Genome wide association study of body weight, body mass index, adiposity, and fasting glucose in 3,173 outbred rats

Apurva S. Chitre[1], Oksana Polesskaya[1], Katie Holl[2], Jianjun Gao[1], Riyan Cheng[1], Angel Garcia Martinez[3], Tony George[6], Alexander F. Gileta[1,10], Wenyan Han[3], Aidan Horvath[4], Alesa Hughson[4], Keita Ishiwari[6], Christopher P. King[5], Alexander Lamparelli[5], Cassandra L. Versaggi[5], Connor Martin[6], Celine L. St. Pierre[11], Jordan A. Tripi[5], Tengfei Wang [3], Hannah Wladecki[1], Hao Chen[3], Shelly B. Flagel[4], Paul Meyer[5], Jerry Richards[6], Terry E. Robinson[7], Abraham A. Palmer[1,8]*, Leah C. Solberg Woods[9]*

* These authors contributed equally to this study as senior authors

[1]University of California, San Diego, Department of Psychiatry, La Jolla, CA, [2]Medical College of Wisconsin, Human and Molecular Genetic Center, Milwaukee, WI, [3]University of Tennessee Health Science Center, Department of Pharmacology, Memphis, TN, [4]University of Michigan, Department of Psychiatry, Ann Arbor, MI, [5]University at Buffalo, Department of Psychology, Buffalo, NY, [6]University at Buffalo, Research Institute on Addictions, Buffalo, NY, [7]University of Michigan, Department of Psychology, Ann Arbor, MI, [8]Universiy of San Diego, Institute for Genomic Medicine, La Jolla, CA [9]Wake Forest School of Medicine, Department of Internal Medicine, Winston Salem, NC, [10] University of Chicago, Department of Human Genetics, Chicago, IL, [11] Department of Genetics, Washington University, St. Louis, MO

Corresponding Author:

Leah Solberg Woods

Wake Forest School of Medicine

1 Medical Center Drive

Winston Salem, NC 27157

lsolberg@wakehealth.edu

Abstract

Obesity is a global health crisis that is influenced by both genetic and environmental factors. Rodent model organisms can be used to understand the biological and genetic basis of obesity and related morphological traits. A major advantage of model organisms is that they can be studied under uniform environmental conditions, thus reducing the complex role of environment and gene by environment interactions. Furthermore, fat pads and other tissues can be dissected and weighed, so that their role in determining body weight can be precisely defined. Highly recombinant populations allow for genetic fine-mapping of complex traits, greatly reducing the number of plausible candidate genes. We performed the largest rat GWAS ever undertaken, using 3,173 male and female adult N/NIH heterogeneous stock (**HS**) rats, which were created by mixing 8 inbred strains. We identified 31 independent loci for body weight, body length, body mass index, fat pad weight (retroperitoneal, epididymal, and parametrial), and fasting glucose. We observed strong evidence of pleotropic effects across multiple phenotypes. Three loci contained only a single gene (*Epha5, Nrg1 and Klhl14),* whereas others were larger and contained many genes. We replicated a locus containing *Prlhr,* and a second locus containing *Adcy3,* which we had previously identified in a smaller HS rat study. Finally, by subsampling our dataset, we showed an exponential growth of significant loci as sample size increased towards 3,173. Our results demonstrate the potential for rodent studies to add to our understanding of the molecular genetic factors that contribute to obesity-relevant traits and emphasize the importance of sample size.

Introduction

Obesity is a growing health epidemic; over one third of the adult population and almost one fifth of all children in the United States are considered obese. There has been a steady increase in prevalence of obesity since the 1970's, and prevalence is continuing to rise (Hales et al. 2018). Obesity is a major risk factor for multiple diseases including cardiovascular disease, type 2 diabetes, cancer, and stroke (Wang et al. 2011), thereby placing a tremendous burden on society. Obesity is caused by interaction of genetic and environmental factors with genetic factors accounting for up to 70% of the population variance (Maes et al. 1997). Although human genome wide association studies (**GWAS**) of obesity have been extremely productive (Loos 2018), our biological understanding of the genetic architecture of obesity is far from complete.

Model organism studies of body weight, morphometric, and metabolic traits represent a complementary approach to understanding the genetic basis of obesity. GWAS in model organisms have traditionally been limited by the lack of recombination present in laboratory crosses. Heterogeneous stocks (**HS**) rats were created in 1984 by interbreeding eight inbred founder strains and maintaining them as an outbred population in a way that minimizes inbreeding (Hansen and Spuhler 1984). This strategy has made HS rats ideal for fine-mapping genetic loci (Solberg Woods 2014). Furthermore, HS founder strains have been fully sequenced (Baud et al. 2013), such that coding polymorphisms can be rapidly identified (Keele et al. 2018). A third advantage of using HS rats is the ability to collect tissues under controlled environmental conditions. These tissues can be used to measure gene expression, thus permitting the identification of expression QTLs (eQTLs; (Tsaih et al. 2014; Parker et al. 2016; Keele et al. 2018)). Our group has previously mapped adiposity traits in HS rats using 743 male HS rats, which identified two genetic loci for visceral adiposity and body weight (Keele et al. 2018).

Through the NIDA-funded Center for GWAS in outbred rats **(**P50DA037844; see www.ratgenes.org), behavioral traits that are relevant to drug abuse have been assessed in thousands of male and female HS rats at three institutions throughout the United States. To more fully utilize these rats, we have collected phenotypic data on body weight and length (which permit calculation of body mass index (**BMI**)), fat pad weight and fasting glucose levels. This dataset provides an unprecedented opportunity to map adiposity traits in a large number HS rats.

Methods

Animals

The NMcwi:HS colony, hereafter referred to as HS, was initiated by the NIH in 1984 using the following eight inbred founder strains: ACI/N, BN/SsN, BUF/N, F344/N, M520/N, MR/N, WKY/N and WN/N (Hansen and Spuhler 1984). The breeding colony has been maintained at the Medical College of Wisconsin (MCW) since 2006. The colony for the current studies consisted of 64 breeder pairs maintained using a mating scheme that takes into account the kinship coefficient such that closely related animals are not bred together. For this study, rats from generations 73 to 80 were used. Breeding rats at MCW were given ad libitum access to Teklad 5010 diet (Harlan Laboratories).

The animals used for this study are part of a large multi-centered project focused on genetic analysis of behavioral phenotypes related to drug abuse. HS rats from MCW were sent to three institutions throughout the United States: University of Tennessee Health Science Center (TN), University at Buffalo (NY), and University of Michigan (MI). Rats were shipped at 3-6 weeks of age and each site received multiple shipments over more than two years (from 10/27/2014 – 03/07/2017). Once they arrived, rats in TN were fed Teklad Irradiated LM-485 Mouse/Rat Diet; rats in NY were fed Envigo Teklad 18% Protein Rodent Diet, and rats in MI were fed Labdiet Picolab Laboritory Rodent Diet Irradiated.

Rats were exposed to a different battery of behavioral testing at each site (see **Supplemental Table 1)** followed by euthanasia, which occurred at different ages at each site. All phenotypes presented in this paper were collected at the time of euthanasia. Briefly, in MI, rats were housed in trios, exposed to cocaine for five days (15 mg/kg), and euthanized 4-7 days after cocaine exposure at 89 $\pm$ 6 days of age. In NY, rats were housed in pairs, tested for multiple behaviors over 16 weeks, exposed to cocaine for three days (10 mg/kg) and euthanized 7-10 days after cocaine exposure at 198 $\pm$ 13 days of age. In TN, there were two separate cohorts: breeders (sent from MCW) and experimental rats (bred in TN). Female breeders had mostly one, sometimes two litters but underwent no behavioral testing. The experimental rats were tested for multiple behaviors, exposed to nicotine (self-administration) for 12 days, and euthanized at 73 $\pm$ 12 days of age. The number of rats phenotyped at each site, as well as ages when adiposity phenotypes were collected are shown in **Table 1**). Details of the experimental pipelines are shown in the **Supplemental Table 1**.

Table 1. Age and number of rats at time adiposity phenotypes were collected.

Phenotyping site	Males, N	Females, N	Total N	Age, days, mean ±SD
MI	519	514	**1,033**	89 +/- 6
NY	449	443	**892**	198 +/- 13
TN experiment	458	451	**909**	73 +/- 12
TN breeders	182	157	**339**	169 +/- 34
	1,608	1,565	**3,173**	

Phenotyping and Tissue collection

Several days after completion of behavioral experiments (see above), rats were fasted overnight (17 +/-2 hours), and body weight was measured. Under anesthesia (phentobarbital at MI and NY, isoflurane at TN), two measures of body length (from nose to base of the tail (body length_NoTail) and from nose to end of tail (body length_Tail)) were collected, allowing us to calculate two measures of body mass index: BMI_NoTail and BMI_Tail. BMI was calculated as: (body weight/body length^2) x 10. For animals in MI and NY, we also measured fasting glucose levels using the Contour Next EZ system (Bayer, Elkhart, IN). Several tissues were dissected and weighed including retroperitoneal, epididymal (males), and parametrial (females) visceral fat pads, hereafter referred to as RetroFat, EpiFat, and ParaFat respectively. Spleen was dissected and sent to University of California, San Diego (UCSD) for DNA extraction and genotyping (described below). All protocols were approved by the Institutional Animal Care and Use Committees of each institution.

Genotyping

DNA was extracted from spleen using a salting out procedure or the Agencourt DNAdvance kit (Beckmann). Genotypes were determined using genotyping-by-sequencing (**GBS**), as described elsewhere (Parker et al. 2016). GBS is a reduced representation genotyping method that sequences the small fraction of the genome comprised of restriction fragments of a certain size. Briefly, DNA was digested with endonucleases *PstI* and *NlaIII*, then custom adapters barcoded for individual subjects were ligated to the fragments. The correct ligation product selectively amplified for 12 cycles of PCR, the resulting libraries were quantitated and pooled to make 48 samples per sequencing lane, then size-selected using PippinPrep to retain fragments between

300 and 450 base pairs. Purified, pooled libraries were sequenced at 100 bp (single end), on Illumina 4000 instrument, producing ~7 million reads per sample. The reads were processed to remove barcode and adaptor sequences, filtered for quality, and mapped to the rn6 rat genome assembly. Variant calling was done with ANGSD (Korneliussen et al. 2014), missing genotypes within the sample were imputed with BEAGLE. Imputation to a reference panel was done using IMPUTE2. This protocol produced 3,400,759 SNPs with error rate below 1%. Variants for X- and Y-chromosomes were not called. Prior to GWAS, SNPs in high linkage disequilibrium were removed using function LDprune (REF: https://www.cog-genomics.org/plink/1.9/ld) with r^2 cut-off 0.95; this produced set of 128,447 SNPs which were used for GWAS, genetic correlations, and heritability estimates. The unpruned set of SNPs was used for LocusZoom plots.

Rats were removed from subsequent analysis if they were discordant between recorded and genetically determined sex (n=6) or coat color (n=8), indicating errors in record keeping or a sample mix up.

Phenotypic and genetic correlations and heritability estimates

Each trait within each research site was quantile normalized, separately for males and females, using R function *qnorm*. Relevant covariates (including age, batch number and dissector) were identified for each trait, and covariate effects were regressed out if they were significant and explained more than 2% of the variance (see **Supplemental Table 2**). Residuals were then quantile normalized again, after which the data for each sex and site was pooled prior to further analysis (phenotypic and genetic correlations, heritability and GWAS). Phenotypic correlations were determined using Spearman's test. Genetic correlations were calculated using bivariate GREML analysis as implemented by GTCA (Yang et al. 2011; Lee et al. 2012). GCTA-GREML analysis was used to estimate proportion of variance attributable to SNPs.

Genetic mapping

GWAS analysis employed a linear mixed model, as implemented in the software GEMMA (Wang et al. 2016), using dosages and a genetic relatedness matrix (GRM) to account for the complex family relationships within the HS population. We used the Leave One Chromosome Out (LOCO) method to guard against proximal contamination, as previously described (Cheng et al. 2013; Gonzales et al. 2018). Significance thresholds were calculated using permutation, because all traits were quantile normalized, we were able to use the same threshold for all traits (Cheng and Palmer 2013). To identify QTLs, we scanned each chromosome to determine if there was at least one SNP that exceeded the permutation-derived threshold of $-\log(p) > 5.6$,

that was supported by a second SNP within 0.5 Mb to have a p-value that was within 2 –log(p) of the index SNP. This algorithm failed to identify 2 SNPs, which were then identified by using wider supporting interval (chr10:85082795 for body length_tail and chr12:5782829 for RetroFat). Then correlation between the top SNP and all other SNPs in the 6 Mb vicinity was calculated. The QTL was defined as an interval containing SNPs with r2 > 0.6. Other QTLs on the same chromosome were tested to ensure that they were independent of the first. To establish independence, we used top SNP from the first QTL as a covariate, and performed a second GWAS. If the resulting GWAS had a additional SNP (on the same chromosome) with a p-value that exceeded our permutation-derived threshold, it was considered to be a second, independent locus. This process was repeated (including all previously significant SNPs as covariates), until no more QTLs were detected on a given chromosome.

Linkage disequilibrium (LD) intervals for the identified QTL were determined by identifying those markers that had a high correlation coefficient with the peak marker ($r^2 = 0.6$). Credible set analysis (Wellcome Trust Case Control et al. 2012) was also performed for each locus. The credible set analysis uses a Bayesian approach to calculate the posterior probability for each SNP ("probability of being causal"). This method also chooses the credible set of SNPs, that is the smallest set of SNPs that can account for 99% of the posterior probability.

Sample size and QTL detection

To determine the number of QTLs detected by different samples sizes, we subsampled data from three phenotypes chosen to have low (~0.15; fasting glucose), medium (~0.3; BMI) and high (~0.45; body weight) chip heritabilities. For each dataset, we performed 100 random subsamples in which we retained 500, 1,000, 1,500, 2,000, or 2,500 individuals (for fasting glucose we could not include 2,000 and 2,500 because the total sample size was smaller than 2,000). Thus, we produced 1,300 total subsamples for the three phenotypes. We then performed a GWAS for each subsampled dataset, using covariates and procedures identical to those described in the previous section. We recorded the number of significant QTLs in each subsampled dataset using an algorithm to automatically record the number of QTLs detected. First, we scanned each chromosome to determine if there was at least one SNP that exceeded the threshold of –log(p) > 5.6. To avoid situations where only a single, presumably anomalous, SNP showed a significant association, we required at least one other SNP within 0.5 Mb have a p-value that was within 2 –log(p) of the index SNP. If we found a second supporting SNP, we recorded the identification of a QTL for that dataset. Some chromosomes were expected to contain more than one independent QTL, but we were also concerned that we might count a single locus twice. To avoid counting

the same locus twice, we excluded all SNPs with $r^2 > 0.4$ relative to the just identified index SNP. We then rescanned the chromosome to see if any additional SNPs on this chromosome exceeded the threshold of –log(p) > 5.6, if they did and they were supported by a second SNP within 0.5 Mb to have a p-value that was within 2 –log(p) of the index SNP, we recorded an additional QTL for that dataset. We then repeated these steps as often as needed until no further significant QTLs could be identified on a given chromosome. We then continued this process for all subsequent chromosomes. After scanning the last chromosome, we tabulated the number of QTLs detected for that dataset and then repeated this procedure for each of the 1,300 subsampled datasets. In this way, we determined the number of significant QTLs in 100 possible subsamplings of the each of three traits at when using 500, 1,000, 1,500, 2,000, and 2,500 individuals, and in maximal number of individuals. This algorithm is slightly different from our approach to the main analysis presented in this paper because we removed SNPs with r2>0.4 rather than conditioning on the index SNP.

Results

Strong phenotypic and genetic correlations between multiple adiposity traits

We observed strong phenotypic and genetic correlations between adiposity traits (**Figure 1**). Although less strong, fasting glucose levels were also significantly correlated with body weight, BMI_Tail, BMI_NoTail, RetroFