bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license.

Genetic architecture drives seasonal onset of hibernation in the 13-lined ground squirrel

Katharine R. Grabek, Thomas F. Cooke, L. Elaine Epperson, Kaitlyn K. Spees, Gleyce F. Cabral, Shirley C. Sutton, Dana K. Merriman, Sandra L. Martin and Carlos D. Bustamante

Short title: Genetic architecture of hibernation onset

Author Affiliations

Katharine R. Grabek
*Email: krgrabek@stanford.edu
Affiliations: Department of Genetics and Department Biomedical Data Science, Stanford University School of Medicine, Stanford, California, United States of America

Thomas F. Cooke
Affiliation: Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America

L. Elaine Epperson
Affiliation: Center for Genes, Environment and Health, National Jewish Health, Denver, Colorado, United States of America

Kaitlyn K. Spees
Affiliation: Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America

Gleyce F. Cabral
Affiliations: Brazil Scientific Mobility Program, Science Without Borders Fellow/CAPES, CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil; Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America

Shirley C. Sutton
Affiliation: Department of Genetics and Department of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, United States of America

Dana K. Merriman
Affiliation: Department of Biology, University of Wisconsin Oshkosh, Oshkosh, Wisconsin, United States of America

Sandra L. Martin
Affiliation: Department of Cellular and Developmental Biology, University of Colorado School of Medicine, Aurora, Colorado, United States of America

Carlos D. Bustamante

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license.

*Email: cdbustam@stanford.edu
Affiliations: Department of Genetics and Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, California, United States of America; Chan Zuckerberg Biohub, San Francisco, California, United States of America

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

Abstract

Hibernation is a highly dynamic phenotype whose timing, for many mammals, is controlled by a circannual clock and accompanied by rhythms in body mass and food intake. When housed in an animal facility, 13-lined ground squirrels exhibit individual variation in the seasonal onset of hibernation, which is not explained by environmental or biological factors, such as body mass and sex. We hypothesized that underlying genetic architecture instead drives variation in this timing. After first increasing the contiguity of the genome assembly, we therefore employed a genotype-by-sequencing approach to characterize genetic variation in 153 13-lined ground squirrels. Combining this with datalogger records, we estimated high heritability (61-100%) for the seasonal onset of hibernation. After applying a genome-wide scan with 46,996 variants, we also identified 21 loci significantly associated with hibernation immergence, which alone accounted for 54% of the variance in the phenotype. The most significant marker (SNP 15, $p=3.81x10^{-6}$) was located near *prolactin-releasing hormone receptor* (*PRLHR*), a gene that regulates food intake and energy homeostasis. Other significant loci were located near genes functionally related to hibernation physiology, including *muscarinic acetylcholine receptor M2* (*CHRM2*), involved in the control of heart rate, *exocyst complex component 4* (*EXOC4*) and *prohormone convertase 2* (*PCSK2*), both of which are involved in insulin signaling and processing. Finally, we applied an expression quantitative loci (eQTL) analysis using existing transcriptome datasets, and we identified significant ($q<0.1$) associations for 9/21 variants. Our results highlight the power of applying a genetic mapping strategy to hibernation and present new insight into the genetics driving its seasonal onset.

Introduction

Hibernation is a highly dynamic phenotype that maximizes energy savings during periods of low resource availability. For a number of mammals, such as the 13-lined ground squirrel, *Ictidomys tridecemlineatus*, an endogenous circannual clock controls the timing of winter hibernation, along with rhythms in reproductive behavior, body mass, and food intake [1-3]. These hibernators partition their year between two distinct states, homeothermy and heterothermy (a.k.a hibernation, Fig 1A) that are distinguished by dramatic differences in behavior and physiology. While physiology during homeothermy resembles that of a non-hibernating mammal, squirrels spend most of their hibernation time in an energy-conserving state called torpor (Fig 1B, top right). Here, metabolic, respiratory and heart rates are

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

dramatically reduced to 1-9% of homeothermic baselines, while body temperature is lowered to near freezing [4]. However, torpor is not continuous, but instead punctuated by brief, metabolically intense, arousals that largely restore baseline physiology, including near-homeothermic body temperature [5,6]. Thus, hibernation is a period of heterothermy composed of cycles between torpor and arousal.

The seasonal transition from homeothermy to heterothermy occurs during the autumn of each year. Successful hibernation requires preparation, most notably the storage of large amounts of energy in the form of fat, because this species fasts throughout the heterothermic period. While post-reproduction homeothermy is marked by increased food intake, as the onset of heterothermy approaches, the squirrel's metabolic rate slows, peak body mass is achieved, and food intake ceases [7]. At the cellular level, glucose-based metabolism is switched to one that is primarily lipid-based, and lipogenesis is swapped for lipolysis [8]. While peak plasma insulin concentration occurs during this period, paradoxically, animals also become transiently insulin-resistant [9]. At the mRNA and protein levels, sweeping changes in expression are observed [10-14]. However, the genetic factors driving this transition remain largely unknown.

The commencement of torpor, defined by a criterion drop in body temperature, is one readily quantifiable outcome of the transition that marks the start of seasonal heterothermy. When housed under standard laboratory conditions in an animal facility (Fig 1B, bottom), 13-lined ground squirrels exhibit individual variation in the timing of their first bout of torpor. This variation is neither accounted for by environmental signals, such as food withdrawal, shortened photoperiod, or falling ambient temperature, nor by biological factors, such as age, body mass, and sex. All of these variables have little to modest influence on timing [15]. Rather, consistent with being controlled by an endogenous circannual clock, we hypothesized that observed variation in the onset of torpor is due to underlying genetic variation between individuals. If so, applying a genome-wide scan could potentially identify genetic components driving the start of seasonal heterothermy.

Therefore, in this study, we first increased the contiguity of the 13-lined ground squirrel draft genome assembly. We next employed a genotype-by-sequencing strategy to characterize genetic variation in 153 13-lined ground squirrels whose tissues were previously collected for use in transcriptomic, proteomic and biochemical studies [16-22]. Many of these squirrels were surgically implanted with body temperature dataloggers, and from their records, we recorded

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

the first day that torpor occurred in each individual (Fig 1C). We next estimated the heritability of, and identified genetic variants associated with, the onset of autumn torpor in this species. Finally, we integrated data from prior transcriptomic studies to identify transcripts whose expression levels were significantly associated with these variants. Our results present new insight into the genetics driving the transition from homeothermy to heterothermy and illustrate the power of genetic analysis to attack questions of exceptional biological significance in a non-classical genetic model organism.

Results

Long-range scaffolding of the draft genome assembly

At the time this study began, the existing 13-lined ground squirrel genome assembly (like that of many non-model organisms) contained thousands of unordered scaffolds, which could lead to difficulties in identifying causative variants, as peaks in linkage disequilibrium (LD) could be spread across multiple scaffolds. We therefore first sought to increase the genome's contiguity using a long-range scaffolding technique [23].

A single library was constructed using proximity ligation of in vitro reconstituted chromatin. After sequencing, which provided 52.6x physical coverage of the genome (Table 1) and scaffolding, the contiguity of the final HiRise assembly was increased approximately three-fold as compared to the existing draft assembly (N50 of 22.6Mb vs 8.19Mb; Table 1, Fig S1 and Table S1). The longest scaffold increased from 58.28Mb to 73.92Mb. Importantly, 539 original draft assembly scaffolds were reduced to just 33 scaffolds, which now contained half of the genome (Fig S2).

Table 1. Details of the draft assembly compared to the HiRise assembly.

Assembly details	Draft	HiRise
Average coverage	495.1x	52.6x
Total length (Mb)	2,478.4	2,478.4
Contigs		
No. of contigs	153,485	153,521
Contig N50 (kb)	44.137	44.131
Scaffolds		
No. of scaffolds (≥ 0)	12,483	10,007
Longest scaffold (Mb)	58.28	73.93
Scaffold N50 (Mb)	8.19	22.6
No. of scaffolds ≥ N50	80	33
Scaffold N90 (Mb)	1.13	3.33
Gaps		

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

Number of gaps	141,005	143,517
Percent of genome in gaps	6.75%	6.76%

137

138 *Identification of genetic variants*

139 We next applied a modified ddRAD sequencing protocol previously described in [24] to
140 generate libraries for 153 13-lined ground squirrels from which we obtained DNA from frozen
141 tissue. After aligning the resulting sample library reads to the HiRise genome assembly (Table
142 S2), we retained 337,695 loci (50.65 Mbp) that fell between predicted *BglII* and *DdeI* target
143 regions, with coverage of at least one read in one individual. Applying variant calling and
144 filtering to these loci (Fig S3), we next identified 786,453 biallelic variants, which had an overall
145 Ti/Tv ratio of 2.19, comparable to ratios reported within intronic and intergenic regions [25]. For
146 use in downstream analyses, we retained 575,178 variants for which genotypes were present
147 in at least 90% of the individuals. Of these retained variants, 35,257 were indels, whereas
148 539,921 were single nucleotide polymorphisms.

149

150 *Population Structure and Genetic Relatedness*

151 The squirrels genotyped in this study originated from wild stock trapped in disparate
152 geographical locales (Fig 1D). The records for their exact source and relatedness were not
153 always available. This was not due to intentional sampling design, but rather due to the
154 availability of squirrels each year, either trapped from the wild or supplied from a breeding
155 colony, and the biological questions originally being pursued. Therefore, to identify population
156 structure within our sample set, we applied ADMIXTURE clustering with 5-fold cross validation
157 [26] on K=2 through K=10 ancestral populations using a set of 54 unrelated individuals who
158 best represented the ancestries of all squirrels (Fig S4; see Methods). We then applied
159 ADMIXTURE projection to estimate proportions of learned ancestries within the remaining 99
160 squirrels. The lowest cross validation error occurred at K=3 (Fig 2A, top plot), where individuals
161 separated into Colorado (CO), Illinois (IL) and Wisconsin (WI) components. The pairwise
162 genetic distance (F_{ST}) estimates between populations were 0.47 and 0.31 for CO vs. WI and
163 IL, respectively, and 0.30 for WI vs. IL, indicating moderate to strong genetic drift. The
164 individual home range for a 13-lined ground squirrel is 0.01 – 0.05 km^2 [27]. Observed genetic
165 differences may simply be due to isolation by geographic distance [28].

