK analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This research used resources from the Center for Institutional Research Computing at Washington State University. Allan Cornejo Kelley for feedback on figures.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. Stenvinkel P, Jani AH, Johnson RJ: **Hibernating bears (Ursidae): metabolic magicians of definite interest for the nephrologist.** *Kidney Int* 2013, **83:**207-212.
2. Martin SL: **Mammalian hibernation: a naturally reversible model for insulin resistance in man?** *Diab Vasc Dis Res* 2008, **5:**76-81.
3. Hellgren EC: **Physiology of hibernation in bears.** *Ursus, Vol 10 - 1998* 1998, **10:**467-477.
4. Berg von Linde M, Arevstrom L, Frobert O: **Insights from the Den: How Hibernating Bears May Help Us Understand and Treat Human Disease.** *Clin Transl Sci* 2015, **8:**601-605.
5. Welinder KG, Hansen R, Overgaard MT, Brohus M, Sonderkaer M, von Bergen M, Rolle-Kampczyk U, Otto W, Lindahl TL, Arinell K, et al: **Biochemical Foundations of Health and Energy Conservation in Hibernating Free-ranging Subadult Brown Bear Ursus arctos.** *Journal of Biological Chemistry* 2016, **291:**22509-22523.
6. Fedorov VB, Goropashnaya AV, Toien O, Stewart NC, Gracey AY, Chang CL, Qin SZ, Pertea G, Quackenbush J, Showe LC, et al: **Elevated expression of protein biosynthesis genes in liver and muscle of hibernating black bears (Ursus americanus).** *Physiological Genomics* 2009, **37:**108-118.
7. Rigano KS, Gehring JL, Evans Hutzenbiler BD, Chen AV, Nelson OL, Vella CA, Robbins CT, Jansen HT: **Life in the fat lane: seasonal regulation of insulin sensitivity, food intake, and adipose biology in brown bears.** *J Comp Physiol B* 2017, **187:**649-676.
8. Nelson OL, Robbins CT: **Cardiac function adaptations in hibernating grizzly bears (Ursus arctos horribilis).** *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology* 2010, **180:**465-473.
9. Barrows ND, Nelson OL, Robbins CT, Rourke BC: **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.** *Physiological and Biochemical Zoology* 2011, **84:**1-17.
10. Harlow HJ, Lohuis T, Beck TD, Iaizzo PA: **Muscle strength in overwintering bears.** *Nature* 2001, **409:**997.
11. Hershey JD, Robbins CT, Nelson OL, Lin DC: **Minimal seasonal alterations in the skeletal muscle of captive brown bears.** *Physiological and Biochemical Zoology* 2008, **81:**138-147.
12. Nelson R, Jones J, Wahner H, McGill D, Code C: **Nitrogen metabolism in bears: urea metabolism in summer starvation and in winter sleep and role of urinary bladder in water and nitrogen conservation.** In *Mayo Clinic Proceedings*. 1975: 141-146.
13. Nelson RA, Wahner HW, Jones JD, Ellefson RD, Zollman PE: **Metabolism of bears before, during, and after winter sleep.** *Am J Physiol* 1973, **224:**491-496.
14. Palumbo PJ, Wellik DL, Bagley NA, Nelson RA: **Insulin and Glucagon Responses in the Hibernating Black Bear.** *Bears: Their Biology and Management* 1983, **5:**291-296.
15. Jansen HT, Trojahn S, Saxton MW, Quackenbush CR, Evans Hutzenbiler BD, Nelson OL, Cornejo OE, Robbins CT, Kelley JL: **Hibernation induces widespread transcriptional remodeling in metabolic tissues of the grizzly bear.** *Commun Biol* 2019, **2:**336.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

16. Fedorov VB, Goropashnaya AV, Stewart NC, Tøien Ø, Chang C, Wang H, Yan J, Showe LC, Showe MK, Barnes BM: **Comparative functional genomics of adaptation to muscular disuse in hibernating mammals.** *Molecular ecology* 2014, **23:**5524-5537.
17. Fedorov VB, Goropashnaya AV, Toien O, Stewart NC, Chang C, Wang HF, Yan J, Showe LC, Showe MK, Barnes BM: **Modulation of gene expression in heart and liver of hibernating black bears (Ursus americanus).** *Bmc Genomics* 2011, **12**.
18. Cheatham B, Kahn CR: **Insulin action and the insulin signaling network.** *Endocr Rev* 1995, **16:**117-142.
19. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes.** *Nature* 2008, **456:**470-476.
20. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, Ozawa M, Ma J, Yoshida N, Reiter JF, et al: **Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex.** *Cell* 2016, **166:**1147-1162 e1115.
21. Nelson OL, Robbins CT, Wu YM, Granzier H: **Titin isoform switching is a major cardiac adaptive response in hibernating grizzly bears.** *American Journal of Physiology-Heart and Circulatory Physiology* 2008, **295:**H366-H371.
22. Salmov N, Vikhlyantsev I, Ulanova A, Gritsyna Y, Bobylev A, Saveljev A, Makariushchenko V, Maksudov G, Podlubnaya Z: **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 (Moscow)* 2015, **80:**343-355.
23. Nelson CJ, Otis JP, Carey HV: **A role for nuclear receptors in mammalian hibernation.** *J Physiol* 2009, **587:**1863-1870.
24. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park Y, Parsana P, Segrè AV, et al: **Genetic effects on gene expression across human tissues.** *Nature* 2017, **550:**204-213.
25. Li D, Wang F, Samuels HH: **Domain structure of the NRIF3 family of coregulators suggests potential dual roles in transcriptional regulation.** *Mol Cell Biol* 2001, **21:**8371-8384.
26. Chayama Y, Ando L, Sato Y, Shigenobu S, Anegawa D, Fujimoto T, Taii H, Tamura Y, Miura M, Yamaguchi Y: **Molecular Basis of White Adipose Tissue Remodeling That Precedes and Coincides With Hibernation in the Syrian Hamster, a Food-Storing Hibernator.** *Front Physiol* 2018, **9:**1973.
27. Srivastava A, Kumar Sarsani V, Fiddes I, Sheehan SM, Seger RL, Barter ME, Neptune-Bear S, Lindqvist C, Korstanje R: **Genome assembly and gene expression in the American black bear provides new insights into the renal response to hibernation.** *DNA Res* 2019, **26:**37-44.
28. Srere HK, Wang LCH, Martin SL: **Central Role for Differential Gene-Expression in Mammalian Hibernation.** *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89:**7119-7123.
29. Shimozuru M, Nagashima A, Tanaka J, Tsubota T: **Seasonal changes in the expression of energy metabolism-related genes in white adipose tissue and skeletal muscle in female Japanese black bears.** *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* 2016, **196:**38-47.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

30. Gautier C, Bothorel B, Ciocca D, Valour D, Gaudeau A, Dupré C, Lizzo G, Brasseur C, Riest-Fery I, Stephan J-P, et al: **Gene expression profiling during hibernation in the European hamster.** *Scientific Reports* 2018, **8:**13167.
31. Faherty SL, Villanueva-Canas JL, Blanco MB, Alba MM, Yoder AD: **Transcriptomics in the wild: Hibernation physiology in free-ranging dwarf lemurs.** *Mol Ecol* 2018, **27:**709-722.
32. Dankel SN, Grytten E, Bjune JI, Nielsen HJ, Dietrich A, Bluher M, Sagen JV, Mellgren G: **COL6A3 expression in adipose tissue cells is associated with levels of the homeobox transcription factor PRRX1.** *Sci Rep* 2020, **10:**20164.
33. Rivet DR, Nelson OL, Vella CA, Jansen HT, Robbins CT: **Systemic effects of a high saturated fat diet in grizzly bears (Ursus arctos horribilis).** *Canadian Journal of Zoology* 2017, **95:**797-807.
34. Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, Lee H, Oskolkov N, Fadista J, Ehlers K, et al: **Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms.** *Cell* 2014, **156:**343-358.
35. Wang B, Tseng E, Baybayan P, Eng K, Regulski M, Jiao Y, Wang L, Olson A, Chougule K, Buren PV, Ware D: **Variant phasing and haplotypic expression from long-read sequencing in maize.** *Communications Biology* 2020, **3:**78.
36. Ramberg S, Hoyheim B, Ostbye TK, Andreassen R: **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* 2021, **12:**656334.
37. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, Koopmans F, Barres B, Smit AB, Sloan SA, et al: **Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells.** *Nat Biotechnol* 2018.
38. Byrne A, Supple MA, Volden R, Laidre KL, Shapiro B, Vollmers C: **Depletion of Hemoglobin Transcripts and Long-Read Sequencing Improves the Transcriptome Annotation of the Polar Bear (Ursus maritimus).** *Frontiers in Genetics* 2019, **10**.
39. Taylor GA, Kirk H, Coombe L, Jackman SD, Chu J, Tse K, Cheng D, Chuah E, Pandoh P, Carlsen R, et al: **The Genome of the North American Brown Bear or Grizzly: Ursus arctos ssp. horribilis.** *Genes (Basel)* 2018, **9**.
40. Ware JV, Nelson OL, Robbins CT, Jansen HT: **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* 2012, **303:**R890-902.
41. Pertea G, Pertea M: **GFF Utilities: GffRead and GffCompare [version 2; peer review: 3 approved].** *F1000Research* 2020, **9**.
42. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.** *Nature Biotechnology* 2019, **37:**907-915.
43. Bray NL, Pimentel H, Melsted P, Pachter L: **Near-optimal probabilistic RNA-seq quantification.** *Nat Biotechnol* 2016, **34:**525-527.
44. de la Fuente L, Arzalluz-Luque A, Tardaguila M, Del Risco H, Marti C, Tarazona S, Salguero P, Scott R, Lerma A, Alastrue-Agudo A, et al: **tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing.** *Genome Biol* 2020, **21:**119.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.13.452179; this version posted July 14, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

45. Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD: **PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API.** *Nucleic Acids Research* 2020, **49:**D394-D403.