

Giant Island Mice Exhibit Widespread Gene Expression Changes in Key Metabolic Organs

Mark J Nolte ^{1,*}, Peicheng Jing¹, Colin N Dewey ², and Bret A Payseur¹

¹Laboratory of Genetics, University of Wisconsin – Madison

²Department of Biostatistics and Medical Informatics, University of Wisconsin – Madison

*Corresponding author: E-mail: mj nolte@wisc.edu.

Accepted: 6 June 2020

Data deposition: RNA-sequencing data from this project has been deposited at GEO under the accession GSE135007. Whole-genome sequencing data for Gough Island mice has been deposited at SRA under the accession PRJNA587779. [Supplementary tables 3, 7, and 11](#) are available on Dryad: <https://doi.org/10.5061/dryad.cjxksn34> (doi:10.5061/dryad.cjxksn34).

Abstract

Island populations repeatedly evolve extreme body sizes, but the genomic basis of this pattern remains largely unknown. To understand how organisms on islands evolve gigantism, we compared genome-wide patterns of gene expression in Gough Island mice, the largest wild house mice in the world, and mainland mice from the WSB/EiJ wild-derived inbred strain. We used RNA-seq to quantify differential gene expression in three key metabolic organs: gonadal adipose depot, hypothalamus, and liver. Between 4,000 and 8,800 genes were significantly differentially expressed across the evaluated organs, representing between 20% and 50% of detected transcripts, with 20% or more of differentially expressed transcripts in each organ exhibiting expression fold changes of at least 2 \times . A minimum of 73 candidate genes for extreme size evolution, including *Irs1* and *Lrp1*, were identified by considering differential expression jointly with other data sets: 1) genomic positions of published quantitative trait loci for body weight and growth rate, 2) whole-genome sequencing of 16 wild-caught Gough Island mice that revealed fixed single-nucleotide differences between the strains, and 3) publicly available tissue-specific regulatory elements. Additionally, patterns of differential expression across three time points in the liver revealed that *Arid5b* potentially regulates hundreds of genes. Functional enrichment analyses pointed to cell cycling, mitochondrial function, signaling pathways, inflammatory response, and nutrient metabolism as potential causes of weight accumulation in Gough Island mice. Collectively, our results indicate that extensive gene regulatory evolution in metabolic organs accompanied the rapid evolution of gigantism during the short time house mice have inhabited Gough Island.

Key words: island evolution, island rule, body size, house mouse, gene regulatory evolution.

Significance

Reported instances of island-colonizing organisms evolving extreme body sizes are numerous, whereas insights into the genetic changes responsible for this widespread phenomenon are fewer. We found genes that are expressed at different levels in metabolic organs from giant Gough Island mice and smaller mainland mice, including genes with roles in cell cycling, fat cell maturation, the inflammatory response, and nutrient metabolism. The genes and cellular processes we identified help to elucidate the gene regulatory basis of gigantism evolution.

Introduction

The repeated evolution of similar phenotypes is a hallmark of adaptation. Island populations of vertebrates often evolve

extreme body sizes, either gigantism or dwarfism (Foster 1964), a pattern known as the “island rule” (Foster 1964; Van Valen 1973). The genomic basis of body size evolution on islands remains poorly characterized, even while the

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

ecological causes and generality of the island rule have been subject to considerable study and debate (Lomolino 2005; Meiri et al. 2008; Raia and Meiri 2011).

One striking example of extreme body size evolution comes from house mice that colonized remote Gough Island, which is located in the South Atlantic Ocean, roughly 2,100 miles from both South America and South Africa. The mice on Gough Island are the largest wild house mice on record. Adult island mice are, on average, about 45% heavier than mainland counterparts (island/mainland males: 26.5 g/18 g; island/mainland females: 22 g/15.5 g) and their skeleton is 12% longer (Rowe-Rowe and Crafford 1992; Jones et al. 2003; Gray et al. 2015; Parmenter et al. 2016). Moreover, house mice likely colonized the island within the last 200 years (Wace 1961; Gray et al. 2014), suggesting that size evolution was rapid. Gough Island mice belong to the same subspecies as laboratory mice (*Mus musculus domesticus*) (Gray et al. 2014), opening the door for the application of genomic and molecular genetic tools developed for this model system to the question of the genetic basis of an instance of the island rule.

Utilization of classical inbred strains selected for growth differences revealed that mechanisms of weight accumulation are complex, involving different cellular processes throughout ontogeny and underpinned by tens of quantitative trait loci (QTL) (Vaughn et al. 1999; Atchley et al. 2000; Cheverud et al. 2001). Gough Island mice are unique because they provide a rare opportunity to explore the genetics of this complex trait from a population evolving in the wild. Gray et al. (2015) reported that over 16 weeks of measurement, eight QTLs contribute to body weight differences between Gough Island mice and a mainland strain from the same subspecies (WSB/Eij), whereas 11 QTLs contribute to differences in growth rate. The QTL display additive effects of modest magnitude, with the largest for weight being 0.66 g. Most of the QTLs begin to affect growth in the first 6 weeks after birth, a period that coincides with augmented growth rates for laboratory mice (Cheverud 2005). These findings provide clues about the genetic architecture of size evolution and point to candidate regions for further study, as the underlying genetic and molecular mechanisms for differential growth are unknown.

Differential expression analyses can identify candidate genes within QTL that are responsible for trait divergence. In mice and humans, regulatory mutations are known to contribute to standing variation in body size and length (Oliver et al. 2005; Clausnitzer et al. 2015; Castro et al. 2019). Additionally, many of the variants connected to size variation in humans through genome-wide association studies fall outside protein-coding regions (Speliotes et al. 2010; Locke et al. 2015; Shungin et al. 2015), indicating an important role for gene regulation in the determination of this complex trait. We hypothesized that evolution of gene expression in three metabolic organs contributed to the evolution of gigantism in Gough Island mice: the gonadal adipose depot, hypothalamus, and liver. Nutrient uptake and distribution is modulated

by extensive cross-talk between these organs, and developmental, physiological, and immunological processes in both the healthy and diseased state are well documented for these organs (gonadal adipose [Bjørndal et al. 2011; Han et al. 2011; Oh et al. 2015; Lackey and Olefsky 2016], hypothalamus [Lam et al. 2005; Ohlsson et al. 2009; Thaler et al. 2012; Kälén et al. 2015], and liver [Hay 1991; Kimura 1991; Liechty and Lemons 1991; Cuezva et al. 1997; Hart et al. 2009; Ohlsson et al. 2009; Klaassen and Aleksunes 2010; Klochendler et al. 2012; Septer et al. 2012; Gunewardena et al. 2015; Lackey and Olefsky 2016]).

In this article, we report genome-wide patterns of differential expression between Gough Island mice and a mainland strain, with the goal of understanding the evolution of island gigantism. Although expression differences may relate to phenotypic differences other than body size, we discovered over 4,000 differentially expressed (DE) genes each in the gonadal adipose, hypothalamus, and liver. We found strong candidate genes for size evolution by overlapping DE genes with body size QTL and with single-nucleotide differences located in putative regulatory elements. By comparing the differential expression of liver-specific genes across three time points, we identified potential coregulated gene groups, which allowed us to identify genes encoding transcriptional regulators within body size QTL that might coordinate widespread expression differences between Gough Island and mainland mice. Functional enrichment analyses of DE genes pointed to candidate cellular and physiological processes underlying the evolution of extreme size, including differences in cell proliferation and organ maturation. This work builds upon previous studies aimed at illuminating the genetic basis of the evolution of island gigantism and reaffirms that Gough Island mice are a unique resource for investigating the evolution of extreme size in natural settings.

Materials and Methods

Sample Collection

Two mouse strains provided tissues for this study: mice from Gough Island (referred to as “island mice” in this report) and WSB/Eij (The Jackson Laboratory, Bar Harbor, ME; referred to as “mainland mice”). The laboratory colony of Gough Island (“island”) mice was founded from mice live-caught on the island and shipped to our laboratory in 2009 (Gray et al. 2015). RNA was obtained from individuals inbred (brother–sister matings) to the seventh generation. All island females sampled shared a most recent common ancestor in the fourth generation of inbreeding. For island females, three litters contributed to the E16.5 samples, two to the 2-week and four to the 4-week. All island males sampled shared the same most recent common ancestor as the island females. For island males, two litters contributed to the E16.5 samples, three to the 2-week and one to the 4-week. WSB/Eij is a wild-derived,

inbred strain with a body size representative of mainland wild house mice. Among mainland mice sampled, a minimum of two litters and a maximum of four provided individuals for sampling across the five conditions. Each mouse was treated as an independent replicate of its line for analyses of differential expression. Island mice are larger than mainland mice. A general linear model that considered all mice in litters that contributed individuals for RNA collection, with predictor variables litter size, age (2 or 4 weeks), sex, strain and the interaction of strain and age showed that, after controlling for the effects of all predictor variables, strain was significantly related to body weight ($F(1, 102) = 102.9, P < 0.001$, partial $\eta^2 = 0.50$; adjusted mean weights (M) and standard errors (SEs): 2-week: $M_{\text{island}} = 10.6$ g, $SE_{\text{island}} = 0.26$ g; $M_{\text{mainland}} = 7.2$ g, $SE_{\text{mainland}} = 0.28$ g; 4-week: $M_{\text{island}} = 15.9$ g, $SE_{\text{island}} = 0.20$ g; $M_{\text{mainland}} = 10.4$ g, $SE_{\text{mainland}} = 0.20$ g).

RNA was obtained from five conditions: the gonadal adipose depot (“adipose”; 4 weeks); the hypothalamus (4 weeks); and the liver at three time points—embryonic day (E) 16.5, postnatal day 14 (2 weeks), and postnatal day 28 (4 weeks). For each condition, samples from five males and five females of each strain were collected, for a total of 100 samples. All samples were snap frozen in liquid nitrogen and stored at -80°C until RNA extraction.

To collect E16.5 embryos, pregnant females were euthanized via CO_2 asphyxiation between 11:30 AM and 1:30 PM. As mouse embryos are resistant to hypoxia, embryos were dissected out of uteri into fresh phosphate-buffered saline whereupon embryos were decapitated immediately, ensuring rapid retrieval of the liver for RNA preservation. To determine the sex of embryos, we followed the polymerase chain reaction (PCR) protocol in Deng et al. (2011), genotyping for the presence or absence of *Sry*, using *Rapsn* amplicons as a positive control.

2-week- and 4-week old mice were placed in a 1 liter chamber and exposed to nonvaporized isoflurane until nonresponsive. Mice were then euthanized via decapitation. This euthanasia method is rapid (≤ 45 s) and obviates the need for excessive handling, resulting in rapid acquisition of organs for RNA preservation. From the medial border of the liver’s left lobe an ~ 0.5 cm \times 0.5 cm section was taken from 2-week- and 4-week-old individuals. Two-week collections occurred between 11 AM and noon, except for two mainland individuals collected between 9:30 and 10 AM. Collection of 4-week adipose, whole hypothalamus and liver samples occurred between 9 AM and 11 AM.

We collected both right and left gonadal (peri-uterine) fat depot samples from 4-week females. Samples were ~ 0.25 cm \times 0.25 cm and taken from the border between the parametrial and perivesical portions of the abdominopelvic (also called the peri-uterine or gonadal) fat depot (Vitali et al. 2012) (supplementary fig. 1, Supplementary Material online). In 4-week males, an ~ 0.5 cm \times 0.5 cm portion of both testicular fat depots was collected; care was taken not to collect any epididymal or testicular tissue. At the time of RNA extraction, right and left fat samples were combined.

Animal Husbandry

Pups were weaned at 3 weeks; individuals providing tissue for the 2-week time point were housed with their mother and siblings until the time of tissue collection. Weanlings were housed with two to three other individuals of the same sex. All adult and weaned mice were housed in a temperature controlled room (68 – 72°F) set on a 12-hour light/dark cycle. Pregnant and nursing mothers were provided with breeder chow (Envigo 2019 Teklad Global Diet) and water ad libitum, whereas weaned mice were provided with standard rodent chow (Envigo 2020X Teklad Global Diet) and water ad libitum.

RNA-Sequencing Library Preparation

Frozen liver tissue was homogenized in 1 ml of TRIzol Reagent (Ambion by Life Technologies) using a powered Qiagen TissueLyser II homogenizer. Liver RNA was isolated using the protocol accompanying the TRIzol Reagent; after the “RNA resuspension” step, the total RNA was treated with DNase and purified (DNA-free Kit, Ambion by Life Technologies, AM1906). Frozen adipose and hypothalamus samples were processed using the Qiagen RNeasy Lipid Tissue Mini Kit (74804). Additionally, The RNase-Free DNase Set (Qiagen, 79254) was used to treat each adipose and hypothalamus sample. Highest quality RNA samples were selected. Selected RNA samples were verified to have an A260/A280 ratio between 1.8 and 2.1 on a NanoDrop UV–Vis Spectrophotometer (Thermo Scientific). Additionally, sample integrity was verified using a fragment analyzer (RNA quality number: $M = 8.5$, $SD = 0.85$). Subsequently, 1 μg of DNase-treated and purified total RNA from each sample was used for mRNA enrichment and RNA-seq library preparation.

Paired-end, 50-base pair RNA-seq libraries were generated on an Illumina HiSeq 2500 platform (using the Rapid Run mode) in the University of Wisconsin-Madison Biotechnology Center. RNA-seq library preparation and construction included the use of Illumina TruSeq RNA Sample Prep kits (v2, sets A and B; RS-122-2001 and RS-122-2202), which enabled the generation of mRNA-enriched libraries from the total RNA, and Illumina TruSeq Rapid Cluster-HS kits (PE-402-4001), which enabled the formation of clonal template clusters during the HiSeq 2500 runs, and Illumina TruSeq Rapid SBS kits (FC-402-4002) for the sequencing of reads.

Read Mapping

We mapped the RNA-seq reads to a reference set of transcripts from the *Mus musculus domesticus* reference genome (mm10/GRCm38, Ensembl v79 annotation) and estimated transcript abundances and read counts using the software packages RSEM (v1.2.21) (Li and Dewey 2011) and Bowtie (v1.1.1) (Langmead et al. 2009). As the libraries were mRNA

enriched, only long transcripts were kept in the reference set. Reference transcript sequences were modified in one of two ways. When mapping reads from island mice, single-nucleotide polymorphisms (SNPs) fixed in island mouse-coding sequences were used to replace sites in the reference transcript sequences. The fixed island mouse SNPs were identified as those that differed between the island and mainland inbred strains (but were invariant between a male and a female founder of the island mouse strain) from whole-genome sequences (average depth of coverage = 10×) (described below). When mapping reads from mainland mice, fixed SNPs within mainland mouse-coding sequences (obtained from Keane et al. 2011) were used to replace sites in the reference transcript sequences.

Detection of Differentially Expressed Genes

To determine whether a gene was differentially expressed (DE) between island and mainland mice, we used a negative binomial generalized linear model (GLM) in the DESeq2 package (Love et al. 2014), implemented in R (Bioconductor version 3.2). To identify DE genes in adipose and hypothalamus, we modeled read counts using factors strain, sex, and strain by sex interaction (strain:sex). Because we collected transcripts from three time points for the liver, we incorporated age as a factor in the model for determining read counts: strain, sex, age, sex:age, strain:sex, strain:age, and strain:sex:age. DE genes were identified as those with an adjusted *P* value < 0.05 after using the Benjamini–Hochberg procedure for multiple test correction with a false discovery rate = 0.05 (Benjamini and Hochberg 1995).

Our analyses focused on sex-averaged differences between the strains for two reasons. This approach enabled straightforward comparison of the locations of DE genes and QTL (Gray et al. 2015) that confer body size differences between these strains in both sexes. Moreover, whole-body sexual dimorphism does not begin until around 3 weeks of age in mice, with accompanying sex-specific gene expression extending across puberty and into increasingly mature mice (Waxman and Celenza 2003; Cheverud 2005). We did detect minimal to modest sex-specific differential expression from the factor sex in our models: 4-week adipose, 1,656 genes; 4-week hypothalamus, 15 genes; E16.5 liver, 34 genes; 2-week liver, 9 genes; and 4-week liver, 138 genes. The relatively large number of sex-specific differential expression in the adipose may reflect the fact that female and male gonadal adipose depots are less anatomically, and perhaps functionally, homologous than other adipose depots, whereas the increased sex-specific differences in more mature livers may reflect the sexual dimorphism that accompanies liver gene expression in increasingly mature mice (Ohlsson et al. 2009). To obtain sex-averaged gene expression differences, contrast vectors were multiplied by the vector of GLM coefficients to obtain sex-averaged \log_2 fold changes for all assayed

conditions. Throughout our report, higher and lower refer to gene expression levels that were higher or lower in island mice relative to mainland mice.

To construct putative coregulated gene groups (described below) from the expression data collected across three time points in the liver, we multiplied a series of contrast vectors with the vector of GLM coefficients to obtain \log_2 fold changes from one time point to another within a strain (as opposed to across strains): embryonic-to-2-week and 2-week-to-4-week.

Functional Enrichment Analyses

Utilizing the R package Goseq (version 1.24.0), we determined the enrichment of Gene Ontology (GO) annotations and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways among the DE genes in all conditions assayed (Young et al. 2010). In Goseq, enrichment analyses were conducted with the *method* argument in the *goseq* function set to “Hypergeometric.” For each condition, separate enrichment analyses were conducted for genes expressed at higher and lower levels in island mice, and the background set of genes consisted of all genes whose transcription was detected in that condition. To obtain maximum functional information, we determined the quantity of significantly enriched GO and KEGG terms returned for different magnitudes of fold change. We quantified the number of enriched GO and KEGG terms for 9-fold change cutoff classes ($|\log_2$ fold change| > 0, or ≥ 0.5 , 1, 1.5, ..., 4). We report on the terms from the fold change cutoff classes from each condition that provided the greatest and second-greatest quantity of significantly enriched GO and KEGG terms (in all cases, these cutoff classes were $|\log_2$ fold change| > 0 or ≥ 0.5) (supplementary figs. 2–5, Supplementary Material online). This approach incorporated terms associated with stringent cutoff classes and ensured that biological processes represented by relatively small expression differences were considered.

The enrichment analyses for all liver conditions, except higher expressed genes at 2 weeks, returned 115 or more significantly enriched GO terms (maximum returned GO terms = 464, for higher in island 4-week liver). Although the full output of the enrichment analyses is available in supplementary tables 1, 2, and 4, Supplementary Material online; supplementary table 3 available at <https://doi.org/10.5061/dryad.cjxksn34>, to navigate these results we used two arrangements of the significant GO terms to look for patterns in the cellular and biochemical processes that could relate to differences between island and mainland mice. First, we used REVIGO to obtain the specificity of each GO term (Supek et al. 2011). In REVIGO, output files contain a “frequency” column that indicates the percentage of genes in the GO Annotation database annotated to a GO term; we consider the REVIGO frequency score to be a measure of the GO term’s specificity. More specific or detailed GO terms are annotated to fewer

genes, whereas increasingly general GO terms are annotated to a greater number of genes. For the purpose of illuminating differences in specific cellular, biochemical, and biological processes between the strains, we focused on the top 50% of the most specific (per the REVIGO frequency value) enriched GO terms in each condition and fold-change direction. (A visual example of how the first arrangement was implemented can be seen in [supplementary fig. 6, Supplementary Material](#) online.)

In a second arrangement, we extracted the broad functional categories, or GOSlim categories, to which each of the specific GO terms is mapped from the GOSlim resource at Mouse Genome Informatics (The Jackson Laboratory, http://www.informatics.jax.org/gotools/MGI_GO_Slim.html, last accessed June 26, 2020). There are 38 GOSlim categories: 12 each are affiliated with subontologies Cellular Component and Molecular Function, and 14 are affiliated with subontology Biological Function. We then quantified the number of highly specific significantly enriched GO terms (from the first arrangement described above) in each of the GOSlim categories. This allowed us to make observations about the broad patterns differentiating the metabolic organs evaluated in the two mouse strains. In particular, because we assayed three time points in the liver, we were especially interested in noting the dynamism in cellular processes across time. For this purpose, in the liver, we paid special attention to GOSlim categories to which a large proportion of specific GO terms mapped from all three time points ([supplementary table 5, Supplementary Material](#) online). Importantly, merging these two arrangements enabled us to assess the most specific biological processes (provided by the first arrangement) that differ between island and mainland mice, while retaining access to observations on broad differences in liver maturation and function (provided by the second arrangement).

Identifying Coregulated Gene Groups in the Liver

We reasoned that genes whose expression levels changed from one time point to another in a similar manner in the liver might be regulated by the same transcription factor(s) (TFs). Such TFs, we further reasoned, might be key players in the differential expression we discovered because their activity could impact the expression of the downstream genes they regulate. We used two criteria related to differential expression across the three time points in the liver to place a subset of genes into coregulated groups. First, we looked for genes that showed coordinated expression changes throughout liver maturation *within* strains. Second, we looked for genes that showed similar patterns of differential expression *between* strains (throughout liver maturation). Below, we describe each criterion in turn.

Within each mouse strain, genes expressed across all three time points could manifest significant changes in expression level within two time intervals: from E16.5 to 2 weeks, and

from 2 weeks to 4 weeks. Thus, a gene could manifest one of nine possible within-strain temporal trajectories across time intervals one and two: decrease–decrease, decrease–increase, decrease–no change, increase–decrease, increase–increase, increase–no change, no change–decrease, no change–increase, and no change–no change. Placement of genes into the trajectories depended on two filters. First, the change in magnitude had to be significantly different (adjusted P value <0.05) in a contrast from the full DESeq2 GLM that explicitly tested for changes in expression level from one time point to another within a strain. Second, in addition to statistical significance, the change in expression magnitude had to exceed an absolute \log_2 fold change of 0.6 from one time point to another. Note that any given gene may belong to either the same temporal trajectory class in both island and mainland mice, or to a different one. Thus, genes that are expressed throughout liver maturation in both strains could be mapped onto a 9×9 grid, which indicates the within-strain temporal trajectory to which the gene belongs in both strains ([supplementary fig. 7, Supplementary Material](#) online).

In addition to classifying liver-specific expressed genes by their strain-specific temporal trajectories (criterion 1), we also classified them according to their between-strain differential expression patterns (criterion 2). A gene that is expressed in both strains across all three time points may be DE at one, two, or all three time points. Thus, each (presumably) continuously expressed gene could be sorted into one of 27 between-strain differential expression patterns: either higher, lower, or the same at any one of the three time points ($3^3 = 27$). As was done for criterion 1, we considered a gene to be DE if the expression level difference was both statistically significant (adjusted P value <0.05), per the DESeq2 GLM, and if the absolute \log_2 fold change exceeded 0.6.

In summary, a gene could be classified 1) by its within-strain temporal trajectories and 2) by its between-strain differential expression pattern. By combining these two classes of patterns, we could separate continuously expressed genes into putative coregulated groups. To do this, we further divided the genes mapped onto the 9×9 within-strain temporal trajectory grid (from grouping criterion 1) by their classification into one of the 27 differential expression patterns (grouping criterion 2). This procedure resulted in 27 individual 9×9 grids, where each grid is specific to a particular differential expression pattern. We hypothesized that genes located within the same cell, on any of the 27 grids, constitute a putative coregulated group ([supplementary fig. 7, Supplementary Material](#) online).

We sought to identify candidate TFs regulating the putative coregulated groups that contained five or more genes meeting the two criteria described above, which resulted in 4,965 genes spread across 285 coregulated groups. We searched for enrichment of particular transcription factor binding sites (TFBSs), or motifs, in the 5 kb, 5'-upstream sequence of each of the genes within a putative coregulated group using the R

package PWMEnrich (Stojnic and Diez 2018). In each enrichment test, the background sequences used for comparison to the coregulated group's sequences were those 5 kb, 5'-upstream sequences from all genes expressed in both strains in any of the three time points assayed for the liver. Assessed TF binding motifs were obtained from the R package PWMEnrich.Mmusculus.background, which provided the position weight frequencies for 637 motifs. Prior to the enrichment analyses, background sequences with strings of "Ns" were removed (single "N" sites were kept, and assigned a weight of 0), nucleotide base frequencies were determined in the set of background sequences, the 637 position weight frequencies were converted to position weight matrices (PWMs), and the distribution of the PWMs was determined in the background sequences.

Enriched TFBSs upstream of putative coregulated groups were identified as those with an adjusted P value <0.05 after using the Benjamini–Hochberg procedure for multiple test correction with a false discovery rate = 0.05 (Benjamini and Hochberg 1995). The TF gene names corresponding to the significantly enriched binding sites were obtained by matching the motif IDs provided by PWMEnrich with the geneSymbol variable drawn from the R package MotifDb, which was queried with argument "mmusculus."

Identifying Candidate Genes for Gigantism by Intersecting Data Sets

To nominate candidate genes involved in the evolution of body size, we intersected our data set of DE genes with other publicly available data sets related to body size evolution in island mice and gene regulation in metabolic organs. For ease of discussion and interpretation, the sequence space around each DE gene of interest was divided into six sequence lengths upstream and downstream of the gene—500 base pairs, 1.5 kb, 3 kb, 5 kb, 10 kb, and 100 kb—as well as 5'- and 3'-untranslated regions, first and second introns, additional introns, and exons.

We employed the R package GenomicFeatures to create a TxDb object that contained sequence information from the mm10/GRCm38 mouse genome assembly (Lawrence et al. 2013). From this TxDb object, which maps virtual mRNA transcripts to their virtual genomic positions, we acquired the sequence content of each of the genic regions for each transcript variant of the DE genes of interest. We used the GenomicRanges function *reduce* to merge overlapping genic region sequences from different transcript variants, or gene isoforms (Lawrence et al. 2013). To ensure that the sequence coordinates of all of the transcripts for a given gene were considered only once per genic region, we merged genic regions (upstream sequences, untranslated regions, introns, etc.) into nonredundant sets of sequences. Note that if a particular sequence was designated a different genic region in two different transcripts then it was retained in the merged

sets of sequences for the respective genic regions (e.g., if a position was in the first intron of several variants but in the 5'-upstream region of another, then that position would be considered once in both the merged intronic and merged upstream sequences).

Our five-condition DE gene data set was intersected with four additional data sets to identify strong candidates for extreme body size evolution: 1) a published set of QTLs mapped using the phenotypes body weight and growth rate in a cross between the island and mainland strains (Gray et al. 2015); 2) fixed SNPs differentiating the island and mainland strains; 3) Mouse Encyclopedia of DNA Elements (ENCODE) data sets (Yue et al. 2014) and published data sets that catalog various, though not exhaustive, indicators of putative gene regulatory elements; and 4) segments of highly conserved sequences taken from the Phylogenetic Analysis with Space/Time models conservation track (phastCons) on the UCSC Genome Browser (Siepel et al. 2005). By intersecting these data sets, we crafted a way to qualitatively assess and rank candidate DE genes. For example, a DE gene within a body weight QTL with a proximal fixed SNP in an established regulatory element may be more causally connected to organ functions that promote gigantism than DE genes outside the QTL. Information on each of these data sets follows.

1. Gray et al. (2015) employed a large cross between island mice and the mainland strain used here (WSB/EiJ) to map 8 QTL for body weight and 11 QTL for growth rate (a total of 14 distinct QTL) during the first 16 weeks of life. We report on those DE genes for which any of their exonic positions lies within 2 Mb of the QTL peaks ("QTL-DE" genes; [supplementary table 6, Supplementary Material](#) online). We performed a permutation test (10,000 replicates) by permuting genic positions to test whether there are more DE genes in QTL than expected by chance. The number of DE genes in the QTL ($n = 363$) does not differ from chance (median = 363, mean = 362.5, min = 325, and max = 397) ([supplementary fig. 8, Supplementary Material](#) online). This result is expected under the simple assumption that a single mutation is responsible for any given QTL and that the mutation affects gene expression.
2. To obtain island mouse genome sequences for calling SNPs, Qiagen DNeasy Blood & Tissue kits were used to extract DNA from the livers of 16 wild-caught Gough Island mice (5 females/11 males) and four descendants of Gough Island mice (two females/two males) inbred to the fourth generation in the laboratory. Samples were prepared according to the TruSeq PCR Free Sample Preparation kit (Illumina Inc., San Diego, CA) with minor modifications. Libraries were size selected for an average insert size of 550 bp using SPRI-based bead selection. Paired-end, 100-bp sequencing was performed, using Rapid v2 SBS chemistry on an Illumina HiSeq 2500 sequencer. Average sequencing depth was 10 \times per sample. Reads were mapped to the mm10

(GRCm38) mouse reference genome using the BWA 0.7.10 (Li and Durbin 2009), PCR duplicates were marked using Picard 1.119, and indel realignment was done using GATK version 3.3. For each individual, SNPs were called with SAMtools 1.0 (Li 2011) and filtered with QUAL ≥ 20 and coverage ≥ 3 ; SNPs from these samples were then merged using BCFtools 1.0 (supplementary table 7 available at <https://doi.org/10.5061/dryad.cjsxksn34>).

3. Mouse ENCODE data sets were obtained from www.encodeproject.org (last accessed June 26, 2020). We located 16 data sets mapped to the mm10 (GRCm38) genome assembly that were generated using mouse tissue collected at time points that matched our experimental design as closely as possible. We considered nine data sets produced from embryonic liver: six data sets from age E16.5 and three from age E14.5. We obtained five data sets derived from 8-week-old male livers and two from 8-week-old gonadal adipose depots. Despite an imperfect temporal match to our data set, the ENCODE data sets likely provide biological insight into the regulatory landscape around DE genes; the activity of temporally restricted enhancers is known to span a wide range of time points, from several days during embryonic development to weeks in postnatal mice (Nord et al. 2013). [Supplementary table 8, Supplementary Material](#) online, lists the Mouse ENCODE data sets used, including the Experiment IDs for each data set and the file that was downloaded from the website.

As most ENCODE experiments consist of more than one biological replicate, we obtained the “replicated peaks” file for 11 of the 16 data sets; these files contain high-quality overlapping peaks from all replicates. However, replicated peaks files were not available for three data sets cataloging DNase1 hypersensitive sites in E14.5 livers from male 129 mice, and two data sets cataloging DNase1 sites in 8-week-old gonadal adipose depots from male C57BL/6 mice. In these instances, we employed a conservative approach to collapse these sequence ranges into single files most likely to represent biologically relevant peaks—one for DNase1 sites in embryonic liver and one for DNase 1 sites in adult adipose depots. The coordinates for the DNase1 sites in each data set were converted into GenomicRanges objects using the R package GenomicRanges (Lawrence et al. 2013). Then, the reduce function was used to join overlapping peak sequences in each file. Shared peak sequences among replicates (or experiments) generated from similar conditions (three from E14.5 livers and two from 8-week adipose) were then identified using the intersect function. This function ensures the retention of only the smallest sequence range shared among the intersected data sets.

We obtained 16 data sets from publications that identified putative gene regulatory elements in mouse 4-week to 12-week liver tissue or the 3T3-L1 cell line, a model preadipose and adipocyte differentiation cell line

(see references in [supplementary table 9, Supplementary Material](#) online). Adipogenesis in gonadal adipose depots is highly plastic—responding to developmental age, temperature, and diet—indicating that findings from the 3T3-L1 cell line could provide insights into differences in differentiation between island and mainland adipose depots (Han et al. 2011; Wang et al. 2013). Data sets were taken directly from published tables or obtained from personal communication with contributing authors. For our analysis, we considered a wide range of gene regulatory element indicators, including histone modification marks, specific TFBSs, general TFBSs, and enhancer RNAs ([supplementary table 9, Supplementary Material](#) online).

4. Across species, conservation among noncoding sequences might indicate a gene regulatory role for the conserved sequence. We used the UCSC Genome Browser Table Browser tool to obtain coordinates for conserved noncoding sequences among vertebrates (Karolchik et al. 2004). The selected options included genome: mouse, assembly: GRCm38/mm10, group: Comparative Genomics, track: Conservation, and for the table option we selected, in turn, three options that utilize aligned sequence information from three sets of vertebrates—Euarchontoglires (21 rodent and primate species), placental mammals (40 species), and vertebrates (60 species) (actual options selected were phastConsElements60wayEuarchontoglires, phastConsElements60wayPlacental, and phastConsElements60way). All returned conserved sequence elements had conservation scores of 200 (default) or higher.

Results

Organ-Specific Patterns of Gene Expression Evolution

Two broad patterns emerged when looking at differential expression in the adipose, hypothalamus, and liver. First, island and mainland mice exhibit substantial gene regulatory evolution—both in the number of genes and in the magnitude of differential expression. Between 21% and 48% of expressed genes are DE between strains in each condition, and approximately one quarter of these DE genes differ between strains by an absolute \log_2 magnitude of 1 ($\geq 2\times$ or $\leq 1/2\times$) (fig. 1A). Although certainly not all of these expression differences impact weight accumulation, the extensive divergence in transcript levels across three organs provides a rich data set from which hypotheses about the roles of specific genes and pathways in island gigantism evolution can be explored. The second pattern concerns genes expressed throughout liver development. Although there are more DE genes of any magnitude in 4-week livers (fig. 1A) and although genes exhibiting a wide range of expression differences, from low to high, increase in number beyond the

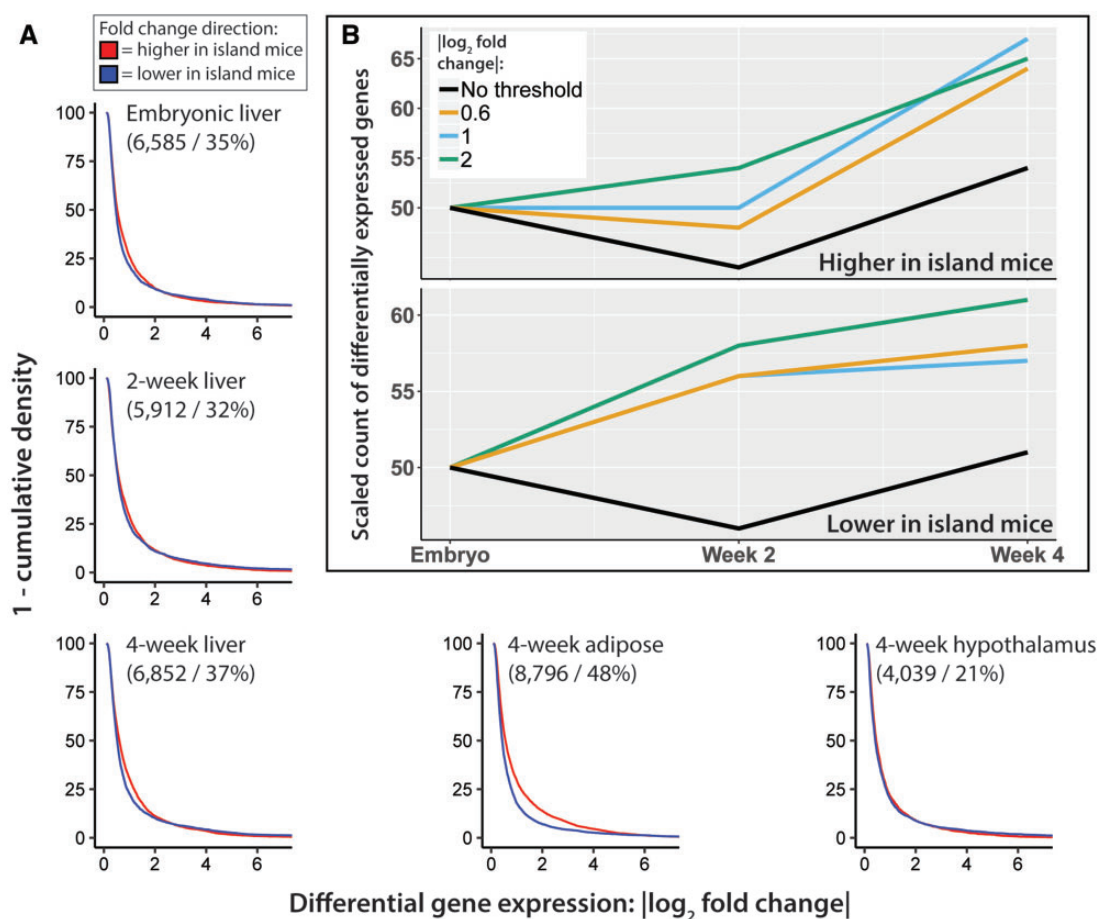


Fig. 1.—Thousands of genes expressed in metabolic organs are transcribed at different levels in island mice relative to mainland mice. (A) Reverse cumulative distribution plots showing the distribution of magnitudes of differential expression for five different conditions. For each condition, the count of DE genes and the percentage DE among all detected transcripts (in parentheses) are provided. (B) Genes expressed throughout liver maturation that are DE (either higher or lower in island mice) at some or all evaluated time points are sorted into four \log_2 fold change classes (absolute value of \log_2 fold change): no threshold (or cutoff), 0.6, 1, and 2. For the liver, the number of genes manifesting large magnitudes of differential expression increases over time. Counts of DE genes are scaled down and given a common origin so that all magnitudes of differential expression can be viewed together. See raw gene counts in [supplementary figure 9, Supplementary Material](#) online.

embryonic time point, it is notable that the percent increase is especially pronounced for genes manifesting large magnitudes of differential expression (fig. 1B). In other words, the elevated number of DE genes in 2- and 4-week livers is not just due to an increase in genes exhibiting relatively small magnitudes of differential expression (tests for a relationship between time point and number of DE genes in a \log_2 fold change class: 1) all DE genes regardless of fold change, which is equivalent to no $|\log_2 \text{fold change}|$ threshold: $X^2(2) = 165.4$, $P < .001$; 2) $|\log_2 \text{fold change}|$ threshold = 0.6: $X^2(2) = 80.9$, $P < 0.001$; 3) $|\log_2 \text{fold change}|$ threshold = 1: $X^2(2) = 42.4$, $P < 0.001$; and 4) $|\log_2 \text{fold change}|$ threshold = 2: $X^2(2) = 16.9$, $P < 0.001$). Among DE genes higher in island mice, this trend is most pronounced from 2 to 4 weeks, where the number of DE genes with a \log_2 fold change ≥ 0.6 or 1 increased by 34% (blue and gold lines in fig. 1B, upper panel). In contrast, among DE genes lower in island mice, this

trend is a bit more subtle; the trend is most pronounced from the embryonic to the 2-week time point, where the number of DE genes with a \log_2 fold change ≤ -2 increased by 16% (green line in fig. 1B, lower panel) (see [supplementary fig. 9, Supplementary Material](#) online, for raw count of DE genes). Our temporal investigation of divergent gene expression in the liver spans two important developmental transitions for house mice (Walthall et al. 2005): the fetal-to-postnatal transition and the parental care-to-weaning transition (includes a suckling-to-solid food shift and increased periods of fasting in the absence of parental care), which occurs at 3 weeks for mice in our colony. Though temporal resolution is low, the accumulation of genes displaying large magnitudes of differential expression at time points flanking the parental care-to-weaning transition (3 weeks) suggests that island mice livers evolved to respond differently to this transition more strongly than the earlier transition. Additionally, this pattern might

simply reflect accumulating differences in transcriptional networks as the liver matures, including through key developmental transitions (Walthall et al. 2005; Hart et al. 2009; Klaassen and Aleksunes 2010; Septer et al. 2012).

Overlap of DE Genes with Body Size QTL

To aid in the nomination of candidate genes and mutations for the evolution of gigantism, we identified the DE genes from all five evaluated conditions that reside within each of the 14 previously identified body weight and growth rate QTL peaks (Gray et al. 2015). We found a total of 363 DE genes within the QTL peaks (subsequently referred to as QTL-DE genes; [supplementary table 10, Supplementary Material online](#)). Although all QTL-DE genes are candidates, we reasoned that strong candidates for gigantism evolution would include those with high magnitudes of differential expression, with fixed SNPs in putative regulatory sequence space, and with a subset of those SNPs overlapping established conserved sequences and/or organ-specific regulatory elements ([fig. 2, table 1 and supplementary tables 8 and 9, Supplementary Material online](#)). These criteria point to molecular genetic changes that could alter gene transcription in genomic regions already known to be involved in body size evolution (Gray et al. 2015) and metabolic organ function.

Partitioning QTL-DE genes in this manner revealed that 180/363 QTL-DE genes exhibit high levels of differential expression with $\geq 1.5\times$ or $\leq 2/3\times$ transcripts (equivalent to an absolute \log_2 fold change ≥ 0.6) in one, two, or more conditions ([table 1 and supplementary table 10, Supplementary Material online](#)). Among these 180 QTL-DE genes with substantial expression differences, 113 harbor a fixed nucleotide difference in their genic regions ([supplementary fig. 10, Supplementary Material online](#)). Among these 113 genes, 73 contain fixed differences between island mice and the mainland strain that overlap one or more putative *cis*-regulatory elements (including Mouse ENCODE elements and highly conserved sequences) characterized in publicly available data sets ([fig. 2, table 1, and supplementary tables 8 and 9, Supplementary Material online; supplementary table 11 available at <https://doi.org/10.5061/dryad.cjxksn34>](#)). Moreover, 28/73 of these genes are annotated to one or more Mammalian Phenotype Ontology (MPO; [Smith et al. 2005](#)) phenotypes related to growth, body size, and metabolism ([table 2](#)). For example, *Irs1* and *Lrp1* are annotated to tens of MPO phenotypes ([table 2](#)), including abnormal fat pad morphology and postnatal growth retardation. Concordantly, *Lrp1* is known to regulate many metabolic processes in the organs in which it is expressed and in which organ-specific regulatory elements overlap its proximal fixed SNPs ([fig. 2 and table 2](#)), including cholesterol storage, glucose and lipid transport, and insulin responsiveness ([Hofmann et al. 2007](#)). Similarly, *Irs1* is one of two major intercellular transmitters of insulin signaling ([Thirone et al. 2006](#)).

Promising candidate genes for extreme size evolution also come from QTL-DE genes that display only a subset of the criteria we used above to learn more about their transcriptional and phenotypic contexts. Four examples illustrate this point. First, none of the fixed SNPs flanking *Egr2* overlap the published regulatory element coordinates we assessed, yet *Egr2* promotes adipocyte differentiation ([Chen et al. 2005](#)) ([table 1 and supplementary fig. 10 and table 10, Supplementary Material online](#)). Second, although *Zbtb16* is not annotated to an MPO phenotype it harbors proximal fixed SNPs that overlap with condition-specific regulatory elements ([fig. 2 and table 1](#)). Polymorphisms in *Zbtb16* are associated with obesity-related traits in humans, whereas in mice this gene influences fat cell-regulated thermogenesis and body weight ([Plaisier et al. 2012; Bendlová et al. 2017](#)). Third, the apolipoprotein gene cluster on chromosome 9 (*Apoa1*, *Apoa3*, *Apoa4*, and *Apoa5*), which harbors high-frequency (but not fixed) island SNPs ([table 1](#)), is associated with plasma triglycerides, obesity risk, and the metabolic syndrome in humans, whereas in mice this gene cluster resides within QTL for body fat and liver lipoprotein metabolism ([Mehrabian et al. 1998; Dallongeville et al. 2008](#)) ([supplementary table 10, Supplementary Material online](#)). Fourth, although *Jmjd1c* exhibits a modest significant fold change magnitude ($|\log_2$ fold change| < 0.6 ; [table 1 and supplementary table 10, Supplementary Material online](#)), its knock down in the 3T3-L1 preadipocyte cell line inhibits adipocyte differentiation ([Buerger et al. 2017](#)).

The majority of QTL-DE genes do not have flanking fixed SNPs that overlap established organ-specific regulatory elements (213/363, 59%; [table 1 and supplementary tables 8 and 9, Supplementary Material online](#)) nor are the majority associated with an MPO phenotype (257/363, 71%; [supplementary table 10, Supplementary Material online](#)). Nonetheless, evidence for the contribution of at least some of the QTL-DE genes to the genesis or maintenance of extreme size comes from the significant enrichment of body size-related phenotypes among the QTL-DE genes when compared with all mouse genes with annotated phenotypes ([table 3](#)). Notably, numerous QTL-DE genes from numerous body size QTL drive the enrichment of each MPO phenotype, providing evidence that at least some of the physiological and metabolic processes that evolved to confer large size in island mice might be fostered by regulatory changes at multiple loci, and possibly in multiple organs, as would be expected for a complex trait like gigantism.

Coregulated Gene Groups in the Liver

We combined patterns of coordinated and differential expression in the liver across three time points to form putative coregulated gene groups. From the perspective of gene regulatory evolution, the differential expression patterns

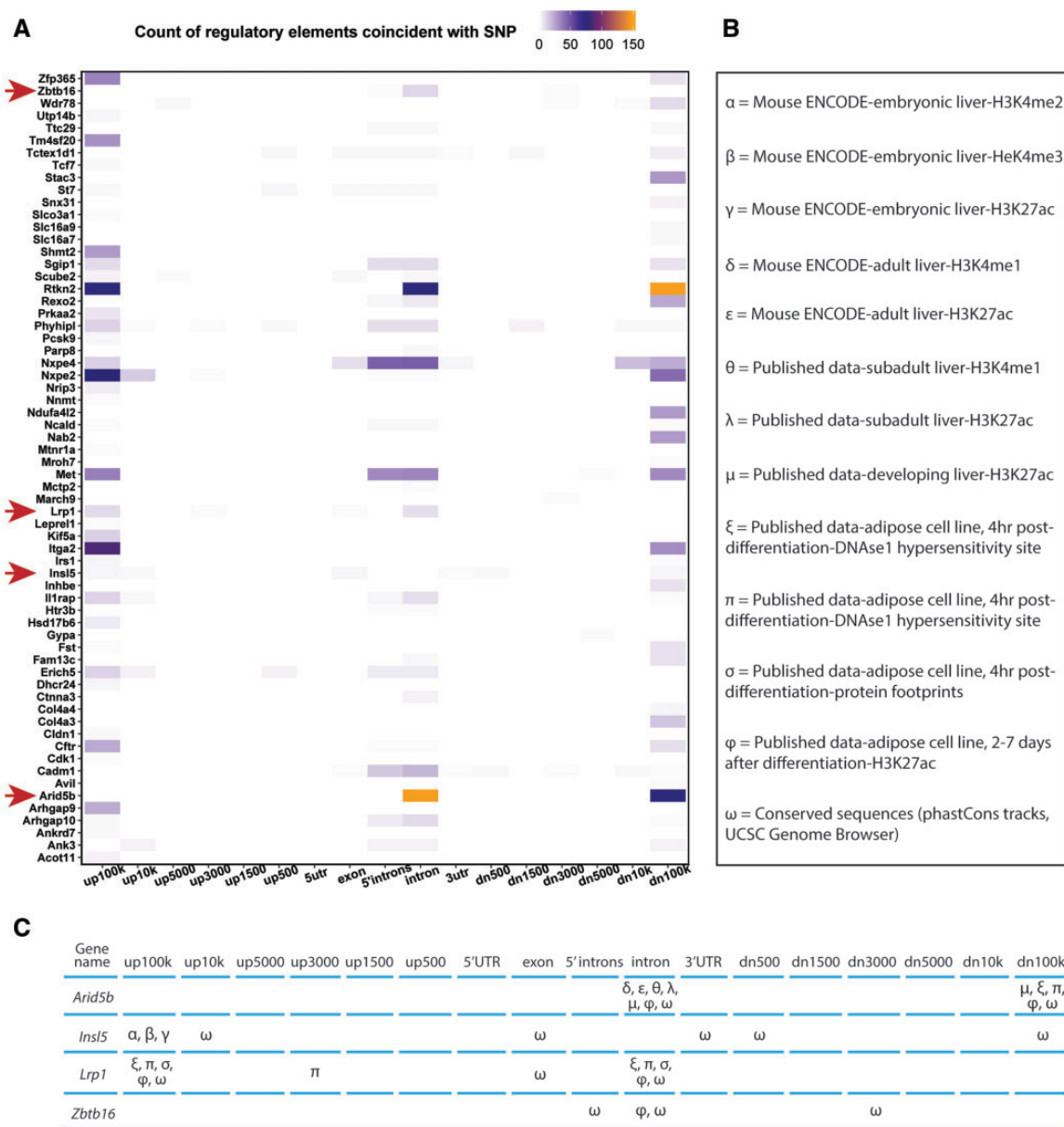


FIG. 2.—A subset of DE genes residing within body size QTL (QTL-DE genes) exhibit fixed SNPs coincident with putative gene regulatory elements. (A) Established gene regulatory elements (REs)—drawn from 29 public data sets (supplementary tables 8 and 9, Supplementary Material online)—overlap with fixed SNPs in 73 QTL-DE genes. Counts of SNP-RE intersections in genic regions are not mutually exclusive: for example, if a SNP occurs in different genic regions in different transcript variants then its coincidence with an RE will be counted for each genic region. Counts are for any position between two adjacent genic regions on the x axis; for example, all SNP-RE co-occurrences between upstream position 501 and 1,500 from the transcriptional start site would be included in the up500 column. Spanning over 1 Mb, gene *Dab1* bears many fixed SNPs and was removed from panel (A) for clarity of the heat map; likewise, genes with long alphanumerical names were removed for clarity. Red arrows in (A) indicate QTL-DE genes highlighted in (C), where genes *Arid5b*, *InsI5*, *Lrp1*, and *Zbtb16* are shown with a subset of fixed SNPs that overlap with putative organ-specific REs from a wide array of public data sets (B), details of which can be found in supplementary tables 8 and 9, Supplementary Material online. Panel (C) does not provide the total number of SNP-RE coincidences, but rather indicates the types of putative REs that overlap with fixed SNPs. In all panels, “5' introns” refers to either the first or second intron of a transcript; 5' intronic counts are also included in the total count for the “intron” genic region. “dn” and “up” refer to down- and upstream, respectively, and the numbers refer to base pairs (k = kilobases).

manifested by the genes in each coregulated group can be traced to two sources, which are not mutually exclusive: 1) differences (e.g., SNPs) in TFBSs in regulatory elements of a

group’s genic members and 2) differential expression of the TFs that modulate the expression levels of a group’s genic members. To evaluate both sources of differential expression,

Table 1
DE Genes (Island vs. Mainland) in Body Size QTL: Sorted According to Magnitude of Differential Expression, SNP Status, and SNP-Regulatory Element Overlap

363 Total DE Genes (±2 Mb around QTL Peaks)									
Larger DE Magnitude: log ₂ fold change ≥ 0.6 (n = 180)					Smaller DE Magnitude: log ₂ fold change ≤ 0.6 (n = 183)				
Genes with Fixed Flanking SNPs					Genes with Fixed Flanking SNPs				
SNPs in Regulatory Element(s) ^a	Remaining Genes	Genes with High-Frequency Flanking SNPs ^b	SNPs Not in Regulatory Element(s)	Remaining Genes	SNPs in Regulatory Element(s) ^a	Remaining Genes	Genes with High-Frequency Flanking SNPs ^b	SNPs Not in Regulatory Element(s)	Remaining Genes
Acot11	Nnmt	Ampd3	Btbd10	Acs1	Acs3	Mgat4b	Apof	Adamt2	Abce1
Ank3	Nrip3	Ankrd52	Cav2	Ankrd37	Agfg1	Mier1	Bace1	Adm	B3gnt7
Ankrd7	Nxpe2	Apoa1	Ctdsp2 ^c	Bloc1s1	Anapc10	Mocs2	Cep164	Ado	B4galnt1
Arhgap10	Nxpe4	Apoa4	Cyp4v3	Ccdc110	Ankrd46	Mogat1	Cfap97	Api3	Cab39
Arhgap9	Parp8	Apoa5	Dock10	Cd63	Arhgef25	Naca	Cnpy2	Cav1	Esyf1
Arid5b	Pcsk9	Apoc3	Egr2	Dkk3	Arl15	Ndufs4	Cocq10a	Ccct6	Far1
Avil	Phyhip1	Apon	Fam151a	Erbp3	Atp5b	Nr3c2	Cox6c	Cdk2	Gdf11
Cadm1	Prkaa2	Ccl20	Fam174b	Gdf9	Baz2a	Nrbf2	Cs	Cnot6	Hnmph1
Cdk1	Rexo2	Cenpu	Fbxo36	Gm9945	C8b	Oma1	Ctr9	Cul3	Hspa4
Cftr	Rtkn2	Daw1	Fgf12	Gpr55	Canx	Pelo	Hmnpab	Dgka	Leap2
Cldn1	Scube2	Emb	Galnt18	Htr2b	Capza2	Pid1	Leap2	Dnajc14	Lsm6
Col4a3	Sgjp1	Fstf4	Gfp2	Kcne4	Ccdc50	Pip4k2c	Lsm6	F11	N4bp3
Col4a4	Shmt2	Gls2	Gm16025	Kif3a	Cdk4	Ppap2b	Nabp2	Fam149a	Nabp2
Ctnna3	Slc16a7	Il23a	Gm28046	Lrig3	Cdkb	Prim1	Ormdl2	Klff10	Ormdl2
Deb1	Slc16a9	Irga7	Gm29125	Lrp2bp	Chd2	Prmt10	Poir2k	Mmaa	Poir2k
Dhcr24	Slco3a1	Lyve1	Gm4890	Myf6	Ctdsp2 ^c	R3hdm2	Rad50	Mmp19	Rad50
Erich5	Snx31	Mettl7b	Gmnc	Myf6b	Ctnnb2	Reep3	Rbms2	Os9	Rbms2
Fam13c	St7	Mical2	Gpr182	Osr2	Dctn2	Rhobtb1	Rbmxl1	Pabpc1	Rbmxl1
Fst	Stac3	Micalc1	Ikzf4	Pollim3	Ddit3	Rnf130	Rdh5	Ppp2ca	Rdh5
Gm26809	Tcf7	Rdh19	Ipo7	Psm1	Dennd5a	Rnf19a	Rpl41	Rab5b	Rpl41
Gm9747	Tctex1d1	Rdh7	Klkb1	Sarnp	Dner	Sar1b	Rrm2b	Rdh1	Rrm2b
Gypa	Tm4sf20	Rgs22	Pmel	Serpine2	Dtx3	Scg2	Rufy1	Rhbdd1	Rufy1
Hsd17b6	Ttc29	Sik3	Rps26	Shroom1	Ednra	Slc35d1	Sept8	Sbf2	Sept8
Htr3b	Utp14b	Sp140	Slc16a14	Slc26a10	Epha4	Smarca5	Slc10a7	Skp1a	Slc10a7
Il1rap	Wdr78	Tagln	Slc19a3	Tespa1	Farsb	Sqstm1	Smarcc2	Smad1	Smarcc2
Inhbe	Zbtb16	Tes	Sorbs2	Tmem26	Fat1	Ssbp3	Sowaha	St8sia2	Sowaha
Insl5	Zfp365	Tfec	Sphkap	Uqcrq	Gab1	Stat6	Spag1	Tlr3	Spag1
Irs1	a ^d	Timeless	Suox	Vmn2r84	Hhip	Tbc1d9b	Sp100	Ttc4	Sp100
Irga2	b ^d	Ttc12	Trp63	v ^d	Hrsp12	Tmem194	Sp110	Ubr5	Sp110
Kif5a	c ^d	Vps13b	Zbtb39	w ^d	Inhbc	Tsfm	Spryd4	Usp47	Spryd4
Leprel1	d ^d	Zfp879	Zfp879	x ^d	Irf2	Usp24	Stat2	Vdac1	Stat2
Lrp1	e ^d	f ^d	g ^d		Irga1	Usp28	Tead1	Wdfy1	Tead1
March9			h ^d		Jmjd1c	Wnt2	Zfp2	Wee1	Zfp2
							Zfp354b		Zfp354b

(continued)

Table 1 Continued

363 Total DE Genes (± 2 Mb around QTL Peaks)

Larger DE Magnitude: $|\log_2 \text{ fold change}| \geq 0.6$ ($n = 180$)

Smaller DE Magnitude: $|\log_2 \text{ fold change}| \leq 0.6$ ($n = 183$)

Genes with Fixed Flanking SNPs		Genes with Fixed Flanking SNPs	
SNPs in Regulatory Element(s) ^a	SNPs Not in Regulatory Element(s)	SNPs in Regulatory Element(s) ^a	SNPs Not in Regulatory Element(s)
Remaining Genes	Genes with High-Frequency Flanking SNPs ^b	Remaining Genes	Genes with High-Frequency Flanking SNPs ^b
<i>Mctp2</i>	$f^{c,d}$	<i>Lpp</i>	<i>Zfp354a</i>
<i>Met</i>	$f^{c,d}$	<i>Lsm8</i>	<i>Zfp706</i>
<i>Mroh7</i>	k^d	<i>Ltc4s</i>	<i>Zfp143</i>
<i>Mtnr1a</i>	f^d	<i>Maml1</i>	<i>Zfp354c</i>
<i>Nab2</i>	m^d	<i>Mars</i>	<i>Zfp454</i>
<i>Ncald</i>	n^d	<i>Mettl1</i>	aa^d
<i>Ndufa4l2</i>	o^d	<i>Mff</i>	bb^d
			<i>Xrcc6bp1</i>
			<i>Zfp827</i>
			<i>Zpr1</i>
			cc^d

^aPutative published and publicly available regulatory elements listed in [supplementary tables 8 and 9](#), [Supplementary Material](#) online.

^bHigh-frequency SNP = frequency across 16 sequenced wild-caught Gough Island mouse genomes ≥ 0.9 .

^cIn conditions where $|\log_2 \text{ fold change}| \leq 0.6$ fixed SNPs flanking these genes overlap condition-specific regulatory elements.

^dLonger gene names listed here: (a) BC055111, (b) F420014N23Rik, (c) Gm10384, (d) 1110002J07Rik, (e) 1700040L02Rik, (f) 3010026O09Rik, (g) A230028O05Rik, (h) A930033H14Rik, (i) BC035947, (j) Mettl21b^c, (k) Rasgef1c, (l) 1600010M07Rik, (m) 1700024P16Rik, (n) 5730419F03Rik, (o) 9830004L10Rik, (p) A630001G21Rik, (q) BC089597, (r) D830026I12Rik, (s) G530012D18Rik, (t) 9030616G12Rik, (u) D930048N14Rik, (v) 1700029J07Rik, (w) 2900052N01Rik, (x) 4933407L21Rik, (y) Cdkn2aipnl, (z) Trmem184c, (aa) RP23-32A8.1, (bb) 9530068E07Rik, (cc) Patah1b2, and (dd) A430046D13Rik.