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

166 At *K*=6, we observed separation most consistent with records about sampling (Fig 2A,
167 bottom plot). For instance, the algorithm identified a La Crosse, WI (LaX) ancestral component
168 for the squirrels supplied from the UW Oshkosh breeding colony in 2010, matching the
169 breeding records for that year. Additionally, the algorithm identified two ancestral components
170 for the IL squirrels: those purchased in 2006 ('06) belonged to a single ancestry, while those
171 from 2010 ('10) segregated into another ancestry, suggesting different trapping locales
172 between years. While records about the origins of the UW Oshkosh squirrels supplied prior to
173 2010 were unavailable, the algorithm identified two ancestral components for this breeding
174 colony. The pairwise F_{ST} values were still consistently high (0.33-0.48) among all populations
175 (Table S3), except for the two (non-LaX) within Oshkosh (0.16) and the two within IL (0.23),
176 again supporting the notion of limited gene flow at increased geographical distance.
177 The first three principal components (PCs) from a PCA-based analysis recapitulated
178 both the observed ADMIXTURE *K*=6 clustering and the known geographical sampling locales
179 of the squirrels (Fig 2B). All populations were distinctly separated, except for the two within
180 Oshkosh, whose separation was only observed at the higher PCs (PC17 and PC19, Fig S5).
181

182 *Genetic relatedness within the Oshkosh breeding colony*

183 Due to the strong population structure, and hence large differences in allele
184 frequencies, we limited further analysis to just the Oshkosh population of squirrels (not
185 including LaX, *n*=119), for which we were able to collect the most phenotypic measurements
186 (*n*=72, Table S4) from analysis of the body temperature telemetry data, as opposed to fewer
187 than *n*=10 phenotypic measurements in each of the remaining populations. Our existing
188 records from the breeding colony suggested that many of these squirrels were littermates,
189 although exact relatedness was unknown. We therefore estimated relatedness, adjusting for
190 population substructure with the first PC [29,30], which distinguished the two ancestral
191 components (Fig 2C). Using pedigree reconstruction [31], we identified 19 first-degree families,
192 to which 80% of the squirrels belonged. Consistent with our records, most of these families
193 were composed solely of littermates (Fig 2D, right plot), although in some cases we also
194 identified parent-offspring relationships (Fig 2D, left plot).
195

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

196 *Heritability Estimates for Timing of Autumn Torpor Immersion*

197 We next investigated the effect of genetic architecture on autumn torpor onset within the
198 Oshkosh subset of animals. We estimated heritability of this trait using a linear mixed model, in
199 which we controlled for sex, year of monitoring and date of placement into the hibernaculum
200 (our fixed effects, see methods), and we input the genetic relatedness estimates as the
201 random effect. Unexpectedly, this model converged with no residual error in the variance
202 components, resulting in an estimate of 100% heritability (LMM, Table 2). To confirm this high
203 estimate, we fit a separate Bayesian multivariate general linearized mixed model with the
204 same fixed and random effects. Here, the posterior mode of heritability was 99%, and the
205 confidence intervals were between 61% and 99.9% (MCMCgrm, Table 2); even the lower
206 bound of the estimate still indicated high heritability, confirming our hypothesis that underlying
207 genetic architecture drives variation in the onset of torpor.
208
209 **Table 2. Estimates of heritability for the timing of the first of torpor bout in autumn.**

Trait	Sample Size (*n*)	Method	Fixed Effects	Fixed Effects R^2	Genetic variance	Error variance	Genetic Proportion	C.I. (95%)
Timing of torpor onset	72	LM	Sex, year, T_a	0.2505	--	--	--	--
		LMM		--	57.06	0.00	1	1-1
		MCMCgrm		--	49.8	0.41	0.99	0.61-0.99

210 T_a, ambient temperature ≤ 14°C, denotes day of placement of into hibernaculum.
211
212 *Genome-wide Association Scan*

213 We identified genetic variants associated with the onset of autumn torpor by performing
214 a genome-wide association scan (GWAS) using 46,996 variants with a minor allele frequency
215 (MAF) ≥ 0.05 and the fit from the linear mixed model. As this was an exploratory analysis using
216 a relatively small sample set, we set a significance cut-off at $p \leq 5 \times 10^{-4}$. After accounting for
217 LD, we identified 21 loci that we considered significantly associated with the phenotype (Fig 3A
218 and Table 3). Although none of the variants met strict genome-wide significance after
219 *Bonferroni* correction ($p < 1 \times 10^{-6}$), a plot of the observed vs expected quantiles of log-
220 transformed p-values (Q-Q plot) showed an excess of significant values well above the dashed
221 line in the tail of the distribution (Fig 3B). Furthermore, while the estimated mean allelic effect
222 size was −0.04 days (SD=1.84, *n*=46,996), the effect sizes for these 21 significant variants

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

were all within either the top or bottom 1% of the total distribution, being at least ± 4.25 days for each additional allele (Fig 3C and Table 3).

Table 3. Details about the GWAS variants significantly associated with the onset of torpor

SNP #	Scaffold	Position	Ensembl Scaffold	Ensembl Position	Ref	Alt	MAF	P-value	β	Candidate Gene(s)	Function(s)
1	Scyvm7L_1371	290335	JH393559.1	288921	CA*	C	0.27	2.60E-04	-5.51	RASGEF1B	GTP signal transduction
2	Scyvm7L_2116	23692482	JH393326.1	1868477	C*	T	0.42	4.73E-05	4.27	MFAP3L; CLCN3	Elastic fiber formation; ion channel transport
3	Scyvm7L_471	7123085	JH393365.1	7103865	GT*	G	0.14	3.71E-04	5.64	NKX2-6	Cardiac embryonic development
4	Scyvm7L_100	29226048	JH393613.1	919871	G*	A	0.31	6.02E-05	-4.71	CDK6; SAMD9	Cell cycle regulation; regulation of cell proliferation and apoptosis
5	Scyvm7L_100	29569341	JH393613.1	1263164	G	T*	0.34	1.54E-04	4.86	FAM133B	Poly(A) RNA-binding
6	Scyvm7L_936	9851454	JH393624.1	94391	G	A*+	0.44	1.33E-04	-5.33	DKK2	Wnt signaling regulation
7	Scyvm7L_146	3229252	JH393286.1	25719306	A	G*	0.16	3.70E-05	-7.45	HLA-DPB1	Peptide antigen binding
8	Scyvm7L_146	3843670	JH393286.1	25104888	G*	A	0.38	7.79E-06	5.37	MLN	Regulation of GI contraction and hunger signaling
9	Scyvm7L_146	5440873	JH393286.1	23507685	G*	A	0.30	3.59E-04	-4.47	CLPS; MAPK14	Fat digestion and satiety; p38 signaling in response to stress
10	Scyvm7L_2912	2711274	JH393389.1	2499372	A	C*	0.46	3.90E-06	5.10	EXOC4; CHCHD3	Insulin response, glucose and lipid uptake; maintenance of mitochondrial cristae
11	Scyvm7L_1866	13078233	JH393463.1	779772	C	A*	0.13	4.91E-04	-6.76	KIAA1324l; GRM3	Unknown; glutamatergic neurotransmission
12	Scyvm7L_1164	849226	JH393548.1	1324176	C*	A	0.10	9.33E-05	-7.95	HPSE; SCD5	Wound healing and coagulation; fatty acid synthesis
13	Scyvm7L_1707	29969722	JH393295.1	2605622	G*	C	0.17	1.80E-04	-5.44	PCSK2	Glucose homeostasis, proinsulin and neuropeptide processing
14	Scyvm7L_607	25116481	JH393369.1	5809915	T	C*	0.22	4.39E-04	-4.75	ZNF462	Transcriptional regulation
15	Scyvm7L_301	9767851	JH393296.1	7909312	C	T*	0.14	3.81E-06	7.32	FAM204A; PRLHR	Unknown; feeding and energy homeostasis
16	Scyvm7L_301	17167980	JH393296.1	15309441	C	T*	0.19	2.77E-05	5.75	FOXI2; PTPRE	Transcriptional regulation; regulation of insulin signaling
17	Scyvm7L_30	2415084	JH393398.1	4077332	C*	A	0.44	4.72E-04	-5.17	Vav1; TRIP10	Immune response; insulin regulated lipid metabolism
18	Scyvm7L_30	2868918	JH393398.1	3623498	C	T*+	0.46	2.14E-04	4.25	ACSBG2	Fatty acyl-CoA biosynthesis
19	Scyvm7L_4270	16068966	JH393402.1	1874622	G	A*	0.48	6.19E-05	5.18	DGKI; PTN	phosphatidic acid production;

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

											Neurite outgrowth, adipogenesis
20	Scyvm7L_4270	16503923	JH393402.1	1439665	G	T*	0.37	3.20E-05	-5.08	CHRM2	Regulation of cardiac contractility
21	Scyvm7L_9858	615414	JH393647.1	629692	G*	A	0.37	4.92E-05	5.42	MUC21; DDR1	Cell adhesion; regulation of cell growth

Ref is the allele reported in the reference genome assembly
* Allele for which effect size is estimated; minor allele unless otherwise specified with +
+ Major allele
MAF is the minor allele frequency estimated from the *n*=72 genotypes used in the association scan
β is the effect size estimate in days.
Candidate gene(s) are genes nearest to variant; additional genes listed may be functionally related to hibernation.

We estimated the amount of phenotypic variance explained by the significant loci via linear regression. While the initial model fit with just the fixed effects of sex, year of monitoring and date of hibernaculum placement accounted for 25% of the variance in onset of torpor (Fig 4A), these 21 markers explained 54% of the variance (Fig 4B) and when combined with the fixed effects, accounted for 85% of the total variance in the phenotype (Fig 4C). Furthermore, the most significant GWAS variant (SNP 15, Table 3) alone accounted for 21% of phenotypic variance, while the subset of the top five most significant loci (SNPs 15, 10, 8, 16 and 20) explained 47.5% of the variance, excluding fixed effects. Hence, a small subset of markers accounted for most of the genetic component underlying the timing of autumn torpor in this population of 13-lined ground squirrels.

When we examined the genes located nearest these significant variants, many were functionally related to themes consistent with physiology underlying the transition to hibernation, such as insulin processing and signaling, feeding and satiety, and control of heart rate (Table 3). Intriguingly, the most significant variant, SNP 15, was located nearest the gene *family with sequence similarity 204 member A* (*FAM204A*), whose function is poorly characterized (Fig 5A). However, the *prolactin-releasing hormone receptor* (*PRLHR*), involved in stimulating prolactin release and control of feeding [32], and hence more consistent with roles in circannual timing and hibernation, was approximately 270kb from this marker. The second-most significant marker, SNP 10, was located between two genes that are also functionally relevant within the scope of hibernation: *coiled-coil-helix-coiled-coil-helix domain containing 3* (*CHCHD3*) and *exocyst complex component 4* (*EXOC4*; Fig 5B). While *CHCHD3* maintains the structural integrity of mitochondrial cristae [33], *EXOC4* is a component of the

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

exocyst complex involved in the secretion of insulin [34], as well as lipid and glucose uptake in response to insulin signaling [35,36]. Two weakly-linked variants (r^2=0.33) located approximately 500kb from each other, SNP 19 and SNP 20 (Figs 5C & 5D), were near *pleiotrophin* (*PTN*), a growth factor involved in neurogenesis and axonal outgrowth, angiogenesis and adipogenesis [37-39], and the *muscarinic acetylcholine receptor M2* (*CHRM2*), which mediates bradycardia in response to parasympathetic-induced acetylcholine release [40], a phenomenon well characterized in hibernation [41,42]. Finally, *motilin* (*MLN*), a small peptide hormone that regulates gastrointestinal contractions and stimulates hunger signaling [43] was nearest SNP 8 (Fig 5E), while *prohormone convertase 2* (*PCSK2*), an enzyme that activates hormones and neuropeptides [44-46], including cleavage of proinsulin into its mature form [47], was located in close proximity to SNP 13 (Fig 5F).

Identification of eQTLs using transcriptomic datasets

We hypothesized that these significant loci might be linked to gene regulatory variants. We therefore applied an expression quantitative loci (eQTL) analysis using the EDGE-tag transcript datasets from heart, liver, skeletal muscle (SkM) and brown adipose tissue (BAT) [48,49], since a subset of the squirrels genotyped in this study were assayed for transcriptome expression in these prior studies. Under an additive linear model, we identified significant *cis*-eQTL associations (±500kb, q<0.1) for 9/21 variants (Table 4). The most significant GWAS variant, SNP 15, was also the most significant *cis*-eQTL ($q<4.4X10^{-8}$), where the minor allele, associated with a later onset of torpor (Fig 6A, left plot), was correlated with increased expression of *FAM204A* in BAT (Fig 6A, middle-left and middle-right plots) and SkM. Several variants were associated with expression changes in the previously identified candidate genes (Table 3). SNP 19, associated with a later torpor onset (Fig 6B, left plot), was also correlated with decreased expression of *PTN* in BAT (Fig 6B, middle-left and middle-right plots), while SNP 20, associated with an earlier torpor onset (Fig 6C, left plot), was correlated with increased *CHRM2* expression in heart (Fig 6C, middle-left and middle-right plots). Two variants were correlated with changes in expression of the same transcript in the same direction across tissues: in heart and SkM, SNP 5 correlated with increased expression of *family with sequence similarity 133 member b* (*FAM133B)*, while SNP 21 correlated with decreased expression of *coiled-coil alpha-helical rod protein 1* (*CCHCR1*; Table 4). In contrast, several variants were associated with expression changes of different transcripts depending on

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

the tissue examined, suggesting their linkage to several regulatory loci within the region or to a shared regulatory site that exerts its effects on multiple nearby genes [50]. For example, SNP 2 was associated with decreased *chloride voltage-gated channel 3* (*CLCN3*) expression in SkM, yet also associated with increased *microfibril associated protein 3 like* (*MFAP3L*) expression in heart (Table 4).

