

RESEARCH ARTICLE

Open Access



Whole genome sequence analysis of the TALLYHO/Jng mouse

James Denvir, Goran Boskovic[^], Jun Fan, Donald A. Primerano, Jacaline K. Parkman and Jung Han Kim*

Abstract

Background: The TALLYHO/Jng (TH) mouse is a polygenic model for obesity and type 2 diabetes first described in the literature in 2001. The origin of the TH strain is an outbred colony of the Theiler Original strain and mice derived from this source were selectively bred for male hyperglycemia establishing an inbred strain at The Jackson Laboratory. TH mice manifest many of the disease phenotypes observed in human obesity and type 2 diabetes.

Results: We sequenced the whole genome of TH mice maintained at Marshall University to a depth of approximately 64.8X coverage using data from three next generation sequencing runs. Genome-wide, we found approximately 4.31 million homozygous single nucleotide polymorphisms (SNPs) and 1.10 million homozygous small insertions and deletions (indels) of which 98,899 SNPs and 163,720 indels were unique to the TH strain compared to 28 previously sequenced inbred mouse strains. In order to identify potentially clinically-relevant genes, we intersected our list of SNP and indel variants with human orthologous genes in which variants were associated in GWAS studies with obesity, diabetes, and metabolic syndrome, and with genes previously shown to confer a monogenic obesity phenotype in humans, and found several candidate variants that could be functionally tested using TH mice. Further, we filtered our list of variants to those occurring in an obesity quantitative trait locus, *tabw2*, identified in TH mice and found a missense polymorphism in the *Cidec* gene and characterized this variant's effect on protein function.

Conclusions: We generated a complete catalog of variants in TH mice using the data from whole genome sequencing. Our findings will facilitate the identification of causal variants that underlie metabolic diseases in TH mice and will enable identification of candidate susceptibility genes for complex human obesity and type 2 diabetes.

Keywords: Whole genome sequencing, Obesity, Type 2 diabetes, Mouse model, TALLYHO, Quantitative Trait Loci

Background

The high prevalence of obesity and type 2 diabetes is a serious public health issue that is associated with devastating health consequences such as cardiovascular diseases [1, 2]. The World Health Organization estimated that more than 10 % of the world's adult population (200 million men and 300 million women) was obese (body mass index ≥ 30 kg/m²) in 2008 (<http://www.who.int/mediacentre/factsheets/fs311/en/>) and that, in 2012, 347 million people had diabetes, 90 % of whom had type 2 diabetes (<http://www.who.int/mediacentre/factsheet/fs312/en/>). Genetic predisposition is recognized as a major risk factor for the development of obesity and type 2 diabetes; estimates of the heritability range from 50 % to 60 % for

both diseases [3]. Therefore, identification of underlying susceptibility genes would define potential targets and pathways for intervention and treatment of obesity and type 2 diabetes.

The genetics of human obesity and type 2 diabetes is complex [4, 5]. It involves multiple susceptibility genes and their interactions with environmental factors [6]. Animal models that share both physiologic and genetic similarity with humans are used in obesity and type 2 diabetes research to minimize the confounding effects of heritability, genetic heterogeneity and environment that determine these diseases in humans [7, 8].

The TALLYHO/Jng (TH) mouse is a polygenic inbred model for human obesity and type 2 diabetes [9]. TH mice manifest many of the disease phenotypes observed in human obesity and type 2 diabetes including hyperleptinemia, hyperinsulinemia, insulin resistance, glucose intolerance, hyperlipidemia, and hyperglycemia [10].

* Correspondence: kimj@marshall.edu

[^]Deceased

Department of Biomedical Sciences, Joan C. Edwards School of Medicine, Marshall University, 1700 3rd Ave. #435K BBSC, Huntington, WV 25755, USA



These mice also exhibit increased islet insulin secretion in response to glucose and β -cell mass [11]. The genetic basis for obesity and type 2 diabetes in TH mice has been studied using outcross experiments with normal strains which has led to the identification of multiple quantitative trait loci (QTLs) linked to adiposity and hyperglycemia [10].

In addition to genetic analysis, the TH mouse has been used in the development of therapeutic agents for obesity and type 2 diabetes [12, 13] and has also served as a model system for many diabetes and obesity related abnormalities, including decreased exercise capacity [14], impaired wound healing [15–17], periodontitis [18], tissue susceptibility to hypoxia [19, 20], bone loss [21, 22], circadian disruption [23], and vasculature abnormalities [24–26].

In this study, we sequenced the whole genome of the TH mouse using next-generation sequencing with the goal of identifying genome-wide sets of single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels), a subset of which are present exclusively in the TH mouse (private variants). A complete catalogue of TH genomic variants (public and private) will aid the identification of causative variants and genes underlying the diseases observed in TH mice, which will maximize the relevancy of this model for human obesity and type 2 diabetes. Knowing the distribution of variants across the genome of the TH mouse vs. other classic and wild-derived inbred strains [27] may identify naturally occurring gene variant sets that can represent model genes for the complex human obesity and type 2 diabetes. As a proof of principle, we integrated our genome variant data with genetic mapping data and identified a variant in cell death-inducing DFFA-like effector c (*Cidec*) whose effect on protein function we characterized.

Methods

Description and origin of the TH mouse strain

The TH mouse is originally derived from an outbred colony of the Theiler Original strain in which two male mice spontaneously became polyuric, glucosuric, hyperinsulinemic, and hyperglycemic (Harrow, United Kingdom) [28]. In 1994, male diabetic progeny and apparently normal female mice were imported into The Jackson Laboratory (TJL) (Bar Harbor, ME, USA) and an inbred strain was then established by selecting for male hyperglycemia by Dr. Jürgen Naggert's research group at TJL. In 2001, a sub-colony was initiated in our laboratory with breeding pairs from Dr. Naggert's research colony [10] and used in this study. After arrival at our laboratory, we interbred siblings for 18 generations followed by one generation backcrossing and an additional 14 generations of sibling interbreeding. Since

the generation prior to arrival at our laboratory remains unknown, we refer to the generation used for whole genome sequencing in this study as F? + F18N1F14. All animal studies were carried out with the approval of Marshall University Animal Care and Use Committee.

DNA sequencing, read alignment, and bioinformatics

High quality genomic DNA was extracted from the liver of a male TH mouse from our colony using a Qiagen (Valencia, CA) Genomic-tip 100/G kit. Genomic DNA (1 μ g) was sheared using a Covaris (Woburn, MA) S2 instrument to ~350 bp and used to construct a sequencing library using an Illumina (San Diego, CA) DNA Sample Preparation kit according to the manufacturer's recommended protocol. The library was quantified using ThermoFisher (Waltham, MA) Qubit fluorimetry and sized on an Agilent (Santa Clara, CA) Bioanalyzer DNA chip. The resulting whole genome library (8 pmole) was amplified on a flow cell using an Illumina cBot cluster station and then sequenced on an Illumina HiSeq 1000 in the Marshall University Genomics Core Facility. Three sequencing runs (one 2x50 bp paired-end and two 2x100 bp paired-end) were performed in order to obtain adequate depth of coverage.

Sequencing reads were aligned to the C57BL/6J (B6) reference genome (GRCm38, mm10) using Bowtie2 v2.1.0 [29]. Our variant calling pipeline was based on that developed by Wong et al. [30] to call variants in the FVB/NJ mouse strain. Briefly, duplicate reads were removed with SAMtools v0.1.18 [31] using the "samtools rmdup" command. To improve quality of the variant calling, local realignment around insertions and deletions was performed using GATK v3.2.2 [32] by first running the "RealignerTargetCreator" command and then the "IndelRealigner" command. SNPs and indels were called by generating a pileup using the "samtools pileup" command with options uEDS, and piping the results to the "bcftools view" command with options "-p 0.99 -vbcgN". Variants were filtered using the VCFtools package [33] version 0.1.12. In order to maintain maximal consistency with the Mouse Genome Project (MGP) [34], we used the same options as in Wong et al. [30].

We identified variants in the TH strain that did not occur in any of the 28 mouse strains published in the MGP (Additional file 1: Table S1), which we term "private" variants, following Keane et al. [34]. To do this, we generated a list of all genomic locations of TH variants in the form of a bed file, and then performed variant calling on the bam files for the 28 MGP mouse strains at those locations, using SAMTools' "-l" option along with the same parameters as were used to call TH variants. TH SNPs qualified as private if no MGP strain had the same SNP at the same location, where at least 21 of the

28 strains had a call quality at least 20 and read depth at least 5 at that location. TH indels qualified as private if no MGP strain had a variant at the same genomic location, with the same criteria for call quality and read depth being applied.

We then added functional consequence annotation, including “Sorting Intolerant From Tolerant” (SIFT) [35] prediction of protein function changes of coding variants, using a local installation of Ensembl Variant Effect Predictor (VEP) version 77 [36]. Output from VEP includes classification of each variant using a set of one or more Sequence Ontology (SO) terms [37]. Variants are classified multiple times if they lie within a region intersecting multiple known transcripts. We tabulated the number of variants according to their collection of SO terms, collecting into a single “Multiple Classification Sets” group all variants that had multiple, distinct sets of SO terms due to their presence in multiple transcripts. Since this latter group is large, and may include potentially interesting variants, we identified a set of ten SO terms we call “potentially pathogenic”: *frameshift_variant*, *inframe_deletion*, *inframe_insertion*, *missense_variant*, *stop_gained*, *stop_lost*, *initiator_codon_variant*, *splice_region_variant*, *splice_acceptor_variant*, and *splice_donor_variant*. We tabulated the number of variants in the “multiple classification sets” that classified with each of these SO terms for one or more transcripts. The same classification was performed on the private variants. For missense variants, we additionally used Protein Variation Effect Analyzer v 1.1 (PROVEAN) [38] to provide further prediction of the functional effects of protein changes.

In order to uniquely associate a representative SO term with each variant, we chose a single representative SO term for each distinct set of SO terms identified for a given variant in a given transcript, as shown in Additional file 2: Table S2 and Additional file 3: Table S3. For example, 32 SNPs were classified with the two SO terms “intron_variant” and “splice_region_variant”: for these SNPs we chose “splice_region_variant” as the representative term. Using this process, variants occurring in multiple transcripts that had multiple, distinct sets of SO terms associated with them, resulted in having multiple representative SO terms. For such variants, we chose the “most pathogenic” representative SO term, ordering them in the following priority: “stop_gained”, “stop_lost”, “frameshift_variant”, “missense_variant”, “inframe_insertion”, “inframe_deletion”, “splice_acceptor_variant”, “splice_donor_variant”, “splice_region_variant”, “initiator_codon_variant”, “synonymous_variant”, “5_prime_UTR_variant”, “3_prime_UTR_variant”, “mature_miRNA_variant”, “upstream_gene_variant”, “downstream_gene_variant”, “intron_variant”, “non_coding_transcript_variant”, “intergenic_variant”.

In order to compare our variant sets to known associations between genetic variants and human traits for which TH is a model, we first retrieved the Genome Wide Association Study (GWAS) catalog [39] from the European Bioinformatics Institute (EBI) (<https://www.ebi.ac.uk/gwas/>), and filtered the catalog by the column “DISEASE/TRAIT” for each of the terms “obesity”, “diabetes”, and “metabolic”. For each filtered catalog, we extracted all human Entrez GeneIDs listed in the columns MAPPED_GENE(S), UPSTREAM_GENE_ID, and DOWNSTREAM_GENE_ID. Using the Ensembl interface to BioMart [40], we converted these Entrez GeneIDs to Ensembl human gene IDs. Additionally, we retrieved a list of genes from Pigeyre et al. [41] in which mutations have been previously shown to cause a monogenic obesity phenotype in humans, and found Ensembl IDs for these genes using the same tool. For all these human Ensembl gene IDs, we found Ensembl mouse gene IDs of orthologous genes, again using the Ensembl interface to Biomart. We then filtered our variant sets (SNPs, indels, private SNPs, and private indels) to identify potentially pathogenic variants associated with each of these orthologous gene sets.

Aligned reads were uploaded to the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) and can be accessed via accession number SRP067703.

Cell culture, plasmids, transfections, and microscopy

COS-1 cells (ATCC, Manassas, VA) were grown in Dulbecco’s modified Eagle’s medium with L-glutamine, 4.5 g/L glucose and sodium pyruvate (Mediatech, Manassas, VA) supplemented with 10 % (v/v) bovine serum (Sigma, St. Louis, MO) and 100 units/ml penicillin and 100 mg/ml streptomycin (Mediatech, Manassas, VA).

The murine *Cidec* cDNA was procured from GenScript (Piscataway, NJ). A Green Fluorescent Protein fusion, pAcGFP1-CIDEC was created by cloning the *Cidec* cDNA with 5’ HindIII and 3’ BamHI sites upstream and in-frame with the AcGFP1 sequence in pAcGFP1-N1 (Clontech, Mountain View, CA). The entire *Cidec* coding region was included except for the stop codon [42]. Site-directed mutagenesis was used to introduce the missense polymorphism R46S (GeneScript, Piscataway, NJ). Plasmids, pAcGFP1-CIDEC (R46) and pAcGFP1-CIDEC (S46), were transfected into COS-1 cells cultured on 4-well chamber slides (1×10^5 cells per chamber) (Thermo Scientific, Waltham, MA) using Lipofectamine LTX&PLUS (Life Technologies, Grand Island, NY) following the manufacturer’s instruction [43]. COS cells do not express endogenous CIDEC [44]. After transfection, the cells were cultured for 48 h with and without 400 μ M BSA-complexed oleic acid (Sigma) [45]. Cells were then washed with PBS, fixed by 4 %

paraformaldehyde for 30 min, and permeabilized in 0.05 % (w/v) saponin (Sigma) in PBS for 20 min. Lipids were stained with 1 µg/ml Nile red (Sigma) which partitions with neutral lipids; cells were washed with PBS (6x). Nuclei were labeled by placing coverslips onto slides with Prolong® Gold Antifade Mountant with DAPI (Life Technologies). Coverslips were dried and cells viewed on a Leica SP5 confocal microscope located in the Marshall University Imaging Center. Each experiment was performed in quadruplicate samples. Association of full length *Cidec* with lipid droplets in COS-1 cells was assessed using ImageJ software by performing a co-localization analysis of GFP fluorescence with Nile red labeled lipid droplets in each z-section of a transfected cell [43]; these results were then averaged to calculate the overall co-localization in a given cell (the ratio of Nile red signal specifically associated with GFP signal to GFP signal). The data were from 12 cell clusters in each condition from two independent experiments. Quantitative data were presented as means ± SEM. Two-tailed Student's t-tests were performed on data. Differences of $P < 0.05$ were considered significant.

Results

Identification of SNPs and indels

The whole genome of the TH mouse was sequenced to an average depth of ~64.8X coverage using data from three Illumina paired-end read sequencing runs. The sequencing reads were mapped to the B6 mouse reference genome (GRCm38/mm10). Using a SAMTools-based pipeline [31], we identified 4,370,213 SNPs (4,310,548 of which were homozygous) and 1,213,617 indels (1,065,090 of which were homozygous) genome-wide, relative to the reference strain C57BL/6 (B6). The positional distribution of variants on each chromosome is shown in Fig. 1. To assess the reliability of our data, PCR amplification and Sanger sequencing were applied to 14 homozygous SNPs to determine whether they agreed with next generation sequencing results (Additional file 4: Table S4). We found that all 14 SNPs were consistent with the Illumina sequencing data.

We observed a small percentage of heterozygous SNPs and indels in the TH genome (1.37 % and 12.24 %, respectively). Possible explanations for the apparent presence of heterozygosity include errors in variant calling, true residual heterozygosity, and recent heterozygous mutations not yet fixed to homozygosity [46]. We did not consider heterozygous SNPs and indels in our subsequent analyses.

Functional consequences of SNPs and indels

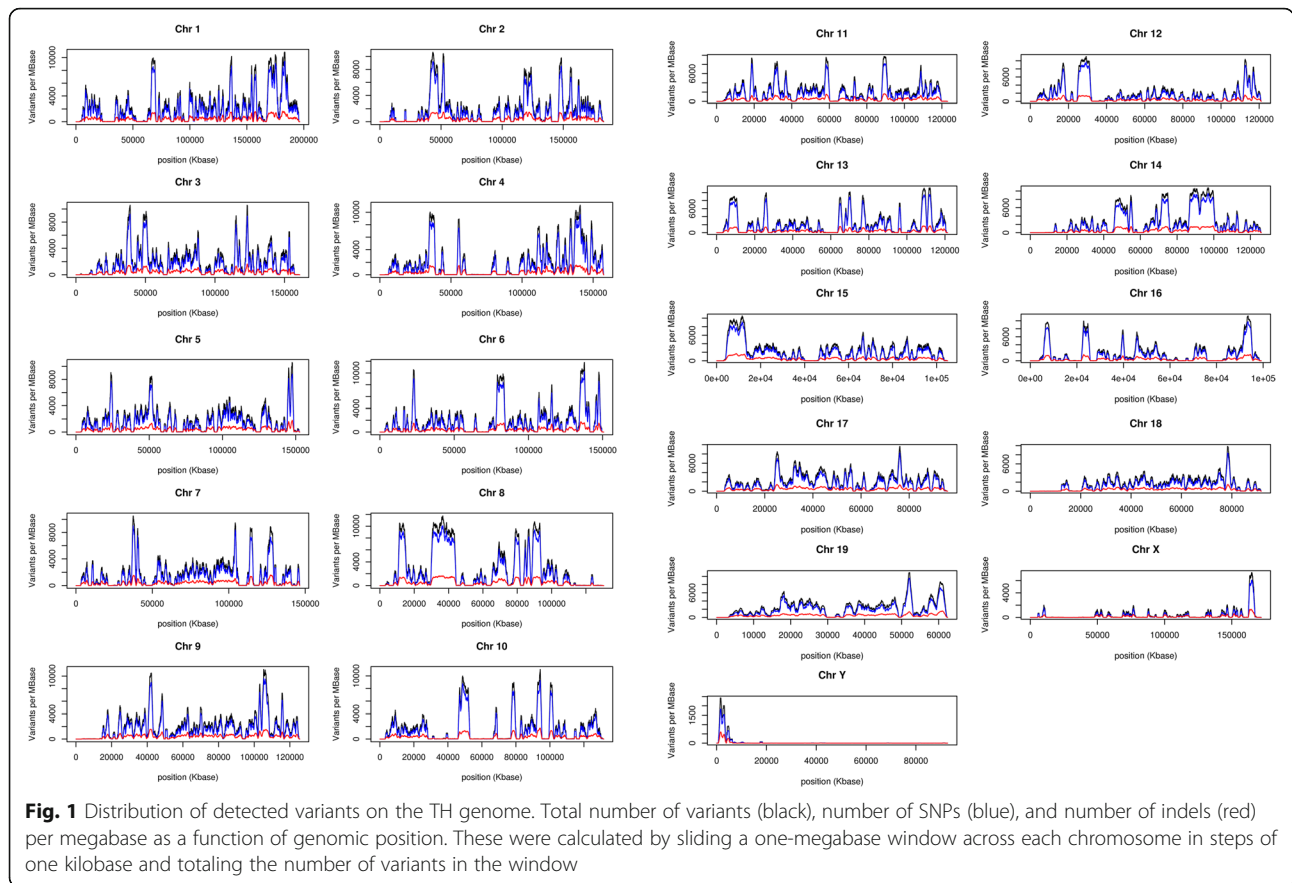
We assigned putative functional consequences to the set of homozygous variants (SNPs and indels) in TH using Variant Effect Predictor (VEP) [36], which classified the

variants using terms from the Sequence Ontology (SO) [37] (Additional file 2: Table S2 and Additional file 3: Table S3). Variants which intersected multiple transcripts, and for which the set of SO terms associated with the variant differs among those transcripts, were categorized as “multiple classification sets”. For each of ten SO terms we designated “potentially pathogenic” (see Methods), we counted the number of variants in the “multiple classification sets” category that classified under that SO term in one or more transcripts.

For each distinct collection of SO terms identified for a variant, we chose a representative term (as shown in Additional file 2: Table S2 and Additional file 3: Table S3). For variants with multiple classification sets this potentially resulted in multiple representative SO terms: for these variants with multiple classifications we categorized the variant as the “most pathogenic” according to the prioritization defined in the methods. This resulted in a unique SO term for each variant: the distributions of these SO terms among SNPs and indels are shown in Figs. 2 and 3, respectively. The majority of SNPs were either intergenic (50.2 %) or intronic (29.2 %). The next two largest SNP groups were those which were located 5 kb upstream (8.95 %) or downstream (7.38 %) from a coding gene. A small number of SNPs (less than 1 % of the total) were located within the protein coding regions of gene bodies. SNP variants resulted in 21,039 synonymous codon changes, 10,829 non-synonymous codon changes, 63 conversions of coding codons to stop codons and 34 conversions of stop codons to coding codons.

There were 10,829 SNPs causing non-synonymous changes to one or more protein-coding transcripts in 4,351 genes. We evaluated the effect of the amino acid substitutions resulting from the SNPs using SIFT [35, 47] (as implemented by VEP) and PROVEAN [38]. These algorithms classify the changes as either deleterious (SIFT score < 0.05 , or PROVEAN score < -2.5) or tolerated/neutral (SIFT/PROVEAN). SIFT classified 14.8 % of the non-synonymous substitutions (1601 substitutions) in 1,148 genes as deleterious. PROVEAN classified 9.6 % (1041 substitutions) in 772 genes as deleterious. 4.8 % of the missense variants (512 substitutions) in 444 genes were classified as deleterious by both algorithms.

As expected, the vast majority (99.83 %) of TH indels fell into noncoding groups (either intergenic, intronic, 5 kb upstream or downstream of coding genes, 5' or 3' UTR, or non-coding transcript variants). The total in-frame and frameshift indels were 362 and 203, respectively (Fig. 3). Proportions of SNPs and indels in each functional class are similar to those found in the FVB/NJ genome [30] (Table 1).



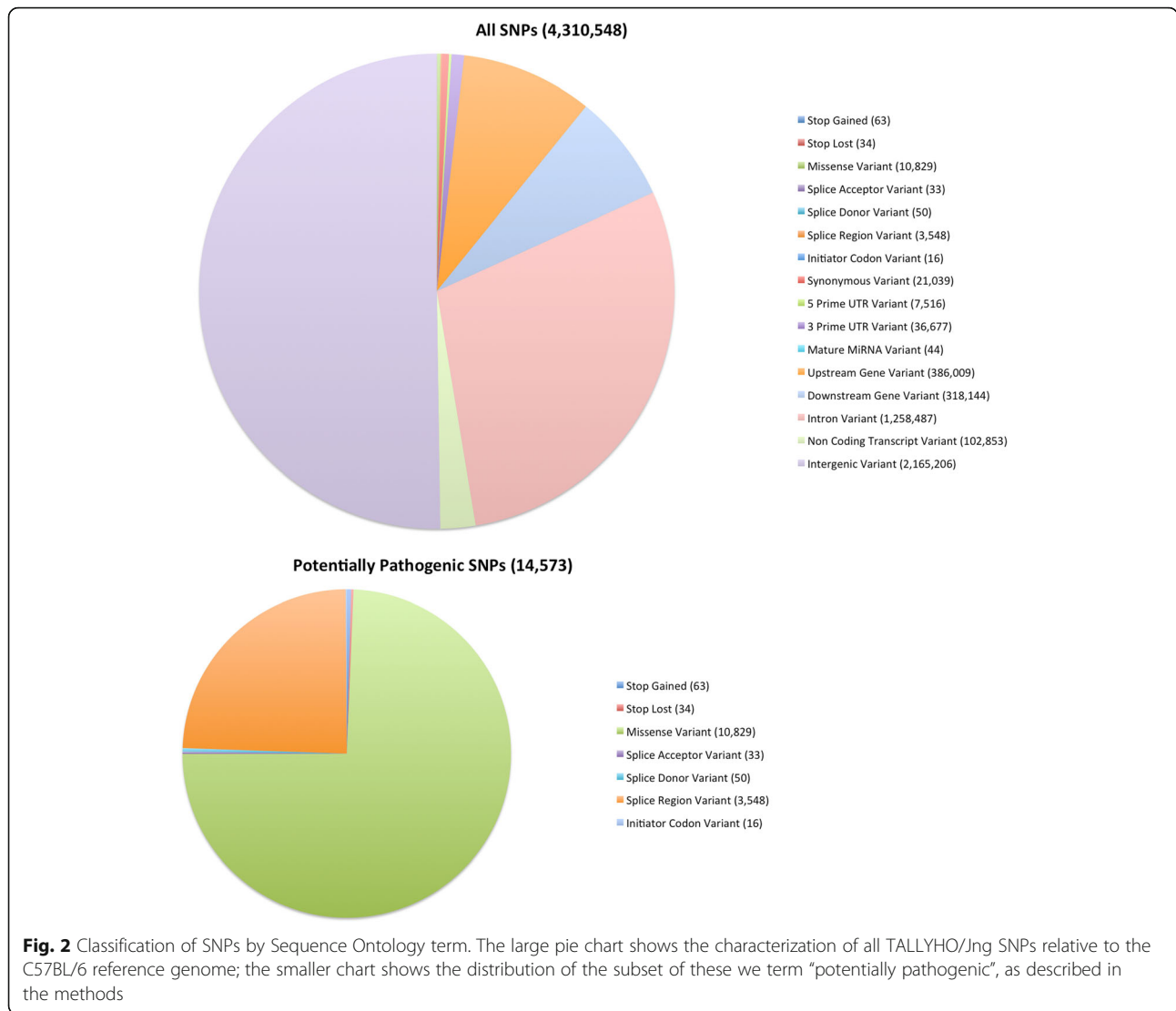
Comparison to the human genome-wide association study catalog and to known mendelian obesity genes

In order to interpret our data in the context of human disease, we collected human gene IDs in which variants were associated with obesity, diabetes, or metabolic syndrome, from the EBI GWAS catalog version 1.0.1 [39]. These gene IDs were mapped to orthologous mouse Ensembl gene IDs using the Ensembl interface to BioMart [40], generating sets of orthologous mouse genes. We additionally collected a list of known Mendelian obesity genes in humans from Pigeyre et al. [41], and generated a set of mouse orthologous genes as aforementioned. We then filtered the potentially pathogenic TH variants against these gene sets to provide lists of potentially pathogenic variants for each of these diseases. We found 26 genes with potentially pathogenic variants in the GWAS obesity gene set, 85 in the GWAS diabetes gene set, and 246 in the GWAS metabolic gene set. Additionally, we found 13 genes with potentially pathogenic variants orthologous to genes known to cause monogenic syndromic obesity, four orthologous to genes known to cause monogenic non-syndromic obesity, and three orthologous to genes known to cause monogenic non-syndromic lipodystrophy (Table 2, Additional file 5: Table S5).

TH private SNPs and indels

We identified variants that were present in the TH strain and absent from 28 MGP strains by passing the bam files for each strain through our variant-calling pipeline at each location where we identified a TH variant. A TH SNP was identified as “private” if it was not called as the same SNP in any of the other 28 strains, and a TH indel was identified as “private” if no variant was identified at the same location in any of the other 28 strains. Additionally, the designation of “private” required a call quality at least 20 and read depth at least 5 in at least 21 of the 28 strains. In this analysis, we identified a total of 98,899 private SNPs and 163,720 private indels, which represented 2.29 % and 15.4 % of the total SNPs and indels, respectively.

We classified private variants by SO term using VEP as described in the analysis for all variants above. Of the 98,899 private SNPs, 358 were missense substitutions, four resulted in the gain of a stop codon, and one in the loss of a stop codon (Additional file 6: Table S6 and Additional file 7: Table S7, Figs. 4 and 5). Among the 163,720 private indels, 42 were frameshift indels and 26 were in-frame indels (Additional file 6: Table S6 and Additional file 7: Table S7, Figs. 4 and 5). For these private SNPs and indels, along with those associated by



VEP with splice regions and initiator codons, we counted the number of such variants occurring in each gene. There were 961 private SNPs meeting one or more of these criteria in 372 genes (Additional file 8: Table S8) and 576 private indels meeting one or more of these criteria in 215 genes (Additional file 9: Table S9). For the private SNPs, we found that 91 of the 372 genes had one or more private SNPs determined by SIFT to be deleterious missense variants, 69 had one or more private SNPs determined by PROVEAN to be deleterious, with 44 genes having SNPs determined to be deleterious by both algorithms.

In the analysis linking our variant lists with human GWAS studies and human Mendelian obesity genes described above, we determined the number of variants in orthologous genes which were private to the TH mouse (Table 2).

Characterization of the *tabw2* obesity QTL interval on chromosome 6

Tabw2 (TALLYHO associated body weight 2) is a major obesity QTL identified in TH mice, and confirmed by a congenic strain on the B6 background [48]. Using sub-congenic analysis we have determined that the effect of *tabw2* on obesity could be attributed to two adjacent loci, *tabw2a* and *tabw2b* [49]. There are 411 and 73 protein coding genes cataloged in the Ensembl mouse annotation database in the *tabw2a* (6:80,217,217-125,356,646 in coordinates relative to GRCm38) and *tabw2b* (6:133,853,029-144,639,629) intervals, respectively. A total of 123,971 SNPs and 28,622 indels were found in the *tabw2a* and *tabw2b* intervals, of which 68,233 SNPs and 15,747 indels were in protein-coding genes (Table 3). Among those, only 5 SNPs (in 5 genes) in the *tabw2a* interval and 2 SNPs (in 2 genes) in the *tabw2b* interval

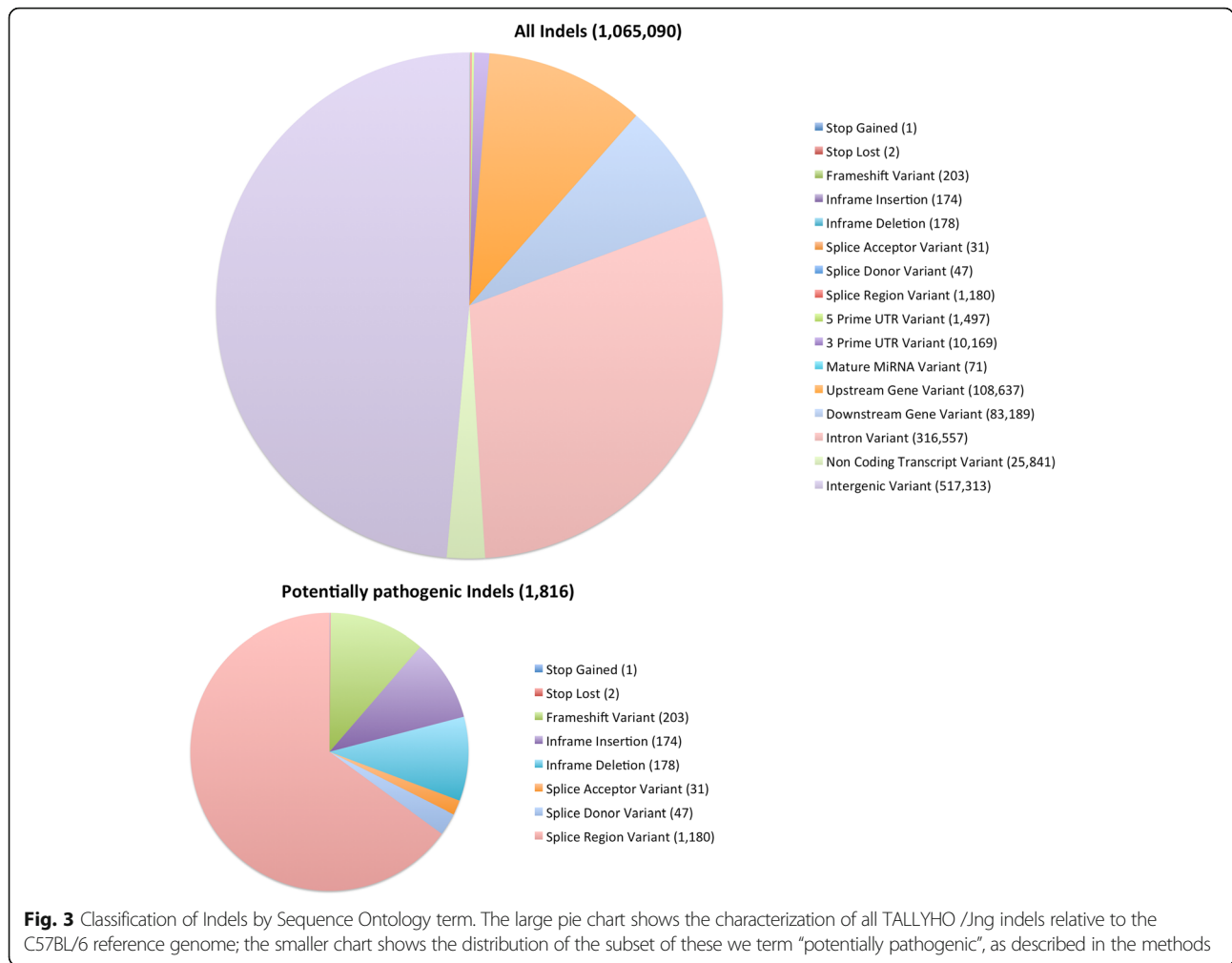


Table 1 Comparison of distribution of variants in TALLYHO/Jng to that in FVB/NJ

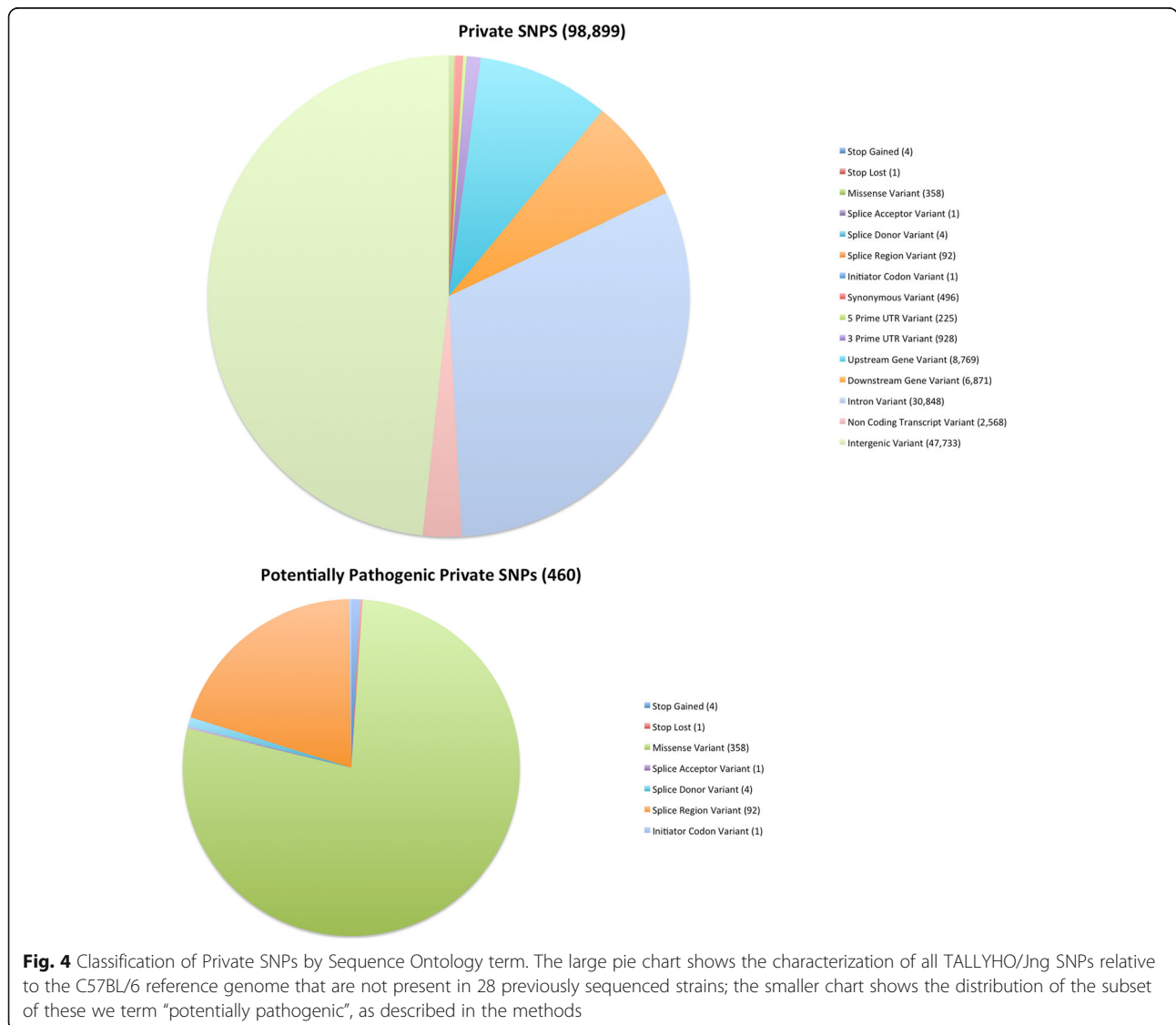
	FVB/NJ		TALLYHO/Jng	
	Percentage of all SNPs	Percentage of all indels	Percentage of all SNPs	Percentage of all indels
5 kb upstream or downstream	7.731 %	9.895 %	7.496 %	8.167 %
5' or 3' UTR	0.434 %	0.433 %	0.291 %	0.298 %
Intronic	19.360 %	19.880 %	18.896 %	19.057 %
Synonymous coding	0.307 %		0.204 %	
Non-synonymous coding	0.176 %		0.103 %	
Essential splice site	0.000 %	0.000 %	0.000 %	0.001 %
Stop gain	0.002 %	0.000 %	0.000 %	0.000 %
Stop lost	0.000 %	0.000 %	0.000 %	0.000 %
In-frame codon insertion or deletion		0.016 %		0.011 %
Frameshift		0.014 %		0.005 %
Two or more consequences	15.436 %	15.710 %	20.401 %	21.467 %
Within, or 5Kb upstream or downstream from, non-coding gene or mature microRNA	6.240 %	6.280 %	2.362 %	2.403 %
Intergenic	50.314 %	47.772 %	50.246 %	48.591 %

Table 2 Variants in genes linked to traits of interest by GWAS or in Mendelian obesity genes

Source	Total genes	SNPs	Indels	Private SNPs	Private Indels
GWAS obesity	26	62	7	3	0
GWAS diabetes	85	208	27	7	2
GWAS metabolic syndrome	246	594	85	27	3
Monogenic syndromic obesity	13	45	2	0	0
Monogenic non-syndromic obesity	4	3	1	0	0
Monogenic non-syndromic lipodystrophy	3	4	0	0	0

were classified as deleterious by both SIFT and PROVEAN (shown in bold in Table 4). We then conducted a literature review of the biological function of the 29 genes in the *tabw2* interval with SNPs classified as deleterious by either algorithm, searching for potential link to obesity (Table 4). From this search, the *Cidec* gene, containing a SNP identified as deleterious by both SIFT and PROVEAN, drew our attention. CIDEC is a lipid droplet protein that is involved in the regulation of cellular lipid droplet size and lipid storage during lipid metabolism in adipocytes [50]. A loss of function mutation in CIDEC causes a familial partial lipodystrophy [43]. Further, weight loss via a low calorie diet was correlated with a reduced gene expression of CIDEC in adipose tissue in humans [51].

There was one nucleotide substitution in the *Cidec* coding sequence, 136 C > A. The substitution of 136



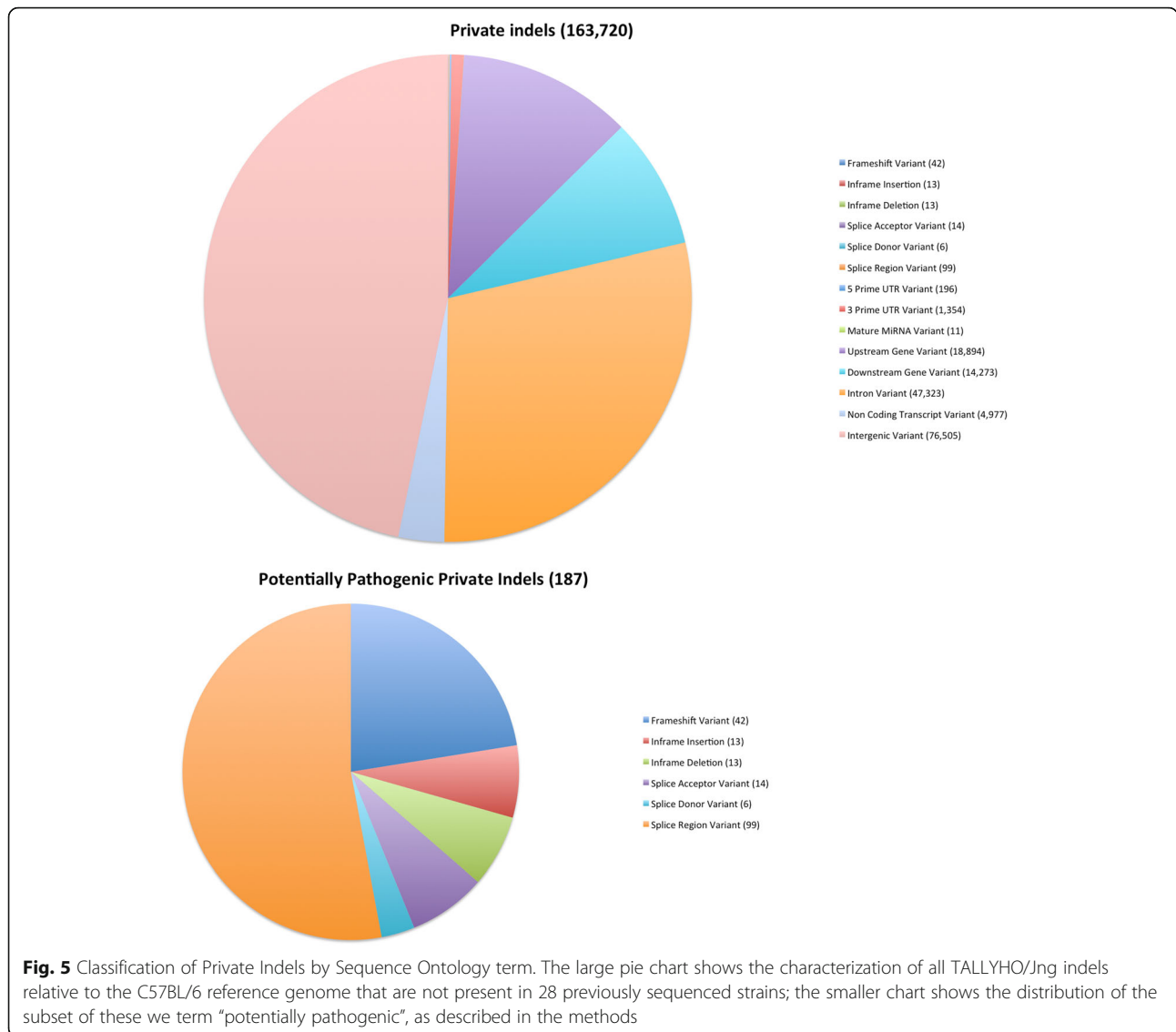


Table 3 Numbers of variants in the *tabw2* locus

	Region	
	<i>Tabw2a</i>	<i>Tabw2b</i>
Total SNPs	81,303	42,668
In protein coding genes	43,878	24,355
Potentially pathogenic	247	107
SIFT deleterious	22	8
PROVEAN deleterious	9	9
Total indels	20,814	7,808
In protein coding genes	11,302	4,445
Potentially pathogenic	38	11

C > A resulted in an amino acid difference of R46S (Arginine 46 Serine) between B6 and TH strains (R46 in B6; S46 in TH). To examine the function of the *Cidec* R46S polymorphism, we created murine plasmids AcGFP1-CIDEC (R46) and AcGFP1-CIDEC (S46) and transiently transfected them into COS-1 cells. COS-1 cells have been used as models for monitoring CIDEC function in prior studies because they lack endogenous Cidec activity [42–44]. It was apparent that AcGFP1-CIDEC was localized to distinct lipid droplets stained with Nile red (Fig. 6). When we considered Nile red lipid staining specifically associated with GFP fluorescence (co-localization), COS-1 cells transfected with the S46 variant exhibited an increase in lipid accumulation compared to cells transfected with the wild type R46 variant in the media both with and without oleic acid (Fig. 6). We

Table 4 SNPs in the *tabw2* locus identified as deleterious by SIFT or PROVEAN

Gene	Location on Chr 6 (bp)	Codon Ref/TH	A.A. Ref/TH	Exon	SIFT score	PROVEAN score	Function	Ref
<i>Gcfc2</i>	81923813	tCt/tTt	S/F	1/18	0.05	-2.5	Unknown	[59]
	81933786	cCa/cAa	P/Q	5/18	0.02	1.05		
	81941369	Cgg/Tgg	R/W	7/17	0	-5.41		
<i>Ccdc142</i>	83107557	cCt/cTt	P/L	7/11	0.15	-2.97	Unknown	
<i>Htra2</i>	83053837	Ggg/Cgg	G/R	1/7	0	N/A	Serine protease	[60]
	83054296	cCt/cTt	P/L	1/7	0	-2.04		
<i>Mogs</i>	83118580	Cgc/Tgc	R/C	4/4	0.03	0.94	Endoplasmic reticulum glucosidase	[61]
<i>Rtkn</i>	831151299	Ccc/Gcc	P/A	12/13	0.14	-4.41	Rho effector protein	[62]
	83151312	gCc/gTc	A/V	11/12	0.03	-1.1		
<i>1700003E16Rik</i>	83162665	aCc/aTc	T/I	4/4	0.01	-1.12	Unknown	
<i>Tet3</i>	83404686	Ctt/Ttt	L/F	3/3	0.01	-1.65	5-methylcytosine oxidase	[63]
<i>Aak1</i>	86965533	aAt/aGt	N/S	14/21	0.09	-2.66	Endocytosis	[64]
<i>Gfpt1</i>	87053776	aCa/aTa	T/I	4/5	0	-0.48	Hexosamine biosynthesis	[65]
<i>D6Ertd527e</i>	87111649	ACc/CTc	T/L	2/2	N/A	-4.67	Unknown	
<i>Chchd6</i>	89467434	Cgt/Tgt	R/C	5/8	0.03	-3.51	Mitochondrial protein	[66]
<i>Zfyve20</i>	92199851	cGg/cAg	R/Q	7/12	0.02	-1.32	Regulate the intracellular route of internalized receptors	[67]
<i>Lmod3</i>	97252294	aGt/aAt	S/N	1/3	0.02	-0.55	Actin thin filament length regulation	[68]
<i>Ttl3</i>	113409684	Ggg/Agg	G/R	2/2	0.04	-1.33	Glycylases	[69]
<i>Cidec</i>	113433131	Cgt/Agt	R/S	3/6	0	-5.85	Lipid storage and lipid droplet formation	[70]
<i>Irak2</i>	113647785	Ctg/Gtg	L/V	2/12	0	-2.4	Toll-like receptor signaling	[71]
<i>Rpl28-ps4</i>	117213939	cAa/cGa	Q/R	1/1	0	-0.09	Unknown	
<i>Ret</i>	118179286	aCc/aTc	T/I	5/20	0	-3.05	Proto-oncogene	[72]
<i>Cacna1c</i>	118602267	aCt/aAt	T/N	42/47	0.43	-2.93	L-type Ca ²⁺ channel	[73]
<i>Cecr5</i>	120510398	Cgt/Tgt	R/C	8/8	0.04	-1.85	Unknown	
<i>A2m</i>	121639361	atG/atA	M/I	3/36	0.02	-0.6	Protease	[74]
	121661523	aCg/aAg	T/K	21/36	0	-5.88		
<i>8430419L09Rik</i>	135209106	tCg/tTg	S/L	2/2	N/A	-6.0	Unknown	
	135209130	aTg/aCg	M/T	2/2	N/A	-6.0		
	135209172	cAg/cGg	Q/R	2/2	N/A	-4.0		
	135209250	cGg/cAg	R/Q	2/2	N/A	-4.0		
	135209265	gCc/gTc	A/V	2/2	N/A	-4.0		
<i>Atf7ip</i>	136561386	tCg/tTg	S/L	3/4	0.01	-0.72	Transcriptional cofactor	[75]
	136561422	aGt/aCt	S/T	3/4	0.05	-0.12		
<i>BC049715</i>	136840037	Act/Cct	T/P	1/1	0.11	-3.22	Unknown	
	136840106	Aga/Gga	R/G	1/1	0.37	-2.86		
	136840191	aTg/aCg	M/T	1/1	0.01	-4.56		
<i>Art4</i>	136857180	Ccc/Acc	P/T	1/3	0.02	-1.15	ADP-ribosyltransferase	[76]
<i>Mgp</i>	136872651	aAa/aGa	K/R	4/4	0.02	-0.33	Calcification inhibitor	[77]
<i>Ptpro</i>	137368083	aCg/aTg	T/M	2/27	0.03	-1.93	Tumor suppressor	[78]
<i>Igpb1b</i>	138658041	aCg/aTg	T/M	1/1	0.05	-1.27	Regulation of phosphatase activity in neuronal cells	[79]
<i>Pik3c2g</i>	139967890	Agt/Tgt	S/C	16/16	0.01	-2.86	Long-term Akt2 activation	[80]

Chr chromosome, A.A. amino acid, Ref reference, TH TALLYHO

SNPs predicted to be deleterious by both SIFT and PROVEAN are shown in bold

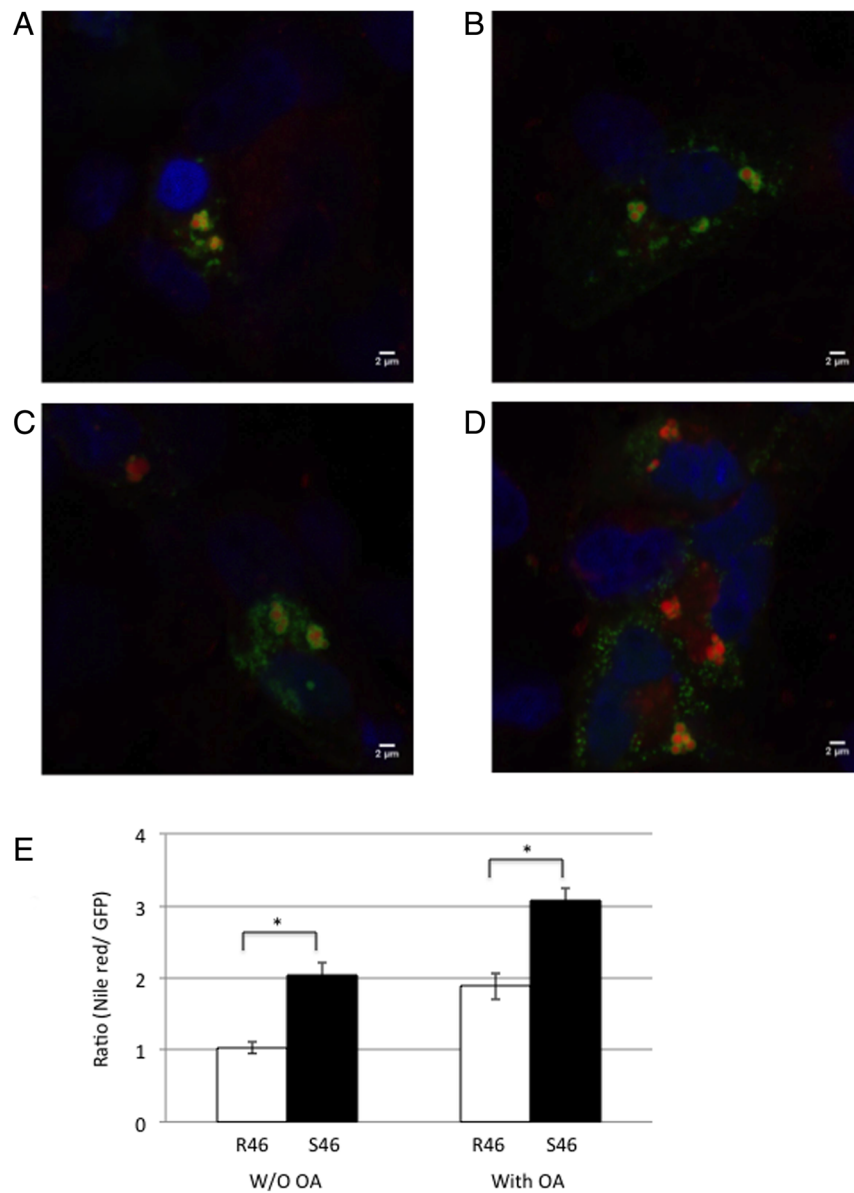


Fig. 6 Representative images of lipid droplets in COS-1 cells transfected with (a and b) AcGFP1-CIDEc (R46) and (c and d) AcGFP1-CIDEc (S46). a and c present cells cultured in media without oleic acid (OA) and b and d present cells cultured in media with OA. Images of nuclei labeled with DAPI, neutral lipid labeled with Nile red and location of AcGFP1 were merged. e Quantification of lipid specifically associated with GFP in cells. Using ImageJ software, the ratio of Nile red/GFP signal was calculated from 34 lipid droplet clusters each for R46 and S46 without OA and 47 lipid droplet clusters for R46 and 46 lipid droplet clusters for S46 with OA. Data are presented as means ± SEM. *P < 0.05

speculate that the S46 variant may be hypermorphic in that it enhances the function of CIDEc in promoting lipid accumulation in lipid droplets.

Discussion

Landscape of the TALLYHO genome

We found a total of 5,375,638 variants (SNPs and indels) in the TH genome, compared to the reference B6 genome (GRCm38). As expected, the large majority of variants occurred in non-protein coding regions of the

genome. The number of SNPs was broadly consistent with that reported by Keane et al. and Wong et al. for 13 laboratory strains [30, 34], and substantially fewer than the four wild-derived strains (PWK/PhJ, CAST/Eij, WSB/Eij, and SPRET/Eij) sequenced in the Keane et al. study [34]. We found 3.5 % to 33.6 % more indels when compared to those 13 laboratory strains, and again considerably fewer than the four wild-derived strains. The proportion of private SNPs we observed was higher than 10 of the laboratory strains from Keane et al. [34], with

the two exceptions being the two strains NOD/ShiLtJ and NZO/HILtJ. We also showed a lower proportion of private SNPs in the TH strain than discovered in FVB/NJ [30]. The proportion of private indels we observed in TH was higher than in these studies. Although the origins of the Theiler Original mouse are unknown, the generally higher proportion of private variants in TH (compared to most laboratory strains) is consistent with the possibility that TH has a unique ancestor. We also note that our sequencing coverage of 64.8X was higher than in these other studies, increasing our general power to detect variants.

Classification of variants

We provided a comprehensive catalog of variants of the TH mouse, relative to the B6 mouse, classified according to the effect of the variant on protein sequence. These effects were predicted by comparing the location of the variant on the reference genome to known genomic annotations provided by the Ensembl gene database [52]. While it is relatively straightforward to automate the classification of a large number of variants via genomic annotations on a coarse scale, providing detailed interpretations of the effect of some mutations requires greater inspection of the context of the mutation that cannot be readily automated and performed on a genome-wide scale. For example, the effect of intronic mutations is not well characterized or understood at present. Our aims in this study include both providing a general overview of the landscape of the TH mouse genome relative to the reference genome, and also providing sufficient detail of the nature of the 5.3 million variants discovered. To this end, we used an automated process to classify variants into a large number of precisely defined categories, and then subsequently reduced the number of these categories, subjectively categorizing them into a smaller set of representative terms. The first categorization provides a fine-grained resource for searching for individual variants that may have phenotypic consequences, and the second provides a general overview of the landscape of the TH genome.

The Sequence Ontology (SO) [37] provides a controlled vocabulary for describing features and annotations associated with a biological sequence, including mutations at the sequence level, along with a formalized description of the relationships between these terms. The Variant Effect Predictor (VEP) [36] is a tool that predicts the effect of genomic variants by analyzing the variant in the context of annotations on the reference genomic sequence. For each variant and each known transcript containing that variant, VEP outputs a subset of SO terms associated with that variant. In many cases, this results in multiple SO terms being associated with a single variant in the context of a single transcript. For

example, we found an A to G single nucleotide polymorphism (SNP) at genomic location 126,585,377 on chromosome 7. This SNP occurs in the second nucleotide of the first codon of the second exon of the apolipoprotein B receptor gene, *Apob*, mutating the GAC codon to GGC. Since this mutation results in a change in amino acid from aspartic acid to glycine, the SNP is annotated with the SO term “missense_variant”. Since this variant also occurs within three bases of a splice site, it is also annotated with the SO term “splice_region_variant”. In order to provide a complete, unbiased characterization of the types of variants discovered, we list all unique combinations of SO terms for a single variant and the number of times those combinations occur (Additional file 2: Table S2 and Additional file 3: Table S3).

In order to provide a concise overview of the relative proportions of occurrences of the major types of variants, for each unique combination of SO terms we selected a “representative” SO term (Additional file 2: Table S2 and Additional file 3: Table S3, second column). These overviews are presented in Figs. 2 and 3.

Due to the presence of multiple transcripts for some genes, and due to the presence of overlapping genes, some individual variants are present in multiple transcripts. In this case, VEP outputs a potentially different combination of SO terms for each of the transcripts in which the variant occurs. When there are multiple, distinct, combinations for a given variant due to its presence in multiple transcripts, these variants are categorized in Additional file 2: Table S2 and Additional file 3: Table S3 as “multiple classification sets”. In order to uniquely classify such variants in the overview (Figs. 2 and 3), we picked a representative SO term for each set of SO terms using the strategy shown in Additional file 2: Table S2 and Additional file 3: Table S3, and then chose the most pathogenic out of these representative SO terms using the priority described in the methods.

Filtering variants by the human GWAS catalog

We employed a translational methodology to identify and prioritize candidate genes for complex disorders for which TH is a model. We examined all the TH variants for potential link to relevant human disease using the EBI GWAS catalog [39] and generated enriched candidate gene lists for obesity, diabetes and metabolic syndrome. We reviewed biological function for selected genes and identify genes we could connect to obesity and/or diabetes. GLIS3 is a transcription factor that plays an important role in pancreatic development and insulin gene expression in beta cells [53]. SORBS1 is involved in insulin stimulated glucose transport in adipocytes [54]. IGF2BP2 is a RNA binding protein, participating in posttranscriptional RNA processing, i.e., RNA

splicing, stabilization, transport, and translation [55]. Inactivation of the *Igf2bp2* gene caused a resistance to diet-induced obesity in mice [56]. Semaphorin 5A (*Sema5A*) is an axon regulator molecule and plays a role during neuronal and vascular development [57]. Collectively, our convergent approaches systematically integrate whole genome sequencing data and genetic information from GWAS-derived findings and provide an opportunity to discover candidate genes for further functional validation.

Missense polymorphism of R46S in CIDEC between B6 and TH mice

Among the genes with SNPs that were identified as deleterious by both SIFT and PROVEAN (Additional file 5: Table S5), were orthologous to Mendelian lipodystrophy genes, and were contained within the *tabw2* interval, we identified *Cidec*, which is an important regulator of energy homeostasis directly involved in promoting the accumulation of triglyceride into intracellular lipid droplets [50]. Lipid droplets are spherical organelles found in many types of eukaryotic cells, including adipocytes, and are composed of a core of neutral lipids, such as sterol esters or triglycerides, surrounded by a monolayer of phospholipids, free cholesterol, and multiple specific proteins including CIDEC [50]. In adipocytes, cellular energy is stored as triglycerides in lipid droplets, and in conditions of fatty acid excess, lipid droplets rapidly increase their volumes [58]. The capacity of CIDEC in accumulating triglycerides into intracellular lipid droplets was demonstrated by transfection experiments using multiple cell types including C pre-adipocytes, 293T cells, and COS cells [42–44]. For example, transfection of full-length murine *Cidec* into COS-7 cells increased total cellular triglycerides by 50 % [44]. We found that the TH allele of CIDEC S46 variant allowed for more lipid accumulation than the B6 allele of CIDEC R46 variant in COS-1 cells. Further functional validation of this missense polymorphism in vivo may provide an opportunity to understand the role of *Cidec* in the context of obesity.

Conclusions

We have sequenced the whole genome of obese type 2 diabetic TH mice by next-generation sequencing and generated a complete catalog of variants classified by location relative to genes and predicted consequences for protein products. We filtered the list of variants to the *tabw2* obesity QTL and identified a missense polymorphism in *Cidec* whose protein function we characterized. Furthermore, by comparing the variant catalog to the human GWAS catalog and to known human Mendelian obesity genes, we were able to identify a list

of susceptibility candidate genes that could be used to dissect the components of polygenic diseases such as obesity and type 2 diabetes.

Additional files

Additional file 1: Table S1. Mouse genome project strains. List of strains from the Mouse Genome Project used in the determination of "private" variants. (DOCX 43 kb)

Additional file 2: Table S2. Classification of TH SNPs by sets of SO terms associated with the SNP. (XLSX 9 kb)

Additional file 3: Table S3. Classification of TH indels by sets of SO terms associated with the indel. (XLSX 10 kb)

Additional file 4: Table S4. PCR primer sequences used for validation of selected variants. (XLSX 56 kb)

Additional file 5: Table S5. Variants in orthologs of GWAS-associated and Mendelian Genes. List of orthologous genes to human genes associated by GWAS studies to obesity, diabetes, and metabolic syndrome, and to Mendelian obesity genes, with variant type and number. (XLSX 74 kb)

Additional file 6: Table S6. Private SNP classification. Classification of private TH SNPs by sets of SO terms associated with the SNP. (XLSX 8 kb)

Additional file 7: Table S7. Private indel classification. Classification of private TH indels by sets of SO terms associated with the indel. (XLSX 8 kb)

Additional file 8: Table S8. Genes with pathogenic private SNPs. List of genes with pathogenic private SNPs, with SIFT scores and SO terms for the corresponding SNP. (XLSX 64 kb)

Additional file 9: Table S9. Genes with pathogenic private indels. List of genes with pathogenic private indels, with SO terms for the corresponding indel. (XLSX 54 kb)

Abbreviations

DNA: Deoxyribonucleic acid; EBI: European Bioinformatics Institute; GWAS: Genome Wide Association Study; Indel: Insertion and/or Deletion; Mbase: Megabase (1,000,000 bases); MGP: Mouse Genome Project; NCBI: National Center for Biotechnology Information; PROVEAN: Protein Variation Effect Analyzer; QTL: Quantitative Trait Locus; SIFT: Sorting Intolerant From Tolerant; SNP: Single Nucleotide Polymorphism; SO: Sequence Ontology; SRA: Sequence Read Archive; *Tabw2*: Tallyho Associated Body Weight 2; TH: TALLYHO/Jng mouse strain; TJL: The Jackson Laboratory; VEP: Variant Effect Predictor

Acknowledgments

This work was supported in part by the NIH/NIDDK-R01DK077202, NIH/NCRR P20 RR016477 and NIH/NIGMS P20GM103434 which funds the IDeA WV-INBRE program and supports the Marshall University Genomics Core Facility and in part by AHA-0855300E. Points of view in this document are those of the authors and do not necessarily represent the official position or views of the NIH or the AHA. We thank Kristy D. Dillon and Taryn P. Stewart for maintaining mouse colonies, and David Neff in the Marshall University Imaging Core Facility for his assistance in confocal microscopy.

Availability of data and materials

Raw sequencing data from this study are deposited in the National Center for Biotechnology Information (NCBI) Sequencing Read Archive and are available via accession number SRP067703.

Authors' contributions

JD performed all data processing and bioinformatics, including alignment, variant calling, identification of private variants, and cross-referencing to the GWAS catalog and Mendelian obesity genes. GB and JF prepared next-generation sequencing libraries and performed the sequencing. JHK and JKP prepared genomic DNA and conducted cell culture and transfection experiments and PCR. JHK, JD and DAP designed the overall research project and drafted the manuscript, tables and figures. All authors except GB, who died unexpectedly during the course of the study, read and approved the final manuscript.

Competing interest

The authors declare that they have no competing interests.

Received: 1 January 2016 Accepted: 2 November 2016

Published online: 11 November 2016

References

- Segula D. Complications of obesity in adults: a short review of the literature. *Malawi Med J.* 2014;26(1):20–4.
- Dailey G, Wang E. A review of cardiovascular outcomes in the treatment of people with type 2 diabetes. *Diabetes Ther.* 2014;5(2):385–402.
- Grarup N, Sandholt CH, Hansen T, Pedersen O. Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia.* 2014;57(8):1528–41.
- Xia Q, Grant SF. The genetics of human obesity. *Ann N Y Acad Sci.* 2013;1281:178–90.
- Ali O. Genetics of type 2 diabetes. *World J Diabetes.* 2013;4(4):114–23.
- Basile KJ, Johnson ME, Xia Q, Grant SF. Genetic susceptibility to type 2 diabetes and obesity: follow-up of findings from genome-wide association studies. *International journal of endocrinology.* 2014;2014:769671.
- McMurray F, Moir L, Cox RD. From mice to humans. *Curr Diab Rep.* 2012;12(6):651–8.
- Joost HG, Schurmann A. The genetic basis of obesity-associated type 2 diabetes (diabesity) in polygenic mouse models. *Mamm Genome.* 2014; 25(9–10):401–12.
- Leiter EH, Strobel M, O'Neill A, Schultz D, Schile A, Reifsnnyder PC. Comparison of two new mouse models of polygenic type 2 diabetes at the Jackson laboratory, NONcNZO10LtJ and TALLYHO/JngJ. *J Diabetes Res.* 2013;2013:165327.
- Kim JH, Saxton AM. The TALLYHO mouse as a model of human type 2 diabetes. *Methods Mol Biol.* 2012;933:75–87.
- Mao X, Dillon KD, McEntee MF, Saxton AM, Kim JH. Islet insulin secretion, β -cell mass, and energy balance in a polygenic mouse model of type 2 diabetes with obesity. *J Inborn Errors Metab Screen.* 2014;2.
- Chen Z, Guo L, Zhang Y, Walzem RL, Pendergast JS, Printz RL, Morris LC, Matafonova E, Stien X, Kang L, et al. Incorporation of therapeutically modified bacteria into gut microbiota inhibits obesity. *J Clin Invest.* 2014;124(8):3391–406.
- Neschen S, Scheerer M, Seelig A, Huypens P, Schultheiss J, Wu M, Wurst W, Rathkolb B, Suhre K, Wolf E, et al. Metformin supports the antidiabetic effect of a sodium glucose cotransporter 2 inhibitor by suppressing endogenous glucose production in diabetic mice. *Diabetes.* 2015;64(1):284–90.
- Ostler JE, Maurya SK, Dials J, Roof SR, Devor ST, Ziolo MT, Periasamy M. Effects of insulin resistance on skeletal muscle growth and exercise capacity in type 2 diabetic mouse models. *Am J Physiol Endocrinol Metab.* 2014;306(6):E592–605.
- Buck 2nd DW, Jin DP, Geringer M, Hong SJ, Galiano RD, Mustoe TA. The TallyHo polygenic mouse model of diabetes: implications in wound healing. *Plast Reconstr Surg.* 2011;128(5):427e–37e.
- Nguyen KT, Seth AK, Hong SJ, Geringer MR, Xie P, Leung KP, Mustoe TA, Galiano RD. Deficient cytokine expression and neutrophil oxidative burst contribute to impaired cutaneous wound healing in diabetic, biofilm-containing chronic wounds. *Wound Repair Regen.* 2013;21(6):833–41.
- Wagner IJ, Szpalski C, Allen Jr RJ, Davidson EH, Canizares O, Saadeh PB, Warren SM. Obesity impairs wound closure through a vasculogenic mechanism. *Wound Repair Regen.* 2012;20(4):512–22.
- Li H, Yang H, Ding Y, Aprecio R, Zhang W, Wang Q, Li Y. Experimental periodontitis induced by *Porphyromonas gingivalis* does not alter the onset or severity of diabetes in mice. *J Periodontol Res.* 2013;48(5):582–90.
- Hong SJ, Jin da P, Buck 2nd DW, Galiano RD, Mustoe TA. Impaired response of mature adipocytes of diabetic mice to hypoxia. *Exp Cell Res.* 2011;317(16):2299–307.
- Sherwani SI, Aldana C, Usmani S, Adin C, Kotha S, Khan M, Eubank T, Scherer PE, Parinandi N, Magalang UJ. Intermittent hypoxia exacerbates pancreatic beta-cell dysfunction in a mouse model of diabetes mellitus. *Sleep.* 2013;36(12):1849–58.
- Devlin MJ, Van Vliet M, Motyl K, Karim L, Brooks DJ, Louis L, Conlon C, Rosen CJ, Boussein ML. Early-onset type 2 diabetes impairs skeletal acquisition in the male TALLYHO/JngJ mouse. *Endocrinology.* 2014;155(10):3806–16.
- Won HY, Lee JA, Park ZS, Song JS, Kim HY, Jang SM, Yoo SE, Rhee Y, Hwang ES, Bae MA. Prominent bone loss mediated by RANKL and IL-17 produced by CD4+ T cells in TallyHo/JngJ mice. *PLoS One.* 2011;6(3):e18168.
- Nascimento NF, Hicks JA, Carlson KN, Hatzidis A, Amaral DN, Logan RW, Seggio JA. Long-term wheel-running and acute 6-h advances alter glucose tolerance and insulin levels in TALLYHO/JngJ mice. *Chronobiol Int.* 2015;1–9.
- Cheng ZJ, Jiang YF, Ding H, Severson D, Triggler CR. Vascular dysfunction in type 2 diabetic TallyHo mice: role for an increase in the contribution of PGH2/TxA2 receptor activation and cytochrome p450 products. *Can J Physiol Pharmacol.* 2007;85(3–4):404–12.
- Didion SP, Lynch CM, Faraci FM. Cerebral vascular dysfunction in TallyHo mice: a new model of Type II diabetes. *Am J Physiol Heart Circ Physiol.* 2007;292(3):H1579–83.
- Li Y, Mihara K, Saifeddine M, Krawetz A, Lau DC, Li H, Ding H, Triggler CR, Hollenberg MD. Perivascular adipose tissue-derived relaxing factors: release by peptide agonists via proteinase-activated receptor-2 (PAR2) and non-PAR2 mechanisms. *Br J Pharmacol.* 2011;164(8):1990–2002.
- Yalcin B, Wong K, Bhomra A, Goodson M, Keane TM, Adams DJ, Flint J. The fine-scale architecture of structural variants in 17 mouse genomes. *Genome Biol.* 2012;13(3):R18.
- Kim JH, Sen S, Avery CS, Simpson E, Chandler P, Nishina PM, Churchill GA, Naggert JK. Genetic analysis of a new mouse model for non-insulin-dependent diabetes. *Genomics.* 2001;74(3):273–86.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9.
- Wong K, Bumpstead S, Van Der Weyden L, Reinholdt LG, Wilming LG, Adams DJ, Keane TM. Sequencing and characterization of the FVB/NJ mouse genome. *Genome Biol.* 2012;13(8):R72.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England).* 2009;25(16):2078–9.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–303.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics (Oxford, England).* 2011;27(15):2156–8.
- Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature.* 2011;477(7364):289–94.
- Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003;31(13):3812–4.
- McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics (Oxford, England).* 2010;26(16):2069–70.
- Eilbeck K, Lewis SE. Sequence Ontology Annotation Guide. *Comparative and Functional Genomics.* 2004;5(8):642–7.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012;7(10):e46688.
- Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(Database issue):D1001–6.
- Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database : the journal of biological databases and curation.* 2011;2011:bar030.
- Pigeyre M, Yazdi FT, Kaur Y, Meyre D. Recent progress in genetics, epigenetics and metagenomics unveils the pathophysiology of human obesity. *Clin Sci.* 2016;130(12):943–86.
- Keller P, Petrie JT, De Rose P, Gerin I, Wright WS, Chiang SH, Nielsen AR, Fischer CP, Pedersen BK, MacDougald OA. Fat-specific protein 27 regulates storage of triacylglycerol. *J Biol Chem.* 2008;283(21):14355–65.
- Rubio-Cabezas O, Puri V, Murano I, Saudek V, Sempke RK, Dash S, Hyden CS, Bottomley W, Vigouroux C, Magre J, et al. Partial lipodystrophy and insulin resistant diabetes in a patient with a homozygous nonsense mutation in CIDEA. *EMBO Mol Med.* 2009;1(5):280–7.
- Grahn TH, Kaur R, Yin J, Schweiger M, Sharma VM, Lee MJ, Ido Y, Smas CM, Zechner R, Lass A, et al. Fat-specific protein 27 (FSP27) interacts with

- adipose triglyceride lipase (ATGL) to regulate lipolysis and insulin sensitivity in human adipocytes. *J Biol Chem.* 2014;289(17):12029–39.
45. Liu K, Zhou S, Kim JY, Tillison K, Majors D, Rearick D, Lee JH, Fernandez-Boyanapalli RF, Barricklow K, Houston MS, et al. Functional analysis of FSP27 protein regions for lipid droplet localization, caspase-dependent apoptosis, and dimerization with CIDEA. *Am J Physiol Endocrinol Metab.* 2009;297(6):E1395–413.
 46. Simecek P, Churchill GA, Yang H, Rowe LB, Herberg L, Serreze DV, Leiter EH. Genetic analysis of substrain divergence in non-obese diabetic (NOD) mice. *G3.* 2015;5(5):771–5.
 47. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073–81.
 48. Kim JH, Stewart TP, Zhang W, Kim HY, Nishina PM, Naggert JK. Type 2 diabetes mouse model TallyHo carries an obesity gene on chromosome 6 that exaggerates dietary obesity. *Physiol Genomics.* 2005;22(2):171–81.
 49. Stewart TP, Mao X, Aqqad MN, Ufford D, Dillon KD, Saxton AM, Kim JH. Subcongenic analysis of tabw2 obesity QTL on mouse chromosome 6. *BMC Genet.* 2012;13:81.
 50. Matsusue K. A physiological role for fat specific protein 27/cell death-inducing DFF45-like effector C in adipose and liver. *Biol Pharm Bull.* 2010;33(3):346–50.
 51. Magnusson B, Gummesson A, Glad CA, Goedecke JH, Jernas M, Lystig TC, Carlsson B, Fagerberg B, Carlsson LM, Svensson PA. Cell death-inducing DFF45-like effector C is reduced by caloric restriction and regulates adipocyte lipid metabolism. *Metabolism.* 2008;57(9):1307–13.
 52. Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. Ensembl 2015. *Nucleic Acids Res.* 2015;43(Database issue):D662–9.
 53. Kang HS, Kim YS, ZeRuth G, Beak JY, Gerrish K, Kilic G, Sosa-Pineda B, Jensen J, Pierreux CE, Lemaigre FP, et al. Transcription factor Glis3, a novel critical player in the regulation of pancreatic beta-cell development and insulin gene expression. *Mol Cell Biol.* 2009;29(24):6366–79.
 54. Baumann CA, Ribon V, Kanzaki M, Thurmond DC, Mora S, Shigematsu S, Bickel PE, Pessin JE, Saltiel AR. CAP defines a second signalling pathway required for insulin-stimulated glucose transport. *Nature.* 2000;407(6801):202–7.
 55. Yisraeli JK. VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. *Biol Cell.* 2005;97(1):87–96.
 56. Dai N, Zhao L, Wrighting D, Kramer D, Majithia A, Wang Y, Cracan V, Borges-Rivera D, Mootha VK, Nahrendorf M, et al. IGF2BP2/IMP2-Deficient mice resist obesity through enhanced translation of Ucp1 mRNA and Other mRNAs encoding mitochondrial proteins. *Cell Metab.* 2015;21(4):609–21.
 57. Sadanandam A, Rosenbaugh EG, Singh S, Varney M, Singh RK. Semaphorin 5A promotes angiogenesis by increasing endothelial cell proliferation, migration, and decreasing apoptosis. *Microvasc Res.* 2010;79(1):1–9.
 58. Sztalryd C, Kimmel AR. Perilipins: lipid droplet coat proteins adapted for tissue-specific energy storage and utilization, and lipid cytoprotection. *Biochimie.* 2014;96:96–101.
 59. Melville SA, Buros J, Parrado AR, Vardarajan B, Logue MW, Shen L, Risacher SL, Kim S, Jun G, DeCarli C, et al. Multiple loci influencing hippocampal degeneration identified by genome scan. *Ann Neurol.* 2012;72(1):65–75.
 60. Liu D, Liu X, Wu Y, Wang W, Ma X, Liu H. Cloning and Transcriptional Activity of the Mouse Omi/HtrA2 Gene Promoter. *Int J Mol Sci.* 2016;17(1):119. <http://www.mdpi.com/1422-0067/17/1/119>.
 61. Chang J, Block TM, Guo JT. Viral resistance of MOGS-CDG patients implies a broad-spectrum strategy against acute virus infections. *Antivir Ther.* 2015;20(3):257–9.
 62. Qu GQ, Lu YM, Liu YF, Liu Y, Chen WX, Liao XH, Kong WM. Effect of RTKN on progression and metastasis of colon cancer in vitro. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie.* 2015;74:117–23.
 63. Jin SG, Zhang ZM, Dunwell TL, Harter MR, Wu X, Johnson J, Li Z, Liu J, Szabo PE, Lu Q, et al. Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Rep.* 2016;14(3):493–505.
 64. Conner SD, Schmid SL. Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J Cell Biol.* 2002;156(5):921–9.
 65. Chen Q, Muller JS, Pang PC, Laval SH, Haslam SM, Lochmuller H, Dell A. Global N-linked Glycosylation is Not Significantly Impaired in Myoblasts in Congenital Myasthenic Syndromes Caused by Defective Glutamine-Fructose-6-Phosphate Transaminase 1 (GFPT1). *Biomolecules.* 2015;5(4):2758–81.
 66. Ding C, Wu Z, Huang L, Wang Y, Xue J, Chen S, Deng Z, Wang L, Song Z, Chen S. Mitofilin and CHCHD6 physically interact with Sam50 to sustain cristae structure. *Sci Rep.* 2015;5:16064.
 67. Stockler S, Corvera S, Lambright D, Fogarty K, Nosova E, Leonard D, Steinfeld R, Ackerley C, Shyr C, Au N, et al. Single point mutation in Rabenosyn-5 in a female with intractable seizures and evidence of defective endocytotic trafficking. *Orphanet J Rare Dis.* 2014;9:141.
 68. Sandaradura S, North KN. LMOD3: the “missing link” in nemaline myopathy? *Oncotarget.* 2015;6(29):26548–9.
 69. Rocha C, Papon L, Cacheux W, Marques Sousa P, Lascano V, Tort O, Giordano T, Vacher S, Lemmers B, Mariani P, et al. Tubulin glycosylases are required for primary cilia, control of cell proliferation and tumor development in colon. *EMBO J.* 2014;33(19):2247–60.
 70. Xu Y, Gu Y, Liu G, Zhang F, Li J, Liu F, Zhang Z, Ye J, Li Q. Cidec promotes the differentiation of human adipocytes by degradation of AMPKalpha through ubiquitin-proteasome pathway. *Biochim Biophys Acta.* 2015;1850(12):2552–62.
 71. Flannery SM, Keating SE, Szymak J, Bowie AG. Human interleukin-1 receptor-associated kinase-2 is essential for Toll-like receptor-mediated transcriptional and post-transcriptional regulation of tumor necrosis factor alpha. *J Biol Chem.* 2011;286(27):23688–97.
 72. Hedayati M, Zarif Yeganeh M, Sheikhol Eslami S, Rezghi Barez S, Hoghooghi Rad L, Azizi F. Predominant RET Germline Mutations in Exons 10, 11, and 16 in Iranian Patients with Hereditary Medullary Thyroid Carcinoma. *J Thyroid Res.* 2011;2011:264248.
 73. Gaber EM, Jayaprakash P, Qureshi MA, Parekh K, Oz M, Adrian TE, Howarth FC. Effects of a sucrose-enriched diet on the pattern of gene expression, contraction and Ca(2+) transport in Goto-Kakizaki type 2 diabetic rat heart. *Exp Physiol.* 2014;99(6):881–93.
 74. Bruno E, Quattrocchi G, Nicoletti A, Le Pira F, Maci T, Mostile G, Andreoli V, Quattrone A, Zappia M. Lack of interaction between LRP1 and A2M polymorphisms for the risk of Alzheimer disease. *Neurosci Lett.* 2010;482(2):112–6.
 75. Sasai N, Saitoh N, Saitoh H, Nakao M. The transcriptional cofactor MCAF1/ATF7IP is involved in histone gene expression and cellular senescence. *PLoS One.* 2013;8(7):e68478.
 76. Parusel I, Kahl S, Braasch F, Glowacki G, Halverson GR, Reid ME, Schawaldner A, Ortolan E, Funaro A, Malavasi F, et al. A panel of monoclonal antibodies recognizing GPI-anchored ADP-ribosyltransferase ART4, the carrier of the Dombrock blood group antigens. *Cell Immunol.* 2005;236(1-2):59–65.
 77. Epstein M. Matrix Gla-Protein (MGP) not only inhibits calcification in large arteries but also may be renoprotective: connecting the dots. *EBioMedicine.* 2016;4:16–7.
 78. Wang X, Yan S, Xu D, Li J, Xie Y, Hou J, Jiang R, Zhang C, Sun B. Aggravated Liver Injury but Attenuated Inflammation in PTPRO-Deficient Mice Following LPS/D-GalN Induced Fulminant Hepatitis. *Cell Physiol Biochem.* 2015;37(1):214–24.
 79. Maeda K, Inui S, Tanaka H, Sakaguchi N. A new member of the alpha4-related molecule (alpha4-b) that binds to the protein phosphatase 2A is expressed selectively in the brain and testis. *European journal of biochemistry / FEBS.* 1999;264(3):702–6.
 80. Braccini L, Ciraolo E, Campa CC, Perino A, Longo DL, Tibolla G, Pregnolato M, Cao Y, Tassone B, Damilano F, et al. PI3K-C2gamma is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. *Nat Commun.* 2015;6:7400.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