Table 2
DE Genes (Island vs. Mainland) in Body Size QTL with Large Fold Changes, Fixed SNPs in Putative Regulatory Elements^a, and Annotated Mammalian Phenotypes

Gene Name	Ensemble Gene ID	Chromosome/ QTL ^b	± Distance QTL Peak (Mb)	QTL-Associated Trait	Condition ^c	Differential Expression (log ₂ fold change)	Mammalian Phenotype Ontology Phenotypes ^d
<i>Met</i>	ENSMUSG00000009376	chr6	1	Body weight, growth rate	wk4a, wk4h	-0.8, -0.44	Abnormal fetal growth/weight/body size and liver development, decreased circulating insulin level, embryo size, fetal size, insulin secretion and liver weight, enlarged liver sinusoidal spaces, hepatic steatosis, impaired glucose tolerance, increased circulating glucose level, postnatal growth retardation
<i>Igf2</i>	ENSMUSG00000015533	chr13	1	Growth rate	el, wk4l	-1.73, 2.72	Homeostasis/metabolism phenotype
<i>Cdk1</i>	ENSMUSG00000019942	chr10	1	Body weight	wk4a, wk4l	0.45, 1.34	Abnormal hepatocyte morphology, decreased hepatocyte number and hepatocyte proliferation
<i>Cldn1</i>	ENSMUSG00000022512	chr16	1	Growth rate	el, wk2l, wk4a, wk4h, wk4l	-0.64, -0.32, -2.54, -0.79, -0.52	Weight loss
<i>Prkaa2</i>	ENSMUSG00000028518	chr4	1	Growth rate	el	1.31	Abnormal gluconeogenesis, decreased body weight, circulating insulin level, glycogen level and insulin secretion, homeostasis/metabolism phenotype, impaired glucose tolerance, increased body weight, circulating free fatty acid level, circulating glucose level, circulating insulin level, circulating leptin level, fatty acid level, susceptibility to diet-induced obesity and triglyceride level, insulin resistance
<i>Dab1</i>	ENSMUSG00000028519	chr4	1	Growth rate	el, wk2l, wk4a, wk4l	1.06, 2.17, 2.12, 2.17	Decreased body size
<i>Rexo2</i>	ENSMUSG00000032026	chr9	1	Body weight	el, wk4a	0.92, 0.41	Increased circulating HDL cholesterol level
<i>Cadm1</i>	ENSMUSG00000032076	chr9	1	Body weight	wk2l, wk4h, wk4l	0.83, -0.16, 0.53	Decreased body size, body weight and fetal size
<i>Cftr</i>	ENSMUSG000000041301	chr6	1	Body weight, growth rate	wk2l	-1.63	Abnormal hepatocyte morphology, decreased body length, body size and body weight, homeostasis/metabolism phenotype, liver/biliary system phenotype, postnatal growth retardation, weight loss
<i>Ncald</i>	ENSMUSG000000051359	chr15	1	Growth rate	wk4a, wk4h	0.97, 0.14	Decreased body length, circulating HDL cholesterol level, circulating glucose level, lean body mass and total body fat amount
<i>Utp14b</i>	ENSMUSG000000079470	chr1	1	Body weight	wk4a	1.16	Growth/size/body region phenotype

(continued)

Table 2 Continued

Gene Name	Ensemble Gene ID	Chromosome/ QTL ^b	± Distance Flanking QTL Peak (Mb)	QTL-Associated Trait	Condition ^c	Differential Expression (log ₂ fold change)	Mammalian Phenotype Ontology Phenotypes ^d
<i>Arid5b</i>	ENSMUSG00000019947	chr10	1–2	Body weight, growth rate	el, wk2l, wk4a	–0.6, –0.74, –1.1	Abnormal fat cell morphology, decreased body length, body size, body weight, brown fat lipid droplet number, susceptibility to diet-induced obesity and white adipose tissue amount, increased food intake, postnatal growth retardation
<i>Nab2</i>	ENSMUSG00000025402	chr10	1–2	Body weight, growth rate	wk2l, wk4l	0.52, –1.09	Decreased body length, circulating HDL cholesterol level and circulating cholesterol level
<i>Scube2</i>	ENSMUSG00000007279	chr7	2	Body weight, growth rate	wk2l, wk4h	–2.55, 0.54	Decreased body length, body size, and body weight
<i>Htr3b</i>	ENSMUSG00000008590	chr9	2	Body weight	el, wk2l, wk4l	–5.88, –8.6, –8.04	Increased body length
<i>Fst</i>	ENSMUSG000000021765	chr13	2	Growth rate	wk2l, wk4a, wk4l	–1.63, –1.21, –1.79	Decreased fetal size and weight, fetal growth retardation, homeostasis/metabolism phenotype
<i>Shmt2</i>	ENSMUSG00000025403	chr10	2	Body weight, growth rate	wk2l, wk4l	–0.56, –0.77	Decreased embryo size
<i>Acot11</i>	ENSMUSG000000034853	chr4	2	Growth rate	el, wk2l, wk4h	–0.72, –0.75, 0.62	Abnormal brown adipose tissue thermogenesis and brown fat cell morphology, decreased circulating HDL cholesterol level, circulating cholesterol level, circulating glucose level, lean body mass, susceptibility to diet-induced obesity and susceptibility to hepatic steatosis, homeostasis/metabolism phenotype, improved glucose tolerance, increased brown adipose tissue amount, circulating triglyceride level, fatty acid level, food intake and oxygen consumption
<i>Dhcr24</i>	ENSMUSG000000034926	chr4	2	Growth rate	wk2l, wk4a, wk4l	–0.44, 1.1, –0.36	Decreased body size, body weight, circulating cholesterol level and circulating triglyceride level, postnatal growth retardation
<i>Lrp1</i>	ENSMUSG000000040249	chr10	2	Body weight, growth rate	el, wk4a, wk4l	–0.55, –0.66, –0.53	Abnormal adipose tissue physiology, brown fat cell morphology, embryo size, fat pad morphology, lipid homeostasis, lipid level and liver physiology, decreased body weight, brown fat lipid droplet number, circulating free fatty acid level, circulating glucose level, circulating insulin level, hepatocyte number, liver triglyceride level, susceptibility to diet-induced obesity and

<i>Stac3</i>	ENSMUSG00000040287	chr10	2	Body weight, growth rate	el, wk2l, wk4a, wk4h, wk4l	0.55, 1.18, 1.06, 0.7, 1.05	Decreased fetal weight	total body fat amount, embryonic growth retardation, homeostasis/metabolism phenotype, improved glucose tolerance, increased food intake and oxygen consumption, postnatal growth retardation
<i>March9</i>	ENSMUSG00000040502	chr10	2	Body weight, growth rate	el, wk2l, wk4l	0.97, 1.08, 0.69	Decreased body weight and circulating HDL cholesterol level	
<i>Pcsk9</i>	ENSMUSG00000044254	chr4	2	Growth rate	wk4a	1.32	Abnormal cholesterol homeostasis, decreased circulating HDL cholesterol level, circulating LDL cholesterol level and circulating cholesterol level, homeostasis/metabolism phenotype, increased circulating HDL cholesterol level	
<i>Irs1</i>	ENSMUSG00000055980	chr1	2	Growth rate	wk4a, wk4l	-1.57, 0.63	Abnormal lipid level, decreased body length, body size, body weight, circulating HDL cholesterol level, circulating glucose level, fetal weight, insulin secretion and total body fat amount, fetal growth retardation, homeostasis/metabolism phenotype, impaired glucose tolerance and lipolysis, increased circulating free fatty acid level, circulating insulin level and circulating triglyceride level, insulin resistance, postnatal growth retardation	
<i>Ins5</i>	ENSMUSG00000066090	chr4	2	Growth rate	el, wk4h	6.47, -0.57	Decreased circulating insulin level, impaired glucose tolerance	
<i>Col4a3</i>	ENSMUSG00000079465	chr1	2	Growth rate	el, wk2l, wk4l	-0.62, -2.08, -2.21	Decreased body weight	
<i>Col4a4</i>	ENSMUSG00000067158	chr1	2	Growth rate	wk2l, wk4l	-0.68, -0.85	Decreased body size and body weight	
<i>Kif5a</i>	ENSMUSG00000074657	chr10	2	Body weight, growth rate	wk4a, wk4h	1.06, 0.27	Decreased body size and body weight, postnatal growth retardation	

^aPutative published and publicly available regulatory elements listed in [supplementary tables 8 and 9](#), [Supplementary Material](#) online.

^bGray et al. (2015).

^cwk4a = 4-week gonadal adipose depot; wk4h = 4-week hypothalamus; el = E16.5 liver; wk2l = 2-week liver; wk4l = 4-week liver.

^dReported phenotypes manually curated to remove verbal redundancy.

Table 3

Significantly Enriched Mammalian Phenotype Ontology Phenotypes among QTL-DE Genes

Mammalian Phenotype Ontology Phenotypes ^a	Count of QTL-DE Genes Linked to Phenotype ^b	Count of Body Size QTL (out of 14) Harboring DE Genes
Abnormal circulating cholesterol level	22	8
Abnormal circulating HDL cholesterol level ^c	16–17	8
Decreased circulating HDL cholesterol level ^c	13–14	7
Abnormal lipid level ^c	40–41	11
Abnormal lipid homeostasis ^c	40–41	11
Abnormal cholesterol level	24	8
Abnormal circulating lipid level ^c	34–35	10
Decreased cholesterol level ^c	16–17	7
Decreased circulating cholesterol level ^c	15–16	7
Abnormal cholesterol homeostasis	24	8
Homeostasis/metabolism phenotype	103	14
Abnormal lipoprotein level ^c	18–19	8
Abnormal circulating lipoprotein level ^c	18–19	8
Abnormal triglyceride level	19	9
Increased circulating triglyceride level	11	7
Abnormal circulating triglyceride level	17	9
Abnormal fatty acid level ^c	15	8
Abnormal body weight	52	14
Abnormal brown adipose tissue thermogenesis	3	2
Abnormal circulating free fatty acids level	10	7
Abnormal free fatty acids level	11	8
Abnormal body length	16	11
Decreased body length	13	9
Decreased body weight	45	14
Decreased total tissue mass	45	14
Abnormal prenatal growth/weight/body size	27	13
Abnormal prenatal body size	24	13
Abnormal cell proliferation ^c	36	13
Decreased cell proliferation	25	12

^aEnrichment carried out in MouseMine: Motenko et al. (2015). All adjusted *P* values < 0.05.^bThe number of QTL-DE genes annotated to the phenotype in a Hypergeometric Test with background equal to all genes with phenotypic annotation for *Mus domesticus*.^cSignificant after multiple test correction using both Bonferroni and Benjamini–Hochberg methods, otherwise significant after Benjamini–Hochberg multiple test correction.

we assessed the enrichment of TF binding motifs in the 5 kb, 5'-upstream sequences of each coregulated group's genes and asked whether TFs whose motifs were enriched were themselves DE. This approach permitted the construction of simple gene regulatory networks that are altered in island mice relative to mainland mice.

We parsed 4,965 genes with dynamic or consistent differential expression patterns within and between strains into one of 285 putative coregulated groups (see Materials and Methods). Among these groups, 224 show statistically significant enrichment for TFBSs in their upstream sequences (supplementary table 12, Supplementary Material online). Overall, there are 253 significantly enriched TFBSs across the 224 coregulated groups (note the following verbal equivalency: enriched TFBS = enriched TF). One hundred and four of these 253 TFs are themselves DE in the liver, and 39 of these DE TFs

are members of one of the coregulated groups with significant upstream TFBS enrichment (supplementary table 12, Supplementary Material online). This suggests that the differential expression of coregulated group genes might be a result of both genetic variation in their putative *cis*-regulatory sequence space (e.g., see fig. 2) and alterations to the expression levels of their upstream-binding TFs (*trans*-regulatory effects).

Of special interest are enriched TFs that reside within body size QTL (QTL-TF; Gray et al. 2015) as they might regulate numerous DE genes that underlie functional differences between island and mainland metabolic organs. Four enriched TFs are QTL-TFs: *Arid5b*, *Egr2*, *Sp100*, and *Tfec*. Except for *Egr2*, these QTL-TFs are themselves DE in the liver (supplementary table 10, Supplementary Material online). *Arid5b*, *Egr2*, *Sp100*, and *Tfec* putatively regulate 40, 1, 3, and 28

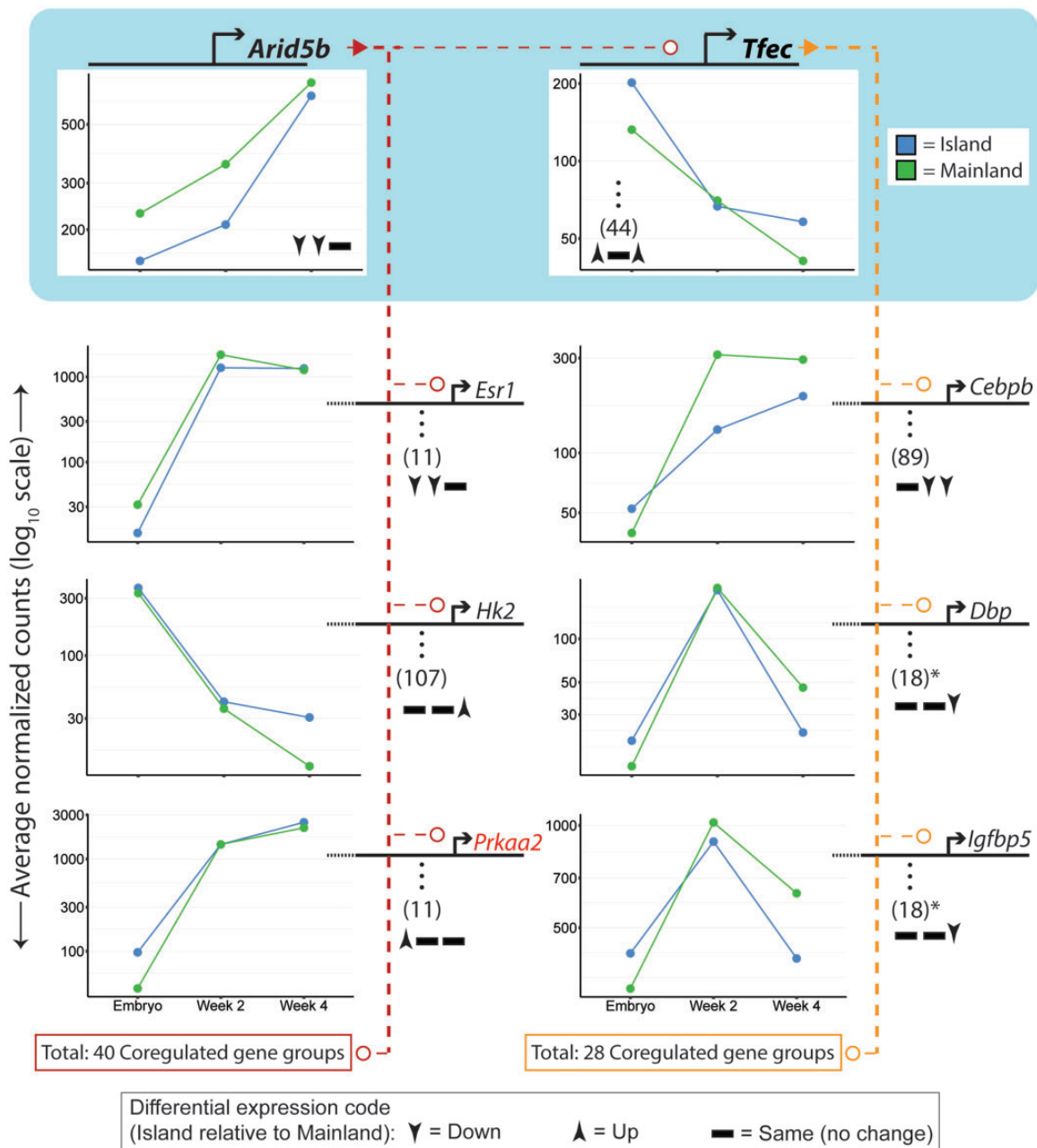


Fig. 3.—Transcriptional regulation of coregulated gene groups throughout liver maturation. TFs *Arid5b* and *Tfec* reside within body size-related QTL (light blue box) mapped from an intercross between island and mainland mice (Gray et al. 2015). Colored dashed lines emanating from either *Arid5b* or *Tfec* connect to coregulated gene groups whose transcription is putatively regulated by these TFs. Circles at the terminus of colored, dashed lines indicate transcriptional regulation as evidenced by enrichment of TF binding motifs in upstream sequences of coregulated group members. Except for one instance (indicated by asterisks), one representative gene is shown for each coregulated gene group displayed; the number of genes within a group is indicated by a vertical ellipsis and associated count within parentheses. Red text indicates a DE gene that also resides within a body size QTL.

coregulated groups, respectively (fig. 3). Among all the 253 enriched TFs, *Arid5b* and *Tfec* contribute regulation to more coregulated groups than 95% and 90% of the other enriched TFs, respectively. Moreover, considering all the genes in their coregulated groups, *Arid5b* and *Tfec* putatively contribute to the regulation 776 and 790 DE genes, respectively (jointly,

32% of the 4,965 DE genes that were partitioned into coregulated groups) (supplementary table 12, Supplementary Material online). This indicates that *Arid5b* and *Tfec* might play a disproportionate role in the total differential expression observed in the maturing liver. *Arid5b* is known to regulate numerous TFs essential for switch points in the adipocyte

differentiation pathway and is also active in hepatocytes (Baba et al. 2011; Claussnitzer et al. 2015).

The proximity of *Arid5b* and *Tfec* to the QTL peaks and their role in regulating hundreds of genes strengthens their nomination as candidate genes with prominent roles in the evolution of gigantism in island mice. Notably, *Tfec* belongs to a coregulated group whose upstream sequences are enriched for *Arid5b* binding sites, and both QTL-TFs regulate genes with well-characterized roles in liver differentiation and metabolism (fig. 3). For example, *Tfec* putatively regulates two genes, *Dbp* and *Igfbp5*, that belong to the same coregulated group and that likely participate in circadian rhythm transcription and body weight regulation, respectively (Wuarin et al. 1992; Salih et al. 2004). *Arid5b* regulates multiple genes from many coregulated groups that have been linked to the liver's ability to process glucose and lipids, including *Hk2* and *Prkaa2*, the latter of which also resides within a body size QTL (Ruderman et al. 2003; Panasyuk et al. 2012) (supplementary table 10, Supplementary Material online). Additional regulatory connections between the QTL-TFs and genes involved in different aspects of metabolic or transcriptional regulation in the liver, including *Cebpb* and *Esr1*, are indicated by the TFBS enrichment analysis (fig. 3 and supplementary table 12, Supplementary Material online) (Westmacott et al. 2006; Qiu et al. 2017).

Functional Enrichment throughout Liver Maturation

To identify cellular and metabolic processes that differ between island and mainland mice throughout liver maturation, we sorted highly specific, significantly enriched GO terms into 38 GO Slim categories (see Materials and Methods). Each GO Slim category encompasses hundreds of GO terms. Figure 4 (and supplementary fig. 11, Supplementary Material online) shows the top one quarter of GO Slim categories containing the greatest quantity of highly specific, enriched GO terms; these include, from highest to lowest, *Other metabolic processes*, *Cell organization and biogenesis*, *Transport*, *Stress response*, *Cell cycle and proliferation*, *Other membranes*, *Developmental processes*, *Signal transduction*, *Mitochondrion*, and *Translational apparatus* (supplementary table 5, Supplementary Material online). Figure 4 simultaneously allows for the visualization of broad functional differences (via the GO Slim color code) and specific alterations in cellular and metabolic processes (via the enriched GO terms along the x axes) between island and mainland livers.

When considering functional enrichment across the evaluated liver time points, three broad patterns are evident (fig. 4 and supplementary fig. 11, Supplementary Material online). First, gene regulatory evolution in the liver was temporally and functionally extensive, affecting a wide range of cellular and metabolic processes from late embryonic development to subadult life stages. In particular, gene regulatory evolution affected a more diverse set of biological functions in

embryonic and 4-week livers (as indicated by the diverse color palette exhibited for E16.5 and 4-week livers in fig. 4) than in 2-week livers (supplementary fig. 11, Supplementary Material online). Second, subsets of biological functions are either consistently associated with higher or lower expressed genes across two assayed time points or they are associated with genes that flip their direction of fold change across time points. For example, immunity and stress response-related functions are enriched among highly expressed genes at both E16.5 and 4 weeks, whereas mitochondrial activity is enriched among higher expressed genes in embryo but enriched among lower expressed genes at 4 weeks. Third, the average difference in expression for genes annotated to these enriched functions ranges from an ~20% to >400% decrease or increase in mRNA levels in island mice relative to mainland mice (fig. 4 and supplementary fig. 11, Supplementary Material online).

Condition-Specific Functional Enrichment

Although some enriched functions may reflect processes that differ among island and mainland mice beyond aspects of body size, below we emphasize significantly enriched GO terms and KEGG pathways in each condition (see Materials and Methods; fig. 4 and supplementary tables 1–4, Supplementary Material online) that might relate to differences in liver maturation and/or body size.

Embryonic Liver

At E16.5, GO terms and KEGG pathways related to mitochondria, metabolism, and immunity are enriched among higher expressed genes in the liver (fig. 4 and supplementary tables 1 and 2, Supplementary Material online). In contrast, general developmental processes are reduced in island mouse embryonic livers, as indicated by lower expression of genes annotated to GO terms such as *Organ morphogenesis* and *Mitotic cell cycle* (fig. 4).

Two-Week Liver

In 2-week livers, the vast majority of significantly enriched GO terms and KEGG pathways are annotated to genes manifesting lower expression in island mice (supplementary fig. 11 and tables 1 and 2, Supplementary Material online). Prominent among these enriched GO terms and KEGG pathways are those related to amino acid catabolism (supplementary fig. 11 and tables 1 and 2, Supplementary Material online). Complementary to this observation is that the KEGG pathway *Lysosome*—a general carbohydrate, lipid, and protein degradation system—is enriched among lower expressed genes. On the other hand, the only enriched KEGG pathway among higher expressed genes is *Ribosome*, indicating potential protein construction (supplementary table 2, Supplementary Material online).

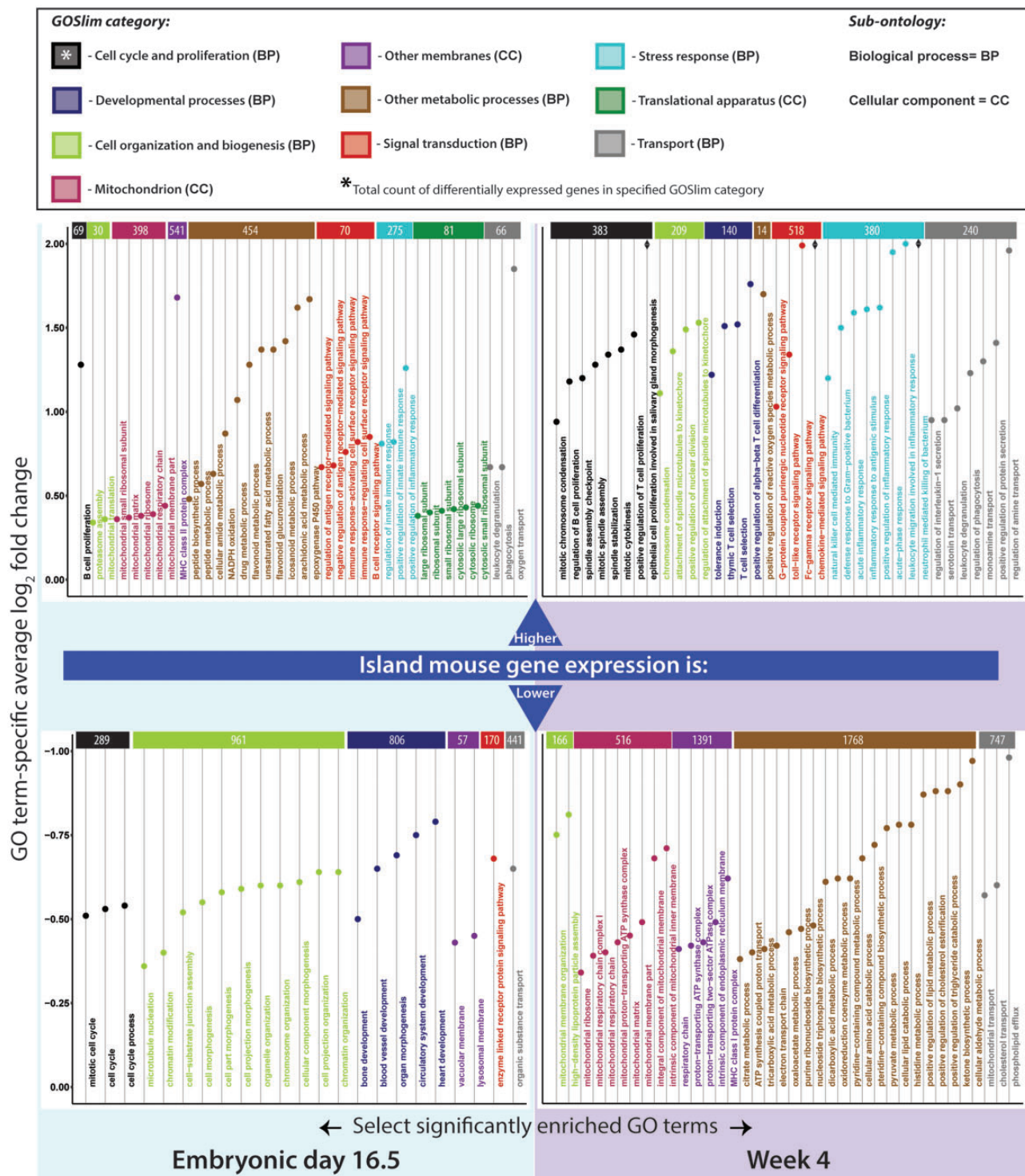


FIG. 4.—Summary of GO functional enrichment for DE genes in the embryonic (E16.5) and 4-week liver. Higher and lower refer to gene expression levels that were higher or lower in island mice relative to mainland mice. Vertical GO terms in all panels are a descriptively specific subset of all significantly enriched GO terms for each time point and fold-change direction (see Materials and Methods). Colors designate GOSlim categories, broad functional categories to which individual GO terms are sorted. The GOSlim color key applies to rectangles at the top of each plot as well; rectangles provide the number of all DE genes pertaining to a GOSlim category regardless if a gene’s annotated GO term is represented in the plot. Plots show the average log₂ fold change for DE genes annotated to an enriched GO term. Black diamonds designate average log₂ fold change values that exceed the maximum value of the plots.

Four-Week Liver

The functions carried out by DE genes in 4-week livers are spread across a wide array of GO Slim categories, indicating that many biochemical and cellular functions differ between island and mainland mice in increasingly mature livers. The majority of enriched GO categories among higher expressed genes are related to the cell cycle and to immune cell functions (fig. 4). This pattern is mirrored in the enriched KEGG pathways, which include *Cell cycle*, *p53 signaling pathway*, *T cell receptor signaling pathway*, and *Cytokine–cytokine receptor interaction* (supplementary table 2, Supplementary Material online). Lower expressed genes provide another prominent signal, that of suppressed metabolic processes, particularly those carried out in the mitochondria. This signal can be seen in the GO terms found within the GO Slim category *Mitochondrion*, and in KEGG pathways *Citrate cycle (TCA cycle)* and *Oxidative phosphorylation* (fig. 4 and supplementary tables 1 and 2, Supplementary Material online).

Four-Week Gonadal Adipose Depot

Numerous significantly enriched KEGG pathways among lower expressed genes in the 4-week adipose are signaling pathways and metabolic processes associated with differentiated fat-storing cells (adipocytes), suggesting the presence of immature preadipocytes in island gonadal adipose depots. Concordantly, these themes—reduced pro-adipocyte signaling pathways and reduced lipid metabolism—are also manifested in the myriad enriched GO categories among lower expressed genes (fig. 5 and supplementary tables 1 and 2, Supplementary Material online). Among the 67 DE genes that contribute to the enrichment of the *Fat cell differentiation* GO term are several with negative, large magnitude fold change values (\log_2 fold change ≤ -1), including *Arid5b*, *Klf5*, *Rgs2*, *Sfrp1*, and *Zbtb16*, all of which play documented roles in adipogenesis (Nishizuka et al. 2001; Oishi et al. 2005; Lagathu et al. 2010; Plaisier et al. 2012; Claussnitzer et al. 2015; Bendlová et al. 2017) (supplementary table 3 available at <https://doi.org/10.5061/dryad.cjsxksn34>). Moreover, both *Arid5b* and *Zbtb16* reside in body size QTL peaks (table 1).

Notably, significantly enriched GO terms among higher expressed genes also contain cellular features related to fat cell differentiation, including enrichment of the Notch signaling pathway and terms related to cilia (20% of enriched GO terms) (fig. 5 and supplementary table 1, Supplementary Material online). Both ciliogenesis and Notch signaling are implicated in early stages of adipocyte differentiation (Garcés et al. 1997; Marion et al. 2009).

Four-Week Hypothalamus

Immunity-related categories comprise the vast majority of significantly enriched GO terms and KEGG pathways annotated to higher expressed genes in the 4-week hypothalamus (fig. 5

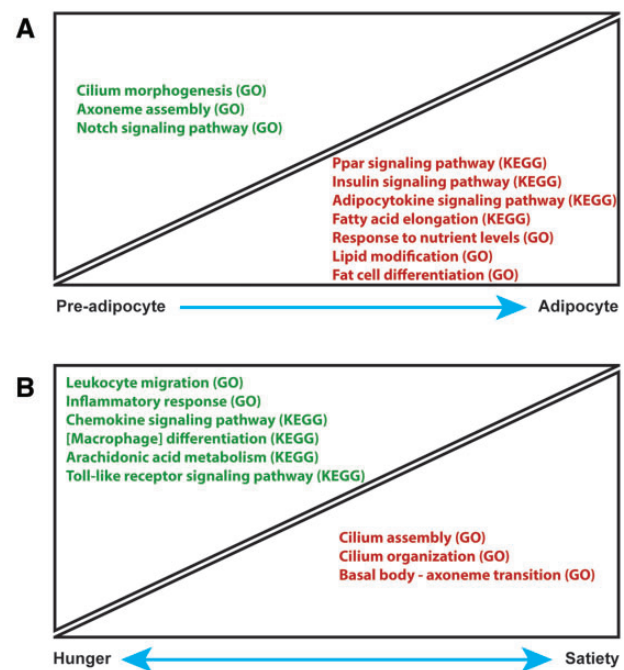


FIG. 5.—GO and KEGG functional enrichment of gonadal adipose depot and hypothalamic DE genes hint at differences in organ function between island and mainland mice. (A) Subset of enriched GO and KEGG terms among gonadal adipose depot DE genes. (B) Subset of enriched GO and KEGG terms among hypothalamic DE genes. In (A) and (B), the height of the triangles (shorter vertical axes) are arbitrary and indicate the qualitative contribution of the functions encapsulated by the displayed enriched GO and KEGG terms to the biological processes denoted along the long axes with blue arrows. Also, in both (A) and (B), green text designates significantly enriched terms among genes with higher transcription in island mice relative to mainland mice, whereas red text indicates enrichment among lower transcribed genes. In (B), the term Macrophage is used in place of Osteoclast as osteoclasts are bone-specific macrophages. Macrophages specific to the central nervous system (i.e., hypothalamus) are called microglia and although tissue-resident macrophages carry out tissue-specific functions and manifest distinctive transcriptional profiles they also have a subset of functions and gene expression profiles in common (Nataf et al. 2005; Gautier et al. 2012; Casano and Peri 2015).

and supplementary tables 1 and 2, Supplementary Material online). Higher expressed genes are also enriched for KEGG pathways *Arachidonic acid metabolism* and *Toll-like receptor (Tlr) signaling pathway* (fig. 5 and supplementary table 2, Supplementary Material online). Notably, cilia-related GO categories are enriched among genes transcribed at lower levels in island mice (fig. 5). Hypothalamic cilia might aid in sensing and regulating fat stores (Kang et al. 2015; Oh et al. 2015). Of particular interest is the lower expression of Bardet–Biedl syndrome genes, *Bbs2/9/10/12*, which are annotated to cilia-related terms (fig. 5 and supplementary table 4, Supplementary Material online). Disruption of these genes in mouse models of human diseases, such as Bardet–Biedl

syndrome, results in obesity (Oh et al. 2015). Relatedly, genes contributing to a particular ciliary protein complex (GO term *TCTN-B9D complex*), which functions in ciliary assembly and resides at the transition of the ciliary basal body and axonemal complex (Garcia-Gonzalo et al. 2011) (referred to as the “Basal body—axoneme transition” in fig. 5), are transcribed at a lower level in island mice (supplementary table 1, [Supplementary Material](#) online; supplementary table 3 available at <https://doi.org/10.5061/dryad.cjsxksn34>).

Discussion

Our study examined genome-wide patterns of gene expression in three key metabolic organs—the adipose, hypothalamus, and liver—to accomplish two goals. First, we sought to identify candidate genes for the evolution of gigantism in Gough Island mice. Second, we aimed to infer changes in cellular and metabolic functions that accompanied the evolution of Gough Island gigantism. Below, we discuss our progress toward these goals.

Candidate Genes

Numerous studies employ comparative transcriptomics (MacManes and Lacey 2012; Vonk et al. 2013; Gallant et al. 2014) and/or sequence variation in characterized gene regulatory elements (Infante et al. 2015; Kvon et al. 2016) to nominate regulatory changes involved in phenotypic evolution. Additionally, others have shown that complex genetic architectures underlie body size evolution via QTL mapping (Vaughn et al. 1999; Atchley et al. 2000; Cheverud et al. 2001). Our research revealed candidate genes for island body size evolution by combining comparative transcriptomics and sequence differentiation data within existing QTL (Gray et al. 2015). Some of the DE genes discovered within QTL for body weight and growth rate differences between Gough Island mice and mainland mice could be responsible for the evolutionary divergence of these traits. We further characterized DE genes within body size QTL by finding fixed nucleotide differences between wild-caught Gough Island mice and a mainland strain, some of which reside in putative regulatory elements. These mutations might have contributed to the expression differences we found. Among candidate genes within the island/mainland body size QTL, two genes stand out: *Arid5b* and *Tfec*. Binding-motif enrichment analysis among potentially coregulated groups of DE genes suggested that *Arid5b* and *Tfec* transcriptionally regulate hundreds of downstream genes in the liver. Our results therefore point to specific genes whose roles in promoting the evolution of island gigantism can be further explored.

Our consideration of candidate genes is accompanied by three caveats. First, much of the differential expression we documented probably affects functions unrelated to body size. Second, some (though not all) differential expression

could reflect changes in the composition or proportion of cell types in the evaluated organs. Third, our search for common transcriptional regulators of liver-specific coregulated groups (such as *Arid5b* and *Tfec*) was restricted to 5 kb of 5′-sequence space. Although these sequences likely include promoter and proximal enhancer elements (Xie et al. 2005; Visel et al. 2009), they do not contain all regulatory inputs.

Differential Gene Expression and Organ Function

The data shared in this report are valuable for generating hypotheses about the role of specific genes and organ functions that evolved to confer Gough Island mice with their giant body size. Below, we share how patterns of differential expression suggest potential mechanisms underlying an instance of the island rule on Gough Island that could be explored in future work.

Maturational Differences

Temporal shifts in developmental processes, or heterochronic shifts, are commonly invoked as mechanisms for evolutionary change (Gould 1985). Results from GO term and KEGG pathway enrichment analyses suggest that the maturational state of both the liver and gonadal adipose differs between island and mainland mice; these differences could impact the development of body size.

Livers in island mouse embryos express genes related to organ growth and development at lower levels, while expressing higher levels of genes related to postnatally maturing livers, including mitochondrial, metabolic, and immune functions (Hay 1991; Kimura 1991; Liechty and Lemons 1991; Cuezva et al. 1997; Thomson and Knolle 2010). Evidence of greater maturation is seen in 2-week livers as well. Attenuation of amino acid degradation is an established biochemical signal of neonatal and early postnatal livers in murid rodents (Conde and Scornik 1977; Blommaert et al. 1995; Gunewardena et al. 2015). At 2 weeks, amino acid metabolism is dissimilar between island and mainland mice, as suggested by the lower expression of lysosomal genes responsible for protein degradation and of genes related to the catabolism and biosynthesis of specific amino acids ([supplementary fig. 11 and tables 1–4, Supplementary Material](#) online). Concordantly, at this time point, there is significant elevation of ribosomal gene transcription in island mice, indicating that proteins are being constructed rather than dismantled. These findings suggest that the liver matures earlier in Gough Island mice and this change may promote metabolic energy for growth and protein synthesis.

Three observations indicate that gonadal adipocytes from island and mainland mice are at different stages of differentiation (and/or proliferation) in the gonadal adipose depot at 4 weeks. First, island mice exhibit lower expression of genes contributing to signaling pathways and metabolic processes characteristic of mature, lipid-storing adipocytes (fig. 5 and

supplementary tables 1, 2, and 4, [Supplementary Material](#) online; supplementary table 3 available at <https://doi.org/10.5061/dryad.cjsxksn34>). Second, island mice show higher expression of genes related to ciliogenesis. During adipogenesis, the presence of a transient primary cilium is an indication of the preadipocyte stage of differentiation, when cilium-modulated signaling pathways suppress differentiation (Marion et al. 2009; Forcioli-Conti et al. 2015; Oh et al. 2015). Third, the Notch signaling pathway was the only enriched signaling pathway among higher expressed genes (fig. 5 and [supplementary table 1, Supplementary Material](#) online). In vitro experiments indicate that adipocyte differentiation is sensitive to the magnitude of Notch signaling (Garcés et al. 1997; Nichols et al. 2004). At 4 weeks, we may have captured part of a population of new, yet young (in terms of differentiation status), preadipocytes. Notably, adipose depot enlargement via proliferation (hyperplasia), rather than expansion (hypertrophy), is correlated with the capacity to attenuate obesogenic conditions (Sun et al. 2011).

Potential Hypothalamic Contributions to Size Evolution

The significant enrichment of signals related to immunoreactivity and inflammation in island mouse hypothalami is akin to observations from mouse models of diet-induced obesity (Thaler et al. 2012; Gao et al. 2014; Kälin et al. 2015). Hypothalamic immunoreactivity is a hallmark of diet-induced obesity and long-term inflammation in the hypothalamus disrupts energy balance by disproportionately reducing satiety (anorexigenic)-promoting POMC neurons (Thaler et al. 2012).

The enhanced expression of immunity-related genes in island mouse hypothalami suggests two ways this organ could have evolved to alter energy balance and increase body size. First, hypothalamic regulation of feeding (Turek et al. 2005; Bechtold and Loudon 2013) may differ between island and mainland mice. Hypothalamic neuronal signaling can promote feeding and in turn nutrient uptake stimulates hypothalamic immunoreactivity (Kälin et al. 2015), raising the possibility that Gough Island mice sampled for RNA collection fed more recently than their mainland counterparts. Second, the hypothalamic immunoreactive environment might be enhanced in island mice, either because there are proportionately more immunoreactive cells, such as microglia, or those cells are more sensitive to circulating nutrient-derived compounds. The latter hypothesis is further supported by enrichment of the KEGG pathway *Toll-like receptor (Tlr) signaling pathway* among higher expressed genes ([supplementary table 2, Supplementary Material](#) online). Tlrs are hypothesized to be triggered by circulating saturated fatty acids (Könner and Brüning 2011) and Tlr signaling in microglia stimulates the transcription of proinflammatory genes (de Git and Adan 2015).

Two additional hints that hypothalamic control of food intake might be different between island and mainland mice at 4 weeks come from considering the function of primary cilia projecting from hypothalamic neurons and the way Arachidonic acid derivatives affect hunger (orexigenic)-promoting Npy/AgRP neurons. Leptin, an anorexigenic hormone secreted by adipose and other tissues, may instigate its signaling cascade through neuronal primary cilia in the hypothalamus (Kang et al. 2015; Oh et al. 2015). The reduced transcription of primary cilia-related genes in island mouse hypothalami indicates that leptin responsiveness may be suppressed; this would reduce suppression of hunger-promoting neurons, possibly increasing food intake (Lopaschuk et al. 2010) (fig. 5 and [supplementary table 1, Supplementary Material](#) online). Relatedly, neurotransmitters synthesized from the cell membrane polyunsaturated fatty acid Arachidonic acid (enriched among higher expressed genes) act on hunger-promoting neurons to increase energy intake (fig. 5) (Lopaschuk et al. 2010).

Island-Mouse-Specific Metabolic Environment in Postweaning Livers

Four-week livers from island mice expressed higher levels of genes related to the cell cycle and lower levels of mitochondria-related genes. At or near the time of weaning, a subset of hepatocytes undergo extensive proliferation and growth via endoreduplication, resulting in polyploid hepatocytes that are larger than their neighboring 2N hepatocytes (Klochender et al. 2012; Pandit et al. 2012; Miettinen et al. 2014). Differently sized hepatocytes manifest mitochondria-related differential gene expression, with larger polyploid hepatocytes showing about a 20% reduction in transcript levels (Miettinen et al. 2014). This reduction is similar in magnitude to that of lower expressed genes annotated to GO terms related to oxidative phosphorylation and mitochondria in our study (fig. 4). Overall, these patterns could indicate distinct sizes for island mouse and mainland mouse hepatocytes or a greater population of large polyploid hepatocytes in island mouse livers, either of which could alter the storage and mobilization of metabolites for growth in island mice.

Increased expression of genes promoting immunoreactive processes was also observed in the 4-week liver (fig. 4, see GO terms in GO Slim categories *Developmental processes*, *Signal transduction*, *Stress response*, and *Transport*). This observation is noteworthy because interactions between the immune system and hepatic cellular environments are key indicators of the progression of metabolic diseases, including obesity (Kim et al. 2004; Schmid et al. 2004; Mighiu et al. 2012; Lackey and Olefsky 2016). In a study that compared hepatic gene expression between lean and obese pigs across time points, the expression of genes regulating immunity was out of sync between the two breeds, suggesting a potential role for

immune system pathways in promoting body size differences (Ponsuksili et al. 2007).

Conclusion

Overall, our results show that extensive gene regulatory evolution in metabolic organs accompanied the evolution of gigantism in Gough Island mice. The intersection of differential expression with QTL mapping, sequence variants, and putative regulatory sequences points to promising candidate genes for body size evolution. Differences in metabolic processes, developmental status, and immunoreactivity in the adipose, hypothalamus, and liver likely promote and/or maintain the extreme body size of Gough Island mice. Our work sets the stage for functional evaluation of the candidate genes, networks, and physiological processes we identified, using the powerful tools available for house mice. Application of comparative transcriptomics to additional island mouse populations with extreme body sizes holds the potential to reveal the generality of gene regulatory mechanisms that underlie the island rule.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

B.A.P. was funded by National Institute of General Medical Sciences grant GM100426. C.N.D. was funded by National Human Genome Research Institute grant HG005232. M.J.N. was supported by a National Human Genome Research Institute training grant to the UW-Madison Genomic Sciences Training Program (5T32HG002760). C.N.D. was also supported in part by the Clinical and Translational Science Award program through the NIH National Center for Advancing Translational Sciences (grant UL1TR002373). We thank the UW-Madison Biotechnology Center for generating the RNA-sequencing libraries. We are grateful to the administrators and staff at the Charman Instructional Facility who aided in caring for the mice in this study. M.J.N. thanks Michelle Parmenter for guidance in working with Gough Island mice and for study-related discussions, Jered Stratton for useful comments on early versions of the manuscript, and Matt Nolte for assistance with figure design.

Literature Cited

- Atchley WR, Wei R, Crenshaw P. 2000. Cellular consequences in the brain and liver of age-specific selection for rate of development in mice. *Genetics* 155(3):1347–1357.
- Baba A, et al. 2011. PKA-dependent regulation of the histone demethylase complex Phf2–Arid5b. *Nat Cell Biol.* 13(6):668–675.
- Bechtold DA, Loudon ASI. 2013. Hypothalamic clocks and rhythms in feeding behavior. *Trends Neurosci.* 36(2):74–82.
- Bendlová B, et al. 2017. ZBTB16 gene variability influences obesity-related parameters and serum lipid levels in Czech adults. *Physiol Res.* 66(Suppl 3):S425–S431.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57:289–300.
- Bjørndal B, Burri L, Staalesen V, Skorve J, Berge RK. 2011. Different adipose depots: their role in the development of metabolic syndrome and mitochondrial response to hypolipidemic agents. *J Obes.* 2011:1–15.
- Blommaert EF, Luiken JJ, Blommaert PJ, van Woerkom GM, Meijer AJ. 1995. Phosphorylation of ribosomal protein S6 is inhibitory for autophagy in isolated rat hepatocytes. *J Biol Chem.* 270(5):2320–2326.
- Buerger F, et al. 2017. Depletion of Jmjd1c impairs adipogenesis in murine 3T3-L1 cells. *Biochim Biophys Acta Mol Basis Dis.* 1863(7):1709–1717.
- Casano AM, Peri F. 2015. Microglia: multitasking specialists of the brain. *Dev Cell* 32(4):469–477.
- Castro JPL, et al. 2019. An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife* 8:e42014.
- Chen Z, Torrens JJ, Anand A, Spiegelman BM, Friedman JM. 2005. Krox20 stimulates adipogenesis via C/EBPbeta-dependent and -independent mechanisms. *Cell Metab.* 1(2):93–106.
- Cheverud JM. 2005. Genetics of growth in the mouse. In: Eisen E, editor. *The mouse in animal genetics and breeding research.* Hackensack (NJ): Imperial College Press. p. 113–130.
- Cheverud JM, et al. 2001. Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mamm Genome* 12(1):3–12.
- Clausnitzer M, et al. 2015. *FTO* obesity variant circuitry and adipocyte browning in humans. *N Engl J Med.* 373(10):895–907.
- Conde RD, Scornik OA. 1977. Faster synthesis and slower degradation of liver protein during developmental growth. *Biochem J.* 166(1):115–121.
- Cuezva JM, et al. 1997. Mitochondrial biogenesis in the liver during development and oncogenesis. *J Bioenerg Biomembr.* 29(4):365–377.
- Dallongeville J, et al. 2008. The APOA5 Trp19 allele is associated with metabolic syndrome via its association with plasma triglycerides. *BMC Med Genet.* 9:84.
- de Git KCG, Adan RAH. 2015. Leptin resistance in diet-induced obesity: the role of hypothalamic inflammation. *Obes Rev.* 16(3):207–224.
- Deng JM, et al. 2011. Generation of viable male and female mice from two fathers. *Biol Reprod.* 84(3):613–618.
- Forcioli-Conti N, Lacas-Gervais S, Dani C, Peraldi P. 2015. The primary cilium undergoes dynamic size modifications during adipocyte differentiation of human adipose stem cells. *Biochem Biophys Res Commun.* 458(1):117–122.
- Foster JB. 1964. Evolution of mammals on islands. *Nature* 202(4929):234–235.
- Gallant JR, et al. 2014. Nonhuman genetics. Genomic basis for the convergent evolution of electric organs. *Science* 344(6191):1522–1525.
- Gao Y, et al. 2014. Hormones and diet, but not body weight, control hypothalamic microglial activity. *Glia* 62(1):17–25.
- Garcés C, et al. 1997. Notch-1 controls the expression of fatty acid-activated transcription factors and is required for adipogenesis. *J Biol Chem.* 272(47):29729–29734.
- Garcia-Gonzalo FR, et al. 2011. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet.* 43(8):776–784.
- Gautier EL, et al. 2012. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol.* 13(11):1118–1128.
- Gould SJ. 1985. *Ontology and phylogeny.* Cambridge: Belknap Press: An Imprint of Harvard University Press.
- Gray MM, et al. 2014. Demographic history of a recent invasion of house mice on the isolated island of Gough. *Mol Ecol.* 23(8):1923–1939.

- Gray MM, et al. 2015. Genetics of rapid and extreme size evolution in island mice. *Genetics* 201(1):213–228.
- Gunewardena SS, et al. 2015. Deciphering the developmental dynamics of the mouse liver transcriptome. *PLoS One* 10(10):e0141220.
- Han J, et al. 2011. The spatiotemporal development of adipose tissue. *Development* 138(22):5027–5037.
- Hart SN, Cui Y, Klaassen CD, Zhong X-B. 2009. Three patterns of cytochrome P450 gene expression during liver maturation in mice. *Drug Metab Dispos.* 37(1):116–121.
- Hay WW. 1991. Glucose metabolism in the fetal-placental unit. In: *Principles of perinatal-neonatal metabolism*. New York: Springer. p. 250–275.
- Hofmann SM, et al. 2007. Adipocyte LDL receptor-related protein-1 expression modulates postprandial lipid transport and glucose homeostasis in mice. *J Clin Invest.* 117(11):3271–3282.
- Infante CR, et al. 2015. Shared enhancer activity in the limbs and phallus and functional divergence of a limb-genital cis-regulatory element in snakes. *Dev Cell* 35(1):107–119.
- Jones AG, Chown SL, Gaston KJ. 2003. Introduced house mice as a conservation concern on Gough Island. *Biodivers Conserv.* 12(10):2107–2119.
- Kälin S, et al. 2015. Hypothalamic innate immune reaction in obesity. *Nat Rev Endocrinol.* 11(6):339–351.
- Kang GM, et al. 2015. Leptin elongates hypothalamic neuronal cilia via transcriptional regulation and actin destabilization. *J Biol Chem.* 290(29):18146–18155.
- Karolchik D, et al. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32(90001):D493–D496.
- Keane TM, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364):289–294.
- Kim S, et al. 2004. Hepatic gene expression profiles in a long-term high-fat diet-induced obesity mouse model. *Gene* 340(1):99–109.
- Kimura RE. 1991. *Lipid metabolism in the fetal-placental unit*. New York: Springer.
- Klaassen CD, Aleksunes LM. 2010. Xenobiotic, bile acid, and cholesterol transporters: function and regulation. *Pharmacol Rev.* 62(1):1–96.
- Klochendler A, et al. 2012. A transgenic mouse marking live replicating cells reveals in vivo transcriptional program of proliferation. *Dev Cell* 23(4):681–690.
- Könner AC, Brüning JC. 2011. Toll-like receptors: linking inflammation to metabolism. *Trends Endocrinol Metab.* 22(1):16–23.
- Kvon EZ, et al. 2016. Progressive loss of function in a limb enhancer during snake evolution. *Cell* 167(3):633–642.e11.
- Lackey DE, Olefsky JM. 2016. Regulation of metabolism by the innate immune system. *Nat Rev Endocrinol.* 12(1):15–28.
- Lagathu C, et al. 2010. Secreted frizzled-related protein 1 regulates adipose tissue expansion and is dysregulated in severe obesity. *Int J Obes.* 34(12):1695–1705.
- Lam TKT, Schwartz GJ, Rossetti L. 2005. Hypothalamic sensing of fatty acids. *Nat Neurosci.* 8(5):579–584.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):R25.
- Lawrence M, et al. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 9(8):e1003118.
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987–2993.
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* 12:323.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Liechty EA, Lemons JA. 1991. Protein metabolism in the fetal-placental unit. In: *Principles of perinatal-neonatal metabolism*. New York: Springer. p. 276–290.
- Locke AE, et al. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518(7538):197–206.
- Lomolino MV. 2005. Body size evolution in insular vertebrates: generality of the island rule. *J Biogeogr.* 32(10):1683–1699.
- Lopaschuk GD, Ussher JR, Jaswal JS. 2010. Targeting intermediary metabolism in the hypothalamus as a mechanism to regulate appetite. *Pharmacol Rev.* 62(2):237–264.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):550.
- MacManes MD, Lacey EA. 2012. The social brain: transcriptome assembly and characterization of the hippocampus from a social subterranean rodent, the colonial tuco-tuco (*Ctenomys sociabilis*). *PLoS One* 7(9):e45524.
- Marion V, et al. 2009. Transient ciliogenesis involving Bardet–Biedl syndrome proteins is a fundamental characteristic of adipogenic differentiation. *Proc Natl Acad Sci U S A.* 106(6):1820–1825.
- Mehrabian M, Wen PZ, Fislser J, Davis RC, Lusis AJ. 1998. Genetic loci controlling body fat, lipoprotein metabolism, and insulin levels in a multifactorial mouse model. *J Clin Invest.* 101(11):2485–2496.
- Meiri S, Cooper N, Purvis A. 2008. The island rule: made to be broken? *Proc R Soc B Biol Sci.* 275(1631):141–148.
- Miettinen TP, et al. 2014. Identification of transcriptional and metabolic programs related to mammalian cell size. *Curr Biol.* 24(6):598–608.
- Mighiu PI, Filippi BM, Lam TKT. 2012. Linking inflammation to the brain-liver axis. *Diabetes* 61(6):1350–1352.
- Motenko H, et al. 2015. MouseMine: a new data warehouse for MGI. *Mamm Genome* 26:325–330.
- Nataf S, et al. 2005. Brain and bone damage in KARAP/DAP12 loss-of-function mice correlate with alterations in microglia and osteoclast lineages. *Am J Pathol.* 166(1):275–286.
- Nichols AM, et al. 2004. Notch pathway is dispensable for adipocyte specification. *Genesis* 40(1):40–44.
- Nishizuka M, Honda K, Tsuchiya T, Nishihara T, Imagawa M. 2001. RGS2 promotes adipocyte differentiation in the presence of ligand for peroxisome proliferator-activated receptor γ . *J Biol Chem.* 276(32):29625–29627.
- Nord AS, et al. 2013. Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 155(7):1521–1531.
- Oh EC, Vasanth S, Katsanis N. 2015. Metabolic regulation and energy homeostasis through the primary cilium. *Cell Metab.* 21(1):21–31.
- Ohlsson C, et al. 2009. The role of liver-derived insulin-like growth factor-I. *Endocr Rev.* 30(5):494–535.
- Oishi Y, et al. 2005. Krüppel-like transcription factor KLF5 is a key regulator of adipocyte differentiation. *Cell Metab.* 1(1):27–39.
- Oliver F, et al. 2005. Regulatory variation at glypican-3 underlies a major growth QTL in mice. *PLoS Biol.* 3(5):e135.
- Panasyuk G, et al. 2012. PPAR γ contributes to PKM2 and HK2 expression in fatty liver. *Nat Commun.* 3:672.
- Pandit SK, et al. 2012. E2F8 is essential for polyploidization in mammalian cells. *Nat Cell Biol.* 14(11):1181–1191.
- Parmenter MD, et al. 2016. Genetics of skeletal evolution in unusually large mice from Gough Island. *Genetics* 204(4):1559–1572.
- Plaisier CL, et al. 2012. Zbtb16 has a role in brown adipocyte bioenergetics. *Nutr Diabetes* 2(9):e46.
- Ponsuksili S, Murani E, Walz C, Schwerin M, Wimmers K. 2007. Pre- and postnatal hepatic gene expression profiles of two pig breeds differing in body composition: insight into pathways of metabolic regulation. *Physiol Genomics.* 29(3):267–279.
- Qiu S, et al. 2017. Hepatic estrogen receptor α is critical for regulation of gluconeogenesis and lipid metabolism in males. *Sci Rep.* 7:1661.
- Raia P, Meiri S. 2011. The tempo and mode of evolution: body sizes of island mammals. *Evolution* 65(7):1927–1934.
- Rowe-Rowe DT, Crafford JE. 1992. Density, body size, and reproduction of feral house mice on Gough Island. *S Afr J Zool.* 27(1):1–5.

- Ruderman NB, et al. 2003. AMPK as a metabolic switch in rat muscle, liver and adipose tissue after exercise. *Acta Physiol Scand.* 178(4):435–442.
- Salih DAM, et al. 2004. Insulin-like growth factor-binding protein 5 (Igfbp5) compromises survival, growth, muscle development, and fertility in mice. *Proc Natl Acad Sci U S A.* 101(12):4314–4319.
- Schmid GM, et al. 2004. Effect of high-fat diet on the expression of proteins in muscle, adipose tissues, and liver of C57BL/6 mice. *Proteomics* 4(8):2270–2282.
- Septer S, et al. 2012. Yes-associated protein is involved in proliferation and differentiation during postnatal liver development. *Am J Physiol Gastrointest Liver Physiol.* 302(5):G493–G503.
- Shungin D, et al. 2015. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518(7538):187–196.
- Siepel A, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15(8):1034–1050.
- Smith CL, Goldsmith C-AW, Eppig JT. 2005. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol.* 6(1):R7.
- Speliotes EK, et al. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 42(11):937–948.
- Stojnic R, Diez D (2018). *PWMErrich: PWM enrichment analysis*. doi: 10.18129/B9.bioc.PWMErrich
- Sun K, Kusminski CM, Scherer PE. 2011. Adipose tissue remodeling and obesity. *J Clin Invest.* 121(6):2094–2101.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7):e21800.
- Thaler JP, et al. 2012. Obesity is associated with hypothalamic injury in rodents and humans. *J Clin Invest.* 122(1):153–162.
- Thirone ACP, Huang C, Klip A. 2006. Tissue-specific roles of IRS proteins in insulin signaling and glucose transport. *Trends Endocrinol Metab.* 17(2):72–78.
- Thomson AW, Knolle PA. 2010. Antigen-presenting cell function in the tolerogenic liver environment. *Nat Rev Immunol.* 10(11):753–766.
- Turek FW, et al. 2005. Obesity and metabolic syndrome in circadian clock mutant mice. *Science* 308(5724):1043–1045.
- Van Valen L. 1973. Pattern and the balance of nature. *Evol Theory* 1:31–49.
- Vaughn TT, et al. 1999. Mapping quantitative trait loci for murine growth: a closer look at genetic architecture. *Genet Res.* 74(3):313–322.
- Visel A, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457(7231):854–858.
- Vitali A, et al. 2012. The adipose organ of obesity-prone C57BL/6J mice is composed of mixed white and brown adipocytes. *J Lipid Res.* 53(4):619–629.
- Vonk FJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A.* 110(51):20651–20656.
- Wace NM. 1961. The vegetation of Gough Island. *Ecol Monogr.* 31(4):337–367.
- Walthall K, Cappon GD, Hurtt ME, Zoetis T. 2005. Postnatal development of the gastrointestinal system: a species comparison. *Birth Defect Res B.* 74(2):132–156.
- Wang QA, Tao C, Gupta RK, Scherer PE. 2013. Tracking adipogenesis during white adipose tissue development, expansion and regeneration. *Nat Med.* 19(10):1338–1344.
- Waxman DJ, Celenza JL. 2003. Sexual dimorphism of hepatic gene expression: novel biological role of KRAB zinc finger repressors revealed. *Genes Dev.* 17(21):2607–2623.
- Westmacott A, Burke ZD, Oliver G, Slack JMW, Tosh D. 2006. C/EBPalpha and C/EBPbeta are early markers of liver development. *Int J Dev Biol.* 50(7):653–657.
- Wuarin J, et al. 1992. The role of the transcriptional activator protein DBP in circadian liver gene expression. *J Cell Sci Suppl.* 16:123–127.
- Xie X, et al. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434(7031):338–345.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11(2):R14.
- Yue F, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515(7527):355–364.

Associate editor: Marta Barluenga