Table 4. Details about the significant *cis*-eQTLs

SNP #	Tissue	Tag ID	Scaffold	Position	Ensembl Scaffold	Ensembl Position	Gene Symbol	Distance (kb)	p-value	q-value	β
2	Heart	Tag_68012	Scyvm7L_2116	23593820	JH393326.1	1967138	MFAP3L	98.7	5.5E-04	0.040	0.76
	SkM	Tag_32453		23802734		1758224	CLCN3	110.3	1.5E-04	0.014	-0.85
5	Heart	Tag_164503	Scyvm7L_100	29572046	JH393613.1	1265869	FAM133B	2.7	1.5E-07	3.1E-05	1.18
	SkM	Tag_78045							1.2E-03	0.071	0.91
8	BAT	Gene_92836	Scyvm7L_146	3382605	JH393286.1	25565906	TAPBP	461.1	2.2E-04	0.011	-0.73
	Heart	Tag_22155		3394532		25554025	ZBTB22	449.1	3.5E-04	0.038	-0.95
13	BAT	Gene_16216	Scyvm7L_1707	30247643	JH393295.1	2327672	LOC106144723	277.9	3.6E-05	3.0E-03	-0.88
15	BAT	Gene_99455	Scyvm7L_301	9806636	JH393296.1	7948097	FAM204A	38.8	1.8E-10	4.4E-08	1.56
	SkM	Tag_16792		9807010		7948471		39.2	1.6E-03	0.071	0.86
17	BAT	Gene_136773	Scyvm7L_30	2440903	JH393398.1	4051486	SH2D3A	25.8	1.4E-05	1.8E-03	-0.69
19	BAT	Gene_137408	Scyvm7L_4270	16236482	JH393402.1	1707078	PTN	167.5	1.1E-04	6.9E-03	-0.67
20	Heart	Tag_114532	Scyvm7L_4270	16416945	JH393402.1	1526642	CHRM2	87.0	1.0E-03	0.056	0.89
	BAT	Gene_79711		299681		299681	MRPS18B	315.7	2.6E-03	0.080	-0.66
21	Heart	Tag_167784	Scyvm7L_9858	810476	JH393647.1	810476	CCHCR1	195.1	1.4E-03	0.059	-0.89
	SkM	Tag_79577							1.1E-04	0.014	-0.97

See Table 3 for SNP #
Tag ID is from the original Edge-tag datasets in [48,49]
β is effect size estimated from quantile-normalized values

None of the variants met significance thresholds (q<0.1) to be identified as *trans*-eQTLs (Tables S5-S8); however, this was likely due to the relatively small sample sizes (*n*=22-23 in heart, SkM and liver; *n*=43 in BAT) and the large number of EDGE-tags tested (25-30K per tissue) in each dataset. We therefore examined the top *trans*-eGene (>500kb from variant) for each significant *cis*-eQTL, hypothesizing that we would identify genes within the same pathway or consistent with the physiology of the *cis*-eGene. Indeed, in heart, the *angiotensin II receptor type 1*, *AGTR1*, involved in regulation of blood pressure [51], was the top *trans*-eGene for SNP 20 ($p=2.97X10^{-5}$, Table S5). In contrast to *CHRM2*, this transcript showed decreased expression in relation to the minor allele (Fig 6C, right plot), consistent with the physiology of

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

311 torpor, where reduced heart rate is coupled with decreased blood pressure [52,53]. In BAT, the
312 top *trans*-eGene for SNP 19 was *FER Tyrosine Kinase* (*FER*; $p=4.73X10^{-5}$, Table S8), whose
313 expression, like *PTN*, was also decreased in relation to the minor allele of its SNP (Fig 6B,
314 right plot). Both *PTN* and *FER* phosphorylate β-catenin [54,55], suggesting a role for this
315 pathway in the start of torpor. Finally in BAT, the top *trans*-eGene for SNP 15 was *PH domain*
316 *and leucine rich repeat protein phosphatase 1* (*PHLPP1*; $p=2.1X10^{-5}$, Table S8), whose
317 expression, like *FAM204A*, was increased in relation to the variant (Fig 6A, right plot). *PHLPP1*
318 is a protein phosphatase that dephosphorylates and inactivates both *Akt2* and *protein kinase C*
319 [56]. *Akt2* is expressed highly in insulin-responsive tissues, including BAT, where it modulates
320 glucose uptake and homeostasis [57], as well as non-shivering thermogenesis (NST) [58].
321 Moreover, increased *PHLPP1* expression is associated with insulin-resistance and
322 hyperinsulinaemia [59]; hence, this gene may play a role in the switch from glucose to fat-
323 based metabolism that occurs at the onset of seasonal heterothermy, and more specifically in
324 BAT, the regulation of NST.
325
326 **Discussion**
327 Mammalian hibernation is a highly dynamic and extraordinary phenotype that remains
328 poorly understood. While it has been characterized at behavioral, whole body, cellular, and
329 molecular levels, a genetic basis of the phenotype has yet to be established. Our study is the
330 first, to our knowledge, to characterize genome-wide variation within a hibernator, the 13-lined
331 ground squirrel. This enabled us to estimate the heritability of, and identify genetic variants
332 associated with, the onset of seasonal heterothermy.
333 Our results of heritability are consistent with those from a study that reported significant
334 heritability in spring emersion from hibernation in wild Columbian ground squirrels [60].
335 However, our estimates for immersion into hibernation are much higher. This is likely due to
336 differences between monitoring animals in an animal facility, where environmental conditions
337 and access to food are relatively constant and social cues are minimal, and monitoring animals
338 in the field, where phenotypic plasticity in response to changing environmental conditions also
339 influences phenological timing [61,62]. In the wild, differences in age (adult vs juvenile) and
340 hibernation timing have been observed [61]. While age did not significantly affect hibernation
341 onset in our study, it is worth noting that 74% of the squirrels in this dataset were juveniles,
342 and therefore hibernation-naïve prior to their first torpor bout. Juveniles likely face a far greater

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

challenge in growing and fattening sufficiently to support winter hibernation in the wild; in contrast, in a relatively constant, resource-rich laboratory environment, hibernation onset for these animals may be particularly driven by endogenous mechanisms, thus increasing heritability estimates.

In contrast to complex human diseases, where often thousands of variants of relatively small effect influence complex physiological phenotype [63], we find that relatively few loci of large effect account for phenotypic differences in the seasonal onset of hibernation. Our results are comparable to what has been observed for adaptive traits in *A. thaliana* [64], morphological variation in domesticated dogs [65] and plate armor of threespine sticklebacks [66]. This may be simply due to limited sample size, as we were underpowered to detect variants with small effect size and/or at low frequency. However, the results from our ADMIXTURE analysis suggest that wild-trapped founders from two distinct populations were crossed to form the breeding colony. One possible explanation is that torpor onset differs between these two populations, and if so, local genetic adaptation in each could explain the relatively few loci of large effect driving phenotypic variation [67].

Although more research is now needed to determine the precise role in torpor immergence for each of these loci, we propose candidate genes, oftentimes the closest to the marker, due to their function being closely related to the physiology of this seasonal transition. In particular, several genes are known to modulate food intake, such as the *prolactin-releasing hormone receptor* (*PRLHR*), *motilin* (*MLN*), and *procolipase* (*CLPS*) [68]. Researchers have hypothesized that mechanisms governing food intake and metabolic suppression are linked, and that hibernation cannot begin until food intake has ceased [7,69]. Our results support this hypothesis, and present new candidates for study in hibernation.

Perhaps the most intriguing candidate gene is *PRLHR*. Sharing a common ancestry with the *NPY* receptors [70], this receptor is expressed primarily in the anterior pituitary, as well as in distinct regions of the brain, including the hypothalamus. Its knockout in mice results in an obese, hyperphagic phenotype, while administration of its agonist, prolactin-releasing hormone (*PRLH*), induces hypophagia and decreases body mass [32]. Both *PRLH* and *PRLHR* appear to mediate the effects of leptin, including activation of NST in BAT [71]. Further, *PRLH* belongs to a class of neuropeptides containing a C-terminal RFamide motif [72]; other RFamide-related peptides are involved in circannual regulation of reproduction [73,74]. Binding of *PRLH* to *PRLHR* may induce prolactin release, a hormone that has

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

established roles in the timing of circannual rhythms, such as seasonal molt [75] and reproduction [76]. Moreover, serum prolactin levels coincide with resumption of post-hibernation feeding in marmots [77]. Thus, we hypothesize that *PRLHR* is involved in the regulation of circannual food intake in hibernators; its role in this rhythm warrants further investigation.

However, we note that the most significant *cis*-eGene for SNP 15 was not *PRLHR*, but rather its near neighbor, *FAM204A*. This may be due to the marker being linked to several regulatory sites within the region or to a single regulatory site that affects multiple nearby genes. Additionally, *PRLHR*, along with other candidate genes, was not expressed in the tissues for which we had transcriptome data, a limiting factor in our analysis. A role for *FAM204A* in hibernation is unclear, as not much is known about this gene. It appears to be expressed in every tissue and localizes to the nucleus (www.proteinatlas.org [78]), where it interacts with a histone acetyltransferase and a methyltransferase [79]. Therefore, it may play a role in epigenetic regulation of gene expression, possibly in response to the onset of fasting and a switch to fatty acid metabolism [80], as animals prepare for hibernation in the fall.

Finally, our results highlight the power of integrating genome data with transcriptome and other high-throughput data to better understand the genetic mechanisms underlying hibernation. Prior “omics” screens (e.g. [10,12,13,22,48,49,81]) have identified hundreds to thousands of genes differentially expressed among the seasonal and physiological states of the hibernator’s year. While leading to insight into the pathways involved, the results of these screens do not distinguish between genes driving vs. those responding to changes in phenotype. They are also limited to the tissue and time-points being examined and may therefore miss important regulators of phenotype. It is worth noting that while several *cis*- and *trans*-eGenes identified here have clear roles in the physiology of torpor, such as *CHRM2* and *AGTR1*, neither of these were identified in their original studies that screened only for differential expression. Thus, by applying complementary genetic mapping approaches, current limitations inherent to gene expression screening strategies will be addressed and enable new insight into the mechanisms driving hibernation. The approaches used here can be extended to a wide variety of hibernators and the quantifiable components that comprise this highly dynamic phenotype.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

Materials and Methods

Animals

Animals were procured and housed at the University of Colorado, Anschutz Medical Campus, as previously described [15]. All animal use was approved by the University of Colorado, Anschutz Medical Campus, Animal Care and Use Committee.

Briefly, 130 colony-bred animals (68 females and 62 males) were obtained from the University of Wisconsin, Oshkosh [82] in the summers of 2007—2010 (Fig 1D, "WI"). These included 73 juveniles naïve to hibernation in the year of study, and 57 adults with at least one year of hibernation. While most from the colony were bred from squirrels originally wild-trapped in northeastern Wisconsin (in and around Oshkosh), several of those obtained from the Oshkosh colony in 2010 were actually bred from either a single or both parents wild-trapped in far western Wisconsin, more than 100 miles away (in and around La Crosse, WI). However, records to identify these specific squirrels were not always maintained. In addition, 17 squirrels (9 females and 8 males; ages unknown), wild-trapped in different locales around central Illinois, were obtained from a commercial supplier (TLS Research, Bloomington, IL) in the summers of 2006 and 2010 (Fig 1D, "IL"). Finally, 6 squirrels (3 females, 3 males; ages unknown) were wild-trapped in the summers of 2006 and 2009 in Elbert County and Larimer County, Colorado (5 and 1, respectively; Fig 1D, "CO").

Upon arrival, animals were housed individually in rodent cages (Fig 1B, bottom) under standard laboratory conditions (20±2°C and 14:10 light-dark cycle, fed cat chow supplemented with sunflower seeds *ad libitum*). In late August or early September, animals not yet euthanized for tissue collection were surgically implanted with an intraperitoneal datalogger (iButton, Embedded Data Systems) and/or a radiotelemeter (VM-FH disks; Mini Mitter, Sunriver, OR) for remote body temperature (T_b) monitoring until tissue collection. The dataloggers recorded T_b±0.5°C every 20, 30 or 60 min, while the radiotelemeters transmitted T_b±0.5°C every 20 sec.

In late September or early October, the squirrels were moved to the hibernaculum to facilitate hibernation. The temperature was lowered stepwise over a two-week period to 4°C. Food was removed as animals became torpid.

Tissue Collection and Telemeter Retrieval

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

Liver samples were collected at different points throughout the year for use in other biochemical studies as previously described [19,81,83,84]. All animals were exsanguinated under isoflurane anesthesia, perfused with ice-cold saline, decapitated, and dissected on ice; tissues were immediately snap frozen in liquid nitrogen and stored at -80°C until processed further. Telemeters were retrieved during tissue collection.

Body Temperature Telemetry Analysis

To identify the first day of torpor, the telemetry data were analyzed in R [85]. T_b was averaged over 4-hour windows. Homeothermic T_b typically ranged from 34-39°C. The first torpor bout was defined as the first point at which T_b fell to or below 25°C (approximately 3-5°C above ambient prior to hibernaculum placement, Fig 1C). Most telemetry data continuously logged T_b from the beginning of September of each year. However, in several cases, telemetry recordings did not start until mid-September. Of these, only cases in which first torpor occurred after a minimum of 10 days of continuous monitoring were included for further analysis. In order to merge data across years, date of first torpor was transformed into date from placement into the hibernaculum, which was centered as day 0; hence, all days prior are negative in value, while post-placement dates are positive.

HiRise Genome Assembly and Annotation

A Chicago library was prepared as described previously [23] from a single 100mg frozen liver sample. Briefly, ~500ng of high molecular weight gDNA (mean fragment length = >50kbp) was reconstituted into chromatin in vitro and fixed with formaldehyde. Fixed chromatin was digested with *DpnII*, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of the library. The library was sequenced on an Illumina HiSeq 2500 (rapid run mode) to produce 150 million 2x101bp paired end reads, which provided 52.6x physical coverage of the genome (1-50kb pairs).

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

The 13-lined ground squirrel draft assembly, shotgun reads, and Chicago library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies [23]. Shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (http://snap.cs.berkeley.edu). The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. After scaffolding, shotgun sequences were used to close gaps between contigs. Table S1 describes the input draft assembly scaffold placement within the HiRise scaffolds.

Gene annotations from the Ensembl (Release 86) and NCBI (Release 101) datasets were lifted over to the HiRise assembly using a custom Python script and Table S1.

Genotype-by-sequencing

We used the modified ddRAD sequencing protocol previously described in [24]. Briefly, high molecular weight DNA was extracted from 8-15 mg of frozen liver with commercially available kits. Digestion and ligation reactions were performed using 200ng of genomic DNA from each sample with *BglII* and *DdeI* and 11-fold excess of sequencing adaptors. Samples were amplified by PCR for 8-12 cycles with a combination of index-containing primers. Between 50-60 samples were pooled in equal amounts according to their concentration of PCR product between 280-480 bp as measured by Tapestation (Agilent Technologies, Santa Clara, CA). Inserts were size-selected on a BluePippin (Sage Science, Beverly, MA) with a target range of 380±100bp and sequenced on the Illumina NextSeq in single-end 151 bp mode using a high output kit.

Variant calling and Filtering

Reads were mapped to the 13-lined ground squirrel HiRise assembly with BWA v. 0.7.12. [86]. Tables of predicted *BglII* and *DdeI* restriction digest fragments were generated as described in [24], and sequencing coverage was measured at these sites. We then defined “target regions” for variant calling using the set of fragments between 125-350 bp long that had non-zero coverage in at least one individual. The mapping data are summarized in Table S2.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

Because publicly available data on 13-lined ground squirrel genetic variation is non-existent, we instead used several variant callers to identify genetic variants and to assess concordance of the genotype calls at each site. Variant calling was performed independently with Sentieon [87], Platypus [88] and Samtools [89,90]. In Sentieon, the pipeline algorithms indel realignment, base quality score recalibration, haplotyper and GVCFtyper were implemented with default settings. In Samtools, variants were called jointly using mpileup to first compute genotype likelihoods and then BCFtools to call genotypes with default parameters. Finally, variants were called jointly in Platypus with the following parameters: minFlank=3, badReadsWindow=5, maxVariants=12, and minReads=6. Only biallelic variants that both passed the filter flags and were identified by all three callers were retained (Fig S3). These variants were next intersected with GATK [91] and compared for genotype call concordance across samples [92]. Those that were ≥ 95% concordant were kept. Basic statistics about the callset, including depth, missingness, heterozygosity, Hardy-Weinberg equilibrium, and TiTv ratio were calculated in VCFtools [93]. Variants with excessive coverage (≥ 65X, approx. 4x the mean coverage, Table S2) and heterozygosity (obs./exp. ratio ≥ 1.2) were removed from the callset (Fig S3). Sample libraries with excessive missingness and/or heterozygosity were removed, remade, and resequenced. Variants then were reiteratively called and filtered as described above. Finally, variants present in ≥ 90% of the sample libraries were used for further downstream analyses.

Population Structure and Genetic Relatedness Estimates

We first inferred relatedness from identity-by-state (IBS) estimates among all genotyped squirrels (*n*=153) using KING software [94]. Due to the expectation that the Colorado squirrels are of a separate subspecies [95], relatedness was calculated independently for this subset. We selected an unrelated subset of 54 squirrels that best represented the ancestries of all squirrels within the dataset using the GENESIS package [96] in R [85]. Variants were pruned for LD in PLINK v. 1.9 [97] using the parameters --indep-pairwise 50 10 0.5, which reduced the dataset to 148,870 variants. We then ran unsupervised ADMIXTURE [26] for K=3 through K=10 with 5-fold cross-validation. To estimate the ancestries of the remaining 99 squirrels, we ran ADMIXTURE's projection analysis using the population structure learned in the initial unsupervised analysis, here with K=2 through K=8 and 5-fold cross-validation.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

We performed principal components analyses (PCA) with PLINK using 90,376 LD pruned variants with MAF > 0.01 for the entire dataset and 30,356 LD pruned variants with MAF > 0.01 for the squirrels within the Oshkosh, WI, population (*n*=119), as identified by ADMIXTURE analysis. We extracted the top 20 principal components in each analysis.

Finally, we calculated genetic relatedness among the 119 Oshkosh WI squirrels using the GENESIS package, adjusting for both population substructure and inbreeding with the first principal component [29]. We used the resulting kinship coefficients and identity-by-descent (IBD) estimates to reconstruct and visualize pedigrees among the 1st degree relatives with PRIMUS [31]. We also constructed a genetic relatedness matrix from the pairwise kinship coefficients.

Genome-Wide Association Scan and Heritability Estimates

All analyses, unless otherwise stated, were performed in R [85]. To identify environmental and biological factors that affected the date of first torpor, we applied a linear regression using variables available from records about the squirrels. In this initial model:

Date of first torpor = f(sex + year of monitoring + date of datalogger implantation + age (juvenile vs. adult) + date of placement into hibernaculum + weight (as last recorded before placement into hibernaculum).

We then pruned factors using step-wise regression until we identified a minimum set that did not significantly reduce the adjusted R-squared value from the initial model, yet also returned a low AIC value. In this final model:

Date of first torpor = f(sex + year of monitoring + date of placement into hibernaculum).

These were our fixed effects.

We carried out a genome-wide association scan (GWAS) on the date of first torpor using GENESIS [96]. We first fit a linear mixed model using the fixed effects and the genetic relatedness matrix as the random effect. We then performed SNP genotype association tests with 46,996 SNPS (MAF≥0.05) and the fit from the linear mixed model. As this was an exploratory analysis, we considered any variant with $p \leq 5 \times 10^{-4}$ to be significantly associated with the phenotype. To account for LD, We calculated the r^2 values for significant SNPs within the same scaffold using PLINK [97,98]. We removed those in moderate to high LD ($r^2 \geq 0.5$), reporting only the most significant variant.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

We estimated heritability of the first day of torpor from the variance components of the linear mixed model. In addition, we also estimated heritability of this phenotype using a separate Bayesian mixed model with the MCMCgrm package in R [99]. Here we input the same fixed and random effects (i.e. genetic relatedness matrix) as in the linear mixed model. For the prior, we used an uninformative inverse-gamma distribution (with variance, V, set to 1 and belief parameter, nu, set to 0.002) on the variance components. We ran three chains, each with a total of 1,000,000 iterations, a burnin of 100,000 rounds and a thinning interval of 200 rounds. Here, all variables had Gelman-Rubin statistics of 1.00 – 1.01, with the absolute value of all autocorrelations < 0.1 and effective sample sizes between 3682.8 and 5294.5. We combined the 3 chains in order to estimate the posterior mode and confidence intervals of the variance components.

Finally, we estimated the effects of the significant GWAS variants on the onset of torpor. Specifically, we used linear regression with the phenotype as the dependent variable and a matrix of significant variant genotypes, either with or without the fixed effects, as the explanatory variables. We also performed forward stepwise regression using genotype combinations from the top 10 significant variants.

EQTL analysis

We applied an eQTL analysis to identify transcripts whose expression levels were significantly affected by the GWAS variants. Here, we used the EDGE-tag datasets from heart, liver, skeletal muscle (SkM) and brown adipose tissue (BAT) previously described in [48,49], where total RNA was digested with *NlaIII*, resulting in the generation of ≈27-nt "EDGE-tags" [100], which mapped to the 3'UTR's of transcripts. The squirrels assayed in these prior transcriptome studies were also genotyped in this study: heart (n=22), liver (n=23), SkM (n=22) and BAT (n=43). Due to small sample sizes, we limited our eQTL association tests to significant GWAS SNPS with MAF ≥ 0.2 in heart, liver and SkM and ≥ 0.1 in BAT, which ensured that a minimum of 9 samples contained at least one minor allele.

Tests for both *cis*- (±500kb) and *trans*-eQTLs were performed with Matrix eQTL [101] under an additive linear model. As the purpose of the original EDGE-tag studies was to identify differentially expressed transcripts among distinct physiological states within the hibernator's year, we included physiological state as a covariate (5 states in heart, liver and SkM: spring cold, SpC; summer active, SA; interbout-aroused in hibernation, IBA; entering torpor in

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

hibernation, En*t*; and late torpor in hibernation, LT; 9 states in BAT, in addition to the those previously mentioned: spring warm, SW; fall transiton, FT; early torpor in hibernation, ET; early in arousal from torpor in hibernation, EAr). In BAT, sequencing platform was also included due to the count bias observed in [49]. Finally, to control for outliers and following recommendations by Matrix eQTL, the counts for each Edge-tag were ranked and quantile-normalized before testing.

Acknowledgements

We thank members of the Bustamante Lab for their helpful discussion while this research was being conducted. We also thank R. Russell, A. Hindle and members of the Martin lab who assisted with the 13-lined ground squirrel care, surgical implantation of dataloggers and collection of tissue.

C. D. Bustamante is a Chan Zuckerberg Biohub investigator.

Funding

This work was supported by the National Science Foundation grant 1642184 to C. D. Bustamante and the National Institutes of Health grant R01HL089049 to S. L. Martin. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Dark J (2005) Annual lipid cycles in hibernators: integration of physiology and behavior. Annu Rev Nutr 25: 469-497.
2. Pengelley ET (1968) Interrelationships of circannian rhythms in the ground squirrel, Citellus lateralis. Comp Biochem Physiol 24: 915-919.
3. Pengelley ET, Asmundson SJ, Barnes B, Aloia RC (1976) Relationship of light intensity and photoperiod to circannual rhythmicity in the hibernating ground squirrel, Citellus lateralis. Comp Biochem Physiol A Comp Physiol 53: 273-277.
4. Ruf T, Geiser F (2015) Daily torpor and hibernation in birds and mammals. Biol Rev Camb Philos Soc 90: 891-926.
5. Carey HV, Andrews MT, Martin SL (2003) Mammalian hibernation: cellular and molecular responses to depressed metabolism and low temperature. Physiol Rev 83: 1153-1181.
6. Hampton M, Nelson BT, Andrews MT (2010) Circulation and metabolic rates in a natural hibernator: an integrative physiological model. Am J Physiol Regul Integr Comp Physiol 299: R1478-1488.
7. Florant GL, Healy JE (2012) The regulation of food intake in mammalian hibernators: a review. J Comp Physiol B 182: 451-467.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

8. Lanaspa MA, Epperson LE, Li N, Cicerchi C, Garcia GE, et al. (2015) Opposing activity changes in AMP deaminase and AMP-activated protein kinase in the hibernating ground squirrel. PLoS One 10: e0123509.
9. Martin SL (2008) Mammalian hibernation: a naturally reversible model for insulin resistance in man? Diab Vasc Dis Res 5: 76-81.
10. Hampton M, Melvin RG, Andrews MT (2013) Transcriptomic analysis of brown adipose tissue across the physiological extremes of natural hibernation. PLoS One 8: e85157.
11. Schwartz C, Hampton M, Andrews MT (2013) Seasonal and regional differences in gene expression in the brain of a hibernating mammal. PLoS One 8: e58427.
12. Hampton M, Melvin RG, Kendall AH, Kirkpatrick BR, Peterson N, et al. (2011) Deep sequencing the transcriptome reveals seasonal adaptive mechanisms in a hibernating mammal. PLoS One 6: e27021.
13. Vermillion KL, Anderson KJ, Hampton M, Andrews MT (2015) Gene expression changes controlling distinct adaptations in the heart and skeletal muscle of a hibernating mammal. Physiol Genomics 47: 58-74.
14. Grabek KR, Martin SL, Hindle AG (2015) Proteomics approaches shed new light on hibernation physiology. J Comp Physiol B 185: 607-627.
15. Russell RL, O'Neill PH, Epperson LE, Martin SL (2010) Extensive use of torpor in 13-lined ground squirrels in the fall prior to cold exposure. J Comp Physiol B 180: 1165-1172.
16. Hindle AG, Martin SL (2014) Intrinsic circannual regulation of brown adipose tissue form and function in tune with hibernation. Am J Physiol Endocrinol Metab 306: E284-299.
17. Hindle AG, Martin SL (2013) Cytoskeletal regulation dominates temperature-sensitive proteomic changes of hibernation in forebrain of 13-lined ground squirrels. PLoS One 8: e71627.
18. Grabek KR, Karimpour-Fard A, Epperson LE, Hindle A, Hunter LE, et al. (2011) Multistate proteomics analysis reveals novel strategies used by a hibernator to precondition the heart and conserve ATP for winter heterothermy. Physiol Genomics 43: 1263-1275.
19. Epperson LE, Rose JC, Carey HV, Martin SL (2010) Seasonal proteomic changes reveal molecular adaptations to preserve and replenish liver proteins during ground squirrel hibernation. Am J Physiol Regul Integr Comp Physiol 298: R329-340.
20. Epperson LE, Rose JC, Russell RL, Nikrad MP, Carey HV, et al. (2010) Seasonal protein changes support rapid energy production in hibernator brainstem. J Comp Physiol B 180: 599-617.
21. Epperson LE, Karimpour-Fard A, Hunter LE, Martin SL (2011) Metabolic cycles in a circannual hibernator. Physiol Genomics 43: 799-807.
22. Hindle AG, Grabek KR, Epperson LE, Karimpour-Fard A, Martin SL (2014) Metabolic changes associated with the long winter fast dominate the liver proteome in 13-lined ground squirrels. Physiol Genomics 46: 348-361.
23. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, et al. (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res 26: 342-350.
24. Cooke TF, Fischer CR, Wu P, Jiang TX, Xie KT, et al. (2017) Genetic Mapping and Biochemical Basis of Yellow Feather Pigmentation in Budgerigars. Cell 171: 427-439 e421.
25. Wang J, Raskin L, Samuels DC, Shyr Y, Guo Y (2015) Genome measures used for quality control are dependent on gene function and ancestry. Bioinformatics 31: 318-323.
26. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19: 1655-1664.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

27. Kurta A, Burt WH (1995) Mammals of the Great Lakes region. Ann Arbor: University of Michigan Press. xii, 376 p. p.
28. Slatkin M (1993) Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution 47: 264-279.
29. Conomos MP, Reiner AP, Weir BS, Thornton TA (2016) Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet 98: 127-148.
30. Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, et al. (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics 28: 3329-3331.
31. Staples J, Qiao D, Cho MH, Silverman EK, University of Washington Center for Mendelian G, et al. (2014) PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. Am J Hum Genet 95: 553-564.
32. Dodd GT, Luckman SM (2013) Physiological Roles of GPR10 and PrRP Signaling. Front Endocrinol (Lausanne) 4: 20.
33. Darshi M, Mendiola VL, Mackey MR, Murphy AN, Koller A, et al. (2011) ChChd3, an inner mitochondrial membrane protein, is essential for maintaining crista integrity and mitochondrial function. J Biol Chem 286: 2918-2932.
34. Tsuboi T, Ravier MA, Xie H, Ewart MA, Gould GW, et al. (2005) Mammalian exocyst complex is required for the docking step of insulin vesicle exocytosis. J Biol Chem 280: 25565-25570.
35. Inoue M, Akama T, Jiang Y, Chun TH (2015) The exocyst complex regulates free fatty acid uptake by adipocytes. PLoS One 10: e0120289.
36. Inoue M, Chang L, Hwang J, Chiang SH, Saltiel AR (2003) The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. Nature 422: 629-633.
37. Wong JC, Krueger KC, Costa MJ, Aggarwal A, Du H, et al. (2016) A glucocorticoid- and diet-responsive pathway toggles adipocyte precursor cell activity in vivo. Sci Signal 9: ra103.
38. Perez-Pinera P, Berenson JR, Deuel TF (2008) Pleiotrophin, a multifunctional angiogenic factor: mechanisms and pathways in normal and pathological angiogenesis. Curr Opin Hematol 15: 210-214.
39. Gonzalez-Castillo C, Ortuno-Sahagun D, Guzman-Brambila C, Pallas M, Rojas-Mayorquin AE (2014) Pleiotrophin as a central nervous system neuromodulator, evidences from the hippocampus. Front Cell Neurosci 8: 443.
40. Fisher JT, Vincent SG, Gomeza J, Yamada M, Wess J (2004) Loss of vagally mediated bradycardia and bronchoconstriction in mice lacking M2 or M3 muscarinic acetylcholine receptors. FASEB J 18: 711-713.
41. Milsom WK, Zimmer MB, Harris MB (1999) Regulation of cardiac rhythm in hibernating mammals. Comp Biochem Physiol A Mol Integr Physiol 124: 383-391.
42. Lyman CP (1982) Hibernation and torpor in mammals and birds. New York: Academic Press. x, 317 p. p.
43. Tack J, Deloose E, Ang D, Scarpellini E, Vanuytsel T, et al. (2016) Motilin-induced gastric contractions signal hunger in man. Gut 65: 214-224.
44. Toorie AM, Cyr NE, Steger JS, Beckman R, Farah G, et al. (2016) The Nutrient and Energy Sensor Sirt1 Regulates the Hypothalamic-Pituitary-Adrenal (HPA) Axis by Altering the Production of the Prohormone Convertase 2 (PC2) Essential in the Maturation of Corticotropin-releasing Hormone (CRH) from Its Prohormone in Male Rats. J Biol Chem 291: 5844-5859.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

45. Taylor NA, Van De Ven WJ, Creemers JW (2003) Curbing activation: proprotein convertases in homeostasis and pathology. FASEB J 17: 1215-1227.
46. Scriba MF, Ducrest AL, Henry I, Vyssotski AL, Rattenborg NC, et al. (2013) Linking melanism to brain development: expression of a melanism-related gene in barn owl feather follicles covaries with sleep ontogeny. Front Zool 10: 42.
47. Smeekens SP, Montag AG, Thomas G, Albiges-Rizo C, Carroll R, et al. (1992) Proinsulin processing by the subtilisin-related proprotein convertases furin, PC2, and PC3. Proc Natl Acad Sci U S A 89: 8822-8826.
48. Bogren LK, Grabek KR, Barsh GS, Martin SL (2017) Comparative tissue transcriptomics highlights dynamic differences among tissues but conserved metabolic transcript prioritization in preparation for arousal from torpor. J Comp Physiol B 187: 735-748.
49. Grabek KR, Diniz Behn C, Barsh GS, Hesselberth JR, Martin SL (2015) Enhanced stability and polyadenylation of select mRNAs support rapid thermogenesis in the brown fat of a hibernator. Elife 4.
50. Tong P, Monahan J, Prendergast JG (2017) Shared regulatory sites are abundant in the human genome and shed light on genome evolution and disease pleiotropy. PLoS Genet 13: e1006673.
51. Ito M, Oliverio MI, Mannon PJ, Best CF, Maeda N, et al. (1995) Regulation of blood pressure by the type 1A angiotensin II receptor gene. Proc Natl Acad Sci U S A 92: 3521-3525.
52. Horwitz BA, Chau SM, Hamilton JS, Song C, Gorgone J, et al. (2013) Temporal relationships of blood pressure, heart rate, baroreflex function, and body temperature change over a hibernation bout in Syrian hamsters. Am J Physiol Regul Integr Comp Physiol 305: R759-768.
53. Lyman CP, O'Brien RC (1963) Autonomic Control of Circulation during the Hibernating Cycle in Ground Squirrels. J Physiol 168: 477-499.
54. Piedra J, Miravet S, Castano J, Palmer HG, Heisterkamp N, et al. (2003) p120 Catenin-associated Fer and Fyn tyrosine kinases regulate beta-catenin Tyr-142 phosphorylation and beta-catenin-alpha-catenin Interaction. Mol Cell Biol 23: 2287-2297.
55. Meng K, Rodriguez-Pena A, Dimitrov T, Chen W, Yamin M, et al. (2000) Pleiotrophin signals increased tyrosine phosphorylation of beta beta-catenin through inactivation of the intrinsic catalytic activity of the receptor-type protein tyrosine phosphatase beta/zeta. Proc Natl Acad Sci U S A 97: 2603-2608.
56. Brognard J, Newton AC (2008) PHLiPPing the switch on Akt and protein kinase C signaling. Trends Endocrinol Metab 19: 223-230.
57. Bouzakri K, Zachrisson A, Al-Khalili L, Zhang BB, Koistinen HA, et al. (2006) siRNA-based gene silencing reveals specialized roles of IRS-1/Akt2 and IRS-2/Akt1 in glucose and lipid metabolism in human skeletal muscle. Cell Metab 4: 89-96.
58. Albert V, Svensson K, Shimobayashi M, Colombi M, Munoz S, et al. (2016) mTORC2 sustains thermogenesis via Akt-induced glucose uptake and glycolysis in brown adipose tissue. EMBO Mol Med 8: 232-246.
59. Andreozzi F, Procopio C, Greco A, Mannino GC, Miele C, et al. (2011) Increased levels of the Akt-specific phosphatase PH domain leucine-rich repeat protein phosphatase (PHLPP)-1 in obese participants are associated with insulin resistance. Diabetologia 54: 1879-1887.
60. Lane JE, Kruuk LE, Charmantier A, Murie JO, Coltman DW, et al. (2011) A quantitative genetic analysis of hibernation emergence date in a wild population of Columbian ground squirrels. J Evol Biol 24: 1949-1959.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

773 61. Sheriff MJ, Kenagy GJ, Richter M, Lee T, Toien O, et al. (2011) Phenological variation in
774 annual timing of hibernation and breeding in nearby populations of Arctic ground
775 squirrels. Proc Biol Sci 278: 2369-2375.
776 62. Sheriff MJ, Richter MM, Buck CL, Barnes BM (2013) Changing seasonality and
777 phenological responses of free-living male arctic ground squirrels: the importance of
778 sex. Philos Trans R Soc Lond B Biol Sci 368: 20120480.
779 63. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the
780 missing heritability of complex diseases. Nature 461: 747-753.
781 64. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al. (2010) Genome-wide
782 association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465:
783 627-631.
784 65. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, et al. (2010) A simple
785 genetic architecture underlies morphological variation in dogs. PLoS Biol 8: e1000451.
786 66. Colosimo PF, Peichel CL, Nereng K, Blackman BK, Shapiro MD, et al. (2004) The genetic
787 architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biol 2:
788 E109.
789 67. Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. Nat
790 Rev Genet 14: 807-820.
791 68. Erlanson-Albertsson C, York D (1997) Enterostatin--a peptide regulating fat intake. Obes
792 Res 5: 360-372.
793 69. Davis DE (1976) Hibernation and circannual rhythms of food consumption in marmots and
794 ground squirrels. Q Rev Biol 51: 477-514.
795 70. Lagerstrom MC, Fredriksson R, Bjarnadottir TK, Fridmanis D, Holmquist T, et al. (2005)
796 Origin of the prolactin-releasing hormone (PRLH) receptors: evidence of coevolution
797 between PRLH and a redundant neuropeptide Y receptor during vertebrate evolution.
798 Genomics 85: 688-703.
799 71. Dodd GT, Worth AA, Nunn N, Korpal AK, Bechtold DA, et al. (2014) The thermogenic
800 effect of leptin is dependent on a distinct population of prolactin-releasing peptide
801 neurons in the dorsomedial hypothalamus. Cell Metab 20: 639-649.
802 72. Elphick MR, Mirabeau O (2014) The Evolution and Variety of RFamide-Type
803 Neuropeptides: Insights from Deuterostomian Invertebrates. Front Endocrinol
804 (Lausanne) 5: 93.
805 73. Kriegsfeld LJ (2006) Driving reproduction: RFamide peptides behind the wheel. Horm
806 Behav 50: 655-666.
807 74. Wood S, Loudon A (2017) The pars tuberalis: The site of the circannual clock in
808 mammals? Gen Comp Endocrinol.
809 75. Lincoln GA, Clarke IJ, Hut RA, Hazlerigg DG (2006) Characterizing a mammalian
810 circannual pacemaker. Science 314: 1941-1944.
811 76. Tortonese DJ (2016) Intrapituitary mechanisms underlying the control of fertility: key
812 players in seasonal breeding. Domest Anim Endocrinol 56 Suppl: S191-203.
813 77. Concannon PW, Castracane VD, Rawson RE, Tennant BC (1999) Circannual changes in
814 free thyroxine, prolactin, testes, and relative food intake in woodchucks, Marmota
815 monax. Am J Physiol 277: R1401-1409.
816 78. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, et al. (2015) Proteomics.
817 Tissue-based map of the human proteome. Science 347: 1260419.
818 79. Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, et al. (2017) The BioGRID
819 interaction database: 2017 update. Nucleic Acids Res 45: D369-D379.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

80. McDonnell E, Crown SB, Fox DB, Kitir B, Ilkayeva OR, et al. (2016) Lipids Reprogram Metabolism to Become a Major Carbon Source for Histone Acetylation. Cell Rep 17: 1463-1472.
81. Hindle AG, Grabek KR, Epperson LE, Karimpour-Fard A, Martin SL (2014) The liver proteome in hibernating ground squirrels is dominated by metabolic changes associated with the long winter fast. Physiol Genomics.
82. Merriman DK, Lahvis G, Jooss M, Gesicki JA, Schill K (2012) Current practices in a captive breeding colony of 13-lined ground squirrels (Ictidomys tridecemlineatus). Lab Anim (NY) 41: 315-325.
83. Nelson CJ, Otis JP, Martin SL, Carey HV (2009) Analysis of the hibernation cycle using LC-MS-based metabolomics in ground squirrel liver. Physiol Genomics 37: 43-51.
84. Rose JC, Epperson LE, Carey HV, Martin SL (2011) Seasonal liver protein differences in a hibernator revealed by quantitative proteomics using whole animal isotopic labeling. Comp Biochem Physiol Part D Genomics Proteomics 6: 163-170.
85. R Core Team (2016) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
86. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.
87. Freed D, Aldana R, Weber J, Edwards J (2017) The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. bioRxiv.
88. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, et al. (2014) Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 46: 912-918.
89. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987-2993.
90. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.
91. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297-1303.
92. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, et al. (2012) Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front Genet 3: 35.
93. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. Bioinformatics 27: 2156-2158.
94. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, et al. (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26: 2867-2873.
95. Armstrong DM (1971) Notes on Variation in Spermophilus tridecemlineatus (Rodentia, Sciuridae) in Colorado andAdjacent States, and Description of a New Subspecies. Journal of Mammalogy, 52: 528-536.
96. Matthew P. Conomos, Timothy Thornton, Gogarten SM (2017) GENESIS: GENetic EStimation and Inference in Structured samples.
97. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.
98. Shaun Purcell, Chang C PLINK 1.9.
99. Hadfield J (2010) MCMC Methods for multi-response generalized linear mixed models: The MCMCglmm R Package. Journal of Statistical Software 33: 1-22.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

100. Hong LZ, Li J, Schmidt-Kuntzel A, Warren WC, Barsh GS (2011) Digital gene expression for non-model organisms. Genome Res 21: 1905-1915.
101. Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28: 1353-1358.
102. IUCN (International Union for Conservation of Nature) 2016. Ictidomys tridecemlineatus. The IUCN Red List of Threatened Species. 2017-2 http://www.iucnredlist.org. Downloaded on 12 June 2017

Supporting Information Captions

Fig S1. Comparison of the contiguity of the input assembly and the final HiRise scaffolds. Each curve shows the fraction of the total length of the assembly in scaffolds of a given length or smaller. The fraction of the assembly is indicated on the Y-axis and the scaffold length in basepairs is given on the X-axis. The two dashed lines mark the N50 and N90 lengths of each assembly. This plot excludes scaffolds less that 1 kb.

Table S1. Details of the draft assembly input and orientation into the final HiRise assembly

Fig S2. HiRise assembly improves the scaffold N50: 539 draft assembly scaffolds are reduced to 33. Each bar represents a HiRise scaffold. Each color within the bar represents a draft assembly scaffold. Note: colors are used to show placement of input scaffolds but are not specific to any one scaffold. Number of draft input scaffolds are listed on right.

Table S2. Summary mapping results for each ddRADseq library

Fig S3. Variant calling pipeline and results. (A) The Venn diagram (top) shows the number of unique and shared variants detected by Platypus, Sentieon and SamTools variant callers. The flow chart beneath outlines the filtering strategy and the number of retained variants after each filtering step. (B) For the 884,092 variants detected by and passing filter flags in each variant caller, plot shows the mean proportion of variant calls that were concordant (Concord), discordant (Discord) or Missing between Sentieon and Platypus (Plat) and Sentieon and SamTools (SamT) among all samples as a function of coverage. Blue box highlights criterion range of ≤ 65x coverage. (C) The observed versus expected heterozygosity ratio of each variant, represented as an open circle and plotted by minor allele frequency (MAF). Those with

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

a ratio >1.2, above the black horizontal line, were filtered from the dataset. Each variant is colored according to its Hardy-Weinberg equilibrium test statistic p-value bin listed in the legend (top left of plot). (D) For the 575,178 variants passing all filtering steps, the plot shows the mean proportion of variant calls that were concordant (Concord), discordant (Discord) or Missing between duplicate libraries of 4 samples (two samples from UW OshKosh, WI, "OK.TR1" and "OK.TR2"; and two samples from IL, "IL.TR1" and "IL.TR2") as a function of coverage.

Fig S4. ADMIXTURE 5-fold cross-validation (CV) error for each value of *K*. Shown are the CV error values for *K*=2 through *K*=8.

Table S3. Pairwise FST estimates for the *K*=6 ADMIXTURE populations.
See text for labeling.

Fig S5. Principal components PC17 and PC19 of all 153 genotyped squirrels reveal population structure within the OshKosh subset of squirrels. Coloring is the same as in Fig 2A and 2C.

Table S4: Biological and environmental data for each Oshkosh squirrel in which torpor onset was recorded

Table S5. Heart *trans*-eQTL results
Table lists results for all *trans*-eqtl associations with $p \leq 1X10^{-5}$
See Table 3 for SNP #
Tag ID is from the original Edge-tag datasets in [48,49]
β is effect size estimated from quantile-normalized values

Table S6. SkM *trans*-eQTL results
Labeling is the same as in Table S5

Table S7. Liver *trans*-eQTL results
Labeling is the same as in Table S5

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

934

935 **Table S8. BAT *trans*-eQTL results**

936 Labeling is the same as in Table S5

937

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

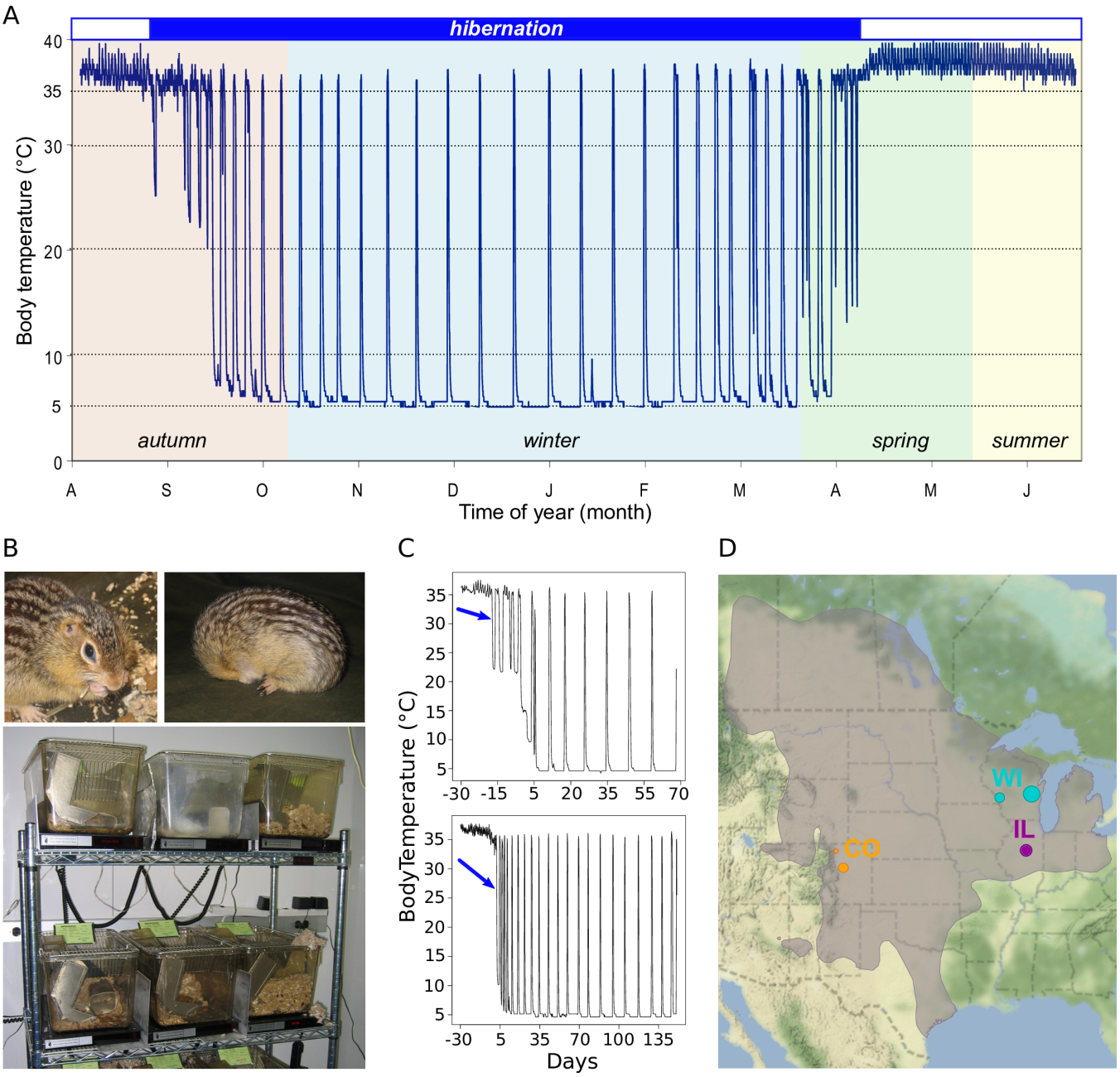

Fig 1. The 13-lined ground squirrel as a model for studying the genetics of hibernation. (A) Body temperature trace showing a 13-lined ground squirrel's year. Hibernating portion is demarcated by blue shaded box above. (B) A non-hibernating (top left) and hibernating (top right) 13-lined ground squirrel, individually housed in standard lab rodent cages in an animal facility (bottom). (C) Representative plots of body temperature telemetry analyses. Arrows point to the first day of torpor, the phenotype measured in this study. Days are ± from hibernaculum placement. (D) Approximate locales of squirrels genotyped in this study. Shaded area indicates the 13-lined ground squirrel's geographic range [102].

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

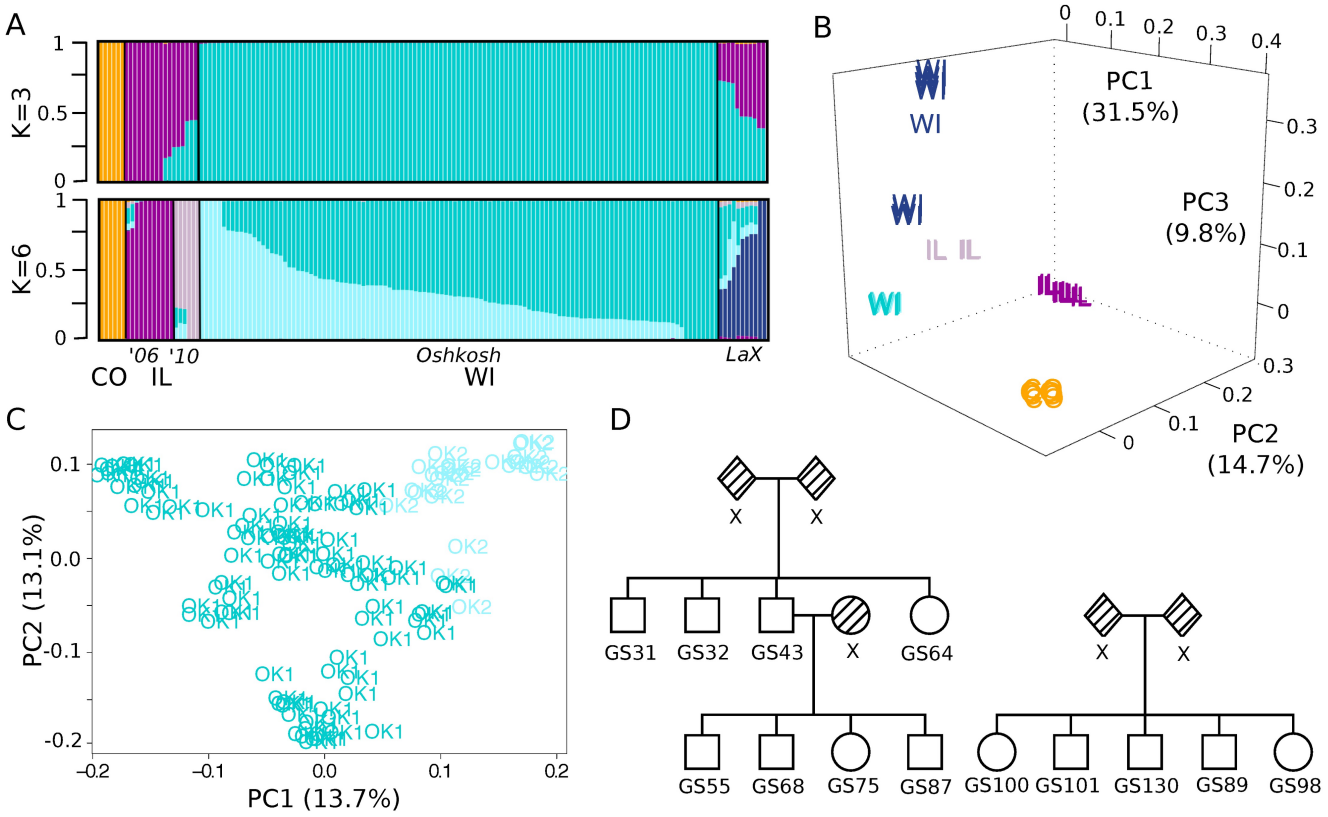

Fig 2. Genotype-by-sequencing reveals population structure and relatedness among sampled squirrels. (A) ADMIXTURE analysis results showing *K=3* or *K=6* genome-wide specific ancestry estimates. Squirrels are shown as vertical bars with proportion of specific ancestry colored within each bar. Populations are clustered and labeled by geographic sampling locales (U.S. state, and for WI, city) and for those from IL, sampling years. (B) Principal components analysis of all 153 genotyped squirrels. The first 3 PCs are plotted, with individuals labeled by state and colored by the population for which they have the greatest proportion of ancestry as determined by *K=6* ADMIXTURE analysis shown in (A). (C) Principal components analysis of 119 squirrels from the Oshkosh WI population. The first 2 PCs are plotted with individuals labeled and colored by Oshkosh sub-population (*OK1* or *OK2*) as determined by *K=6* ADMIXTURE analysis shown in (A). (D) Representative pedigrees reconstructed from identity by descent (IBD) and kinship coefficient estimates of the Oshkosh squirrels. Shaded shapes labeled “x” indicate relatives not genotyped in this study.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

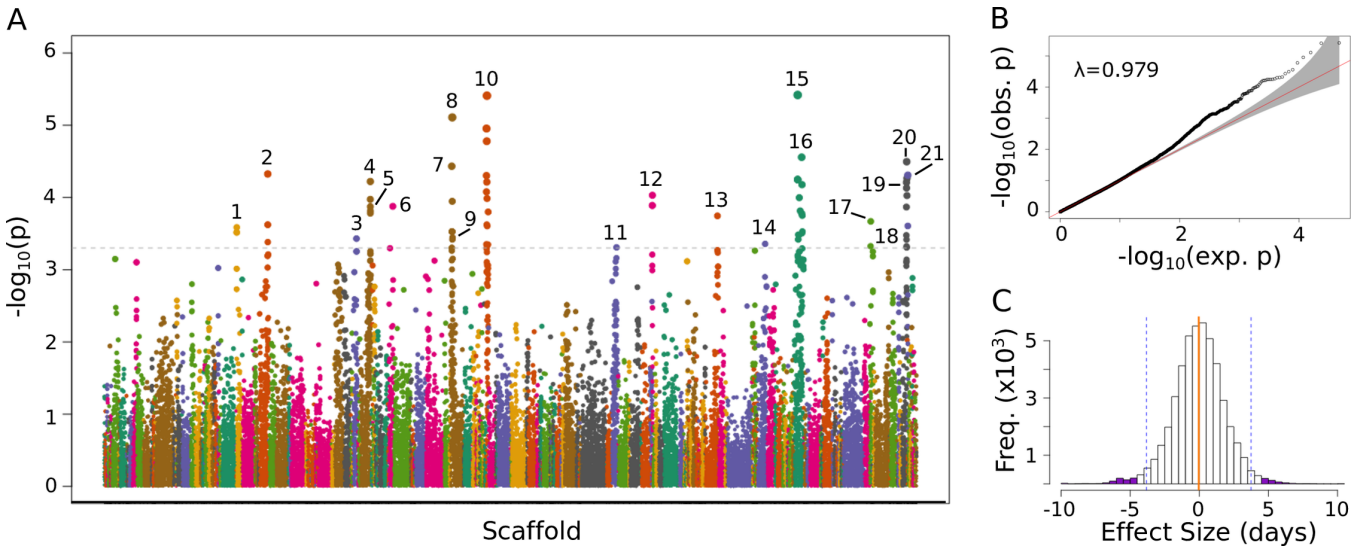

Fig 3. GWAS identifies genetic variants significantly associated with date of first torpor in 13-lined ground squirrels. (A) Manhattan plot shows the negative log-transformed p-values of 46,996 variants (MAF>0.05) tested for association with date of first torpor in 72 squirrels. Variants are ordered by position on scaffold, which are colored along x-axis. Dashed line indicates cutoff for significance ($p<5x10^{-4}$). Significantly associated variants, pruned for LD, are numbered and correspond to those detailed in Table 3 and in Figs 5 and 6. (B) Q-Q plot of the GWAS log-transformed p-values. (C) Histogram of effect sizes of the 46,996 variants on date of first torpor. Orange vertical line marks the mean, dashed vertical lines mark upper and lower bounds of 98th percentile and purple shading indicates effect sizes of the 21 significantly associated variants.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

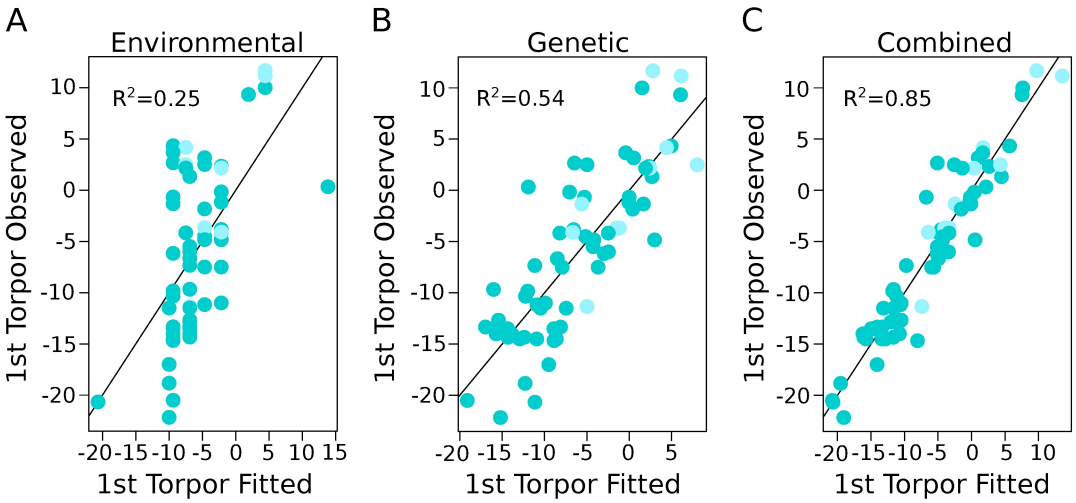

Fig 4. Combined environmental and genetic effects account for variation in the start of torpor. Plots show correlation between fitted and observed values for the start of torpor (in days from hibernaculum placement) using a linear regression model. Adjusted R^2-value is labeled in each. Shading matches Figs 2A and 2C. (A) Linear model fit with the environmental variables of year of monitoring and date of hibernaculum placement, and the biological variable of sex (i.e. the “fixed effects”, see methods). (B) Linear model fit with the 21 significant SNP genotype combinations for each squirrel. (C) Linear model fit with the variables from both (A) and (B).

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

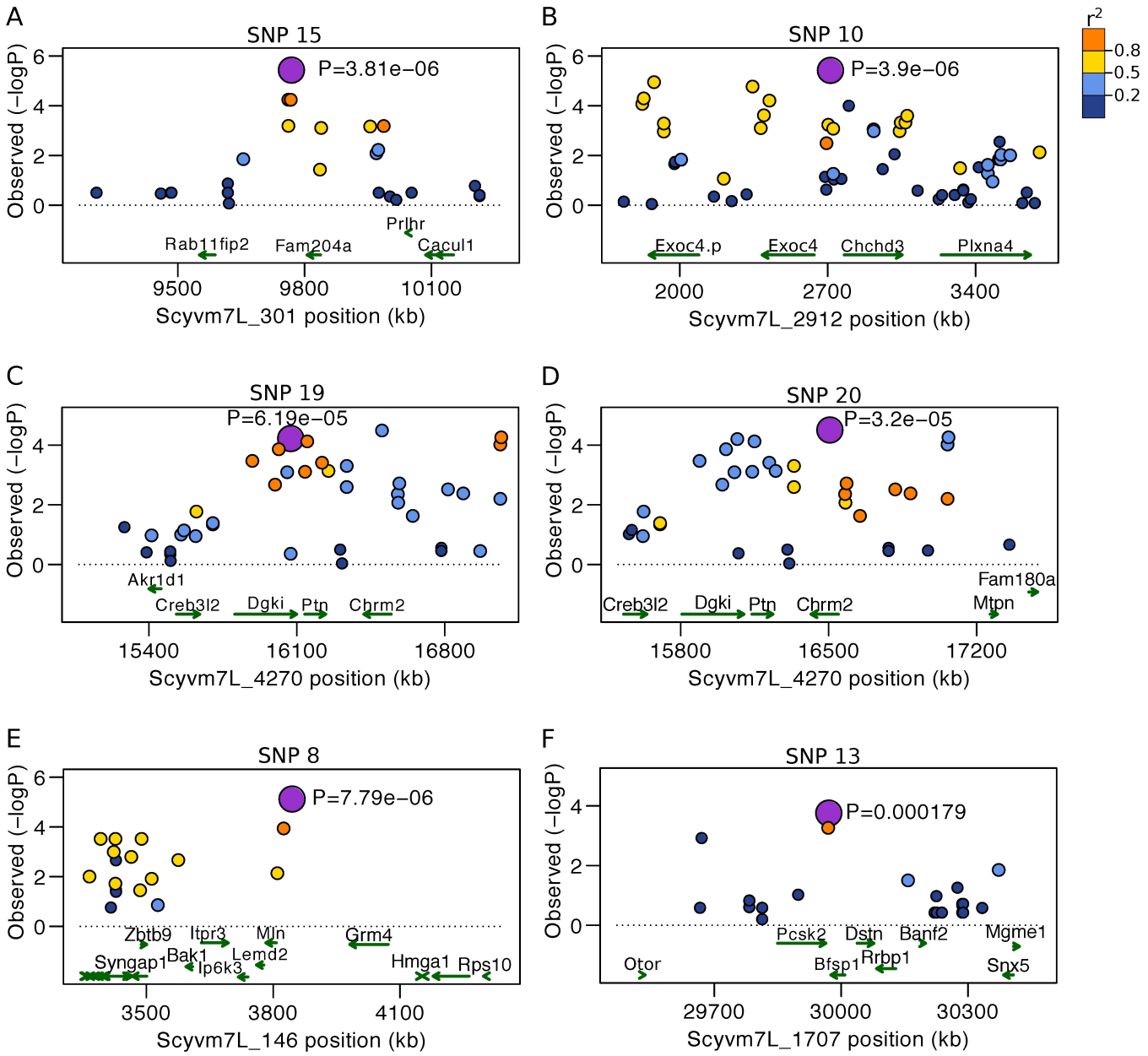

Fig 5. Regional Manhattan plots show locations of selected significant SNPs in proximity to nearest genes. Each plot is centered on one of the significant SNPs, which are labeled on top by number (see Fig 3A and Table 3) and shaded purple. Other variants within region are colored by LD value (r^2) in relation to the significant SNP. Genes are shown below as green arrows and labeled by gene symbol. (A) SNP 15, the most significant SNP in the GWAS, is nearest *FAM204A*. (B) SNP 10, the second most significant SNP, is between *CHCHD3* and *EXOC4*. (C) SNP 19 is located nearest *DGKI* and *PTN*. (D) SNP 20 is located nearest *CHRM2*. (E) SNP 9 is nearest *MLN*. (F) SNP 13 is between *PCSK2* and *BFSP1*.

bioRxiv preprint doi: https://doi.org/10.1101/222307; this version posted December 2, 2017. 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-NC-ND 4.0 International license.

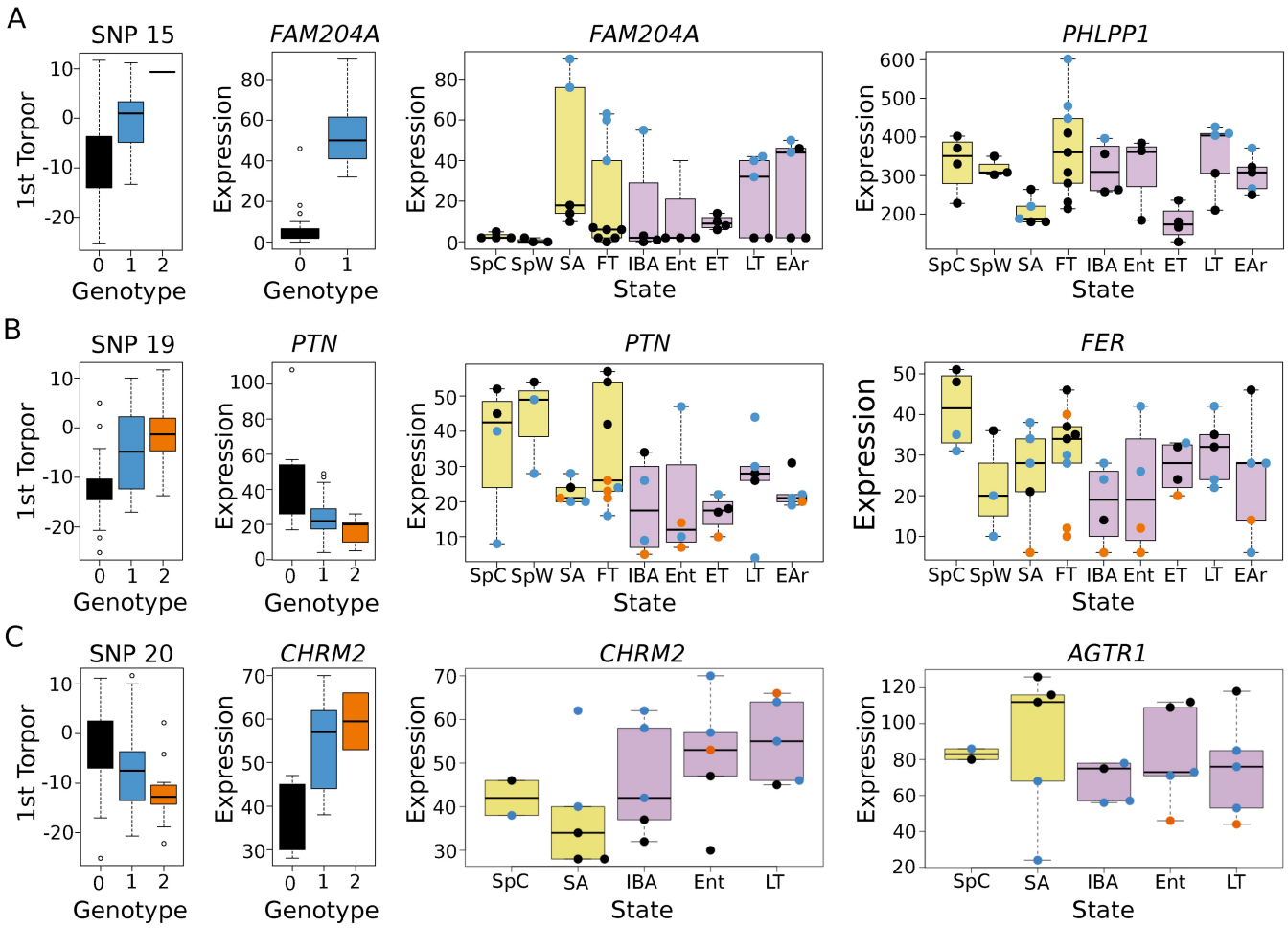

Fig 6. Significant GWAS SNPs are also *cis*- and *trans*-eQTLs that explain variation in mRNA expression. (A) Effect of SNP 15 genotype on date of first torpor (far left) and expression of its *cis*-eGene, *FAM204A*, in BAT (middle left). Middle right plot shows effect of genotype on transcript expression within each of the nine distinct physiological and seasonal states interrogated in the original transcriptome study (labeled below, see methods for explanation of sampling abbreviations). Shaded yellow boxes indicate physiological states from homeothermic and transitional portions of hibernator's year (spring-autumn), while purple shaded boxes are those from within deep hibernation. Far right plot shows effect of genotype on expression of the most significant *trans*-eGene, *PHLPP1*. (B-C) Labeling is as in panel (A). (B) Effect of SNP 19 on *cis*-eGene *PTN* and *trans*-eGene *FER* in BAT. (C) Effect of SNP 20 on *cis*-eGene *CHRM2* and *trans*-eGene *AGTR1* in heart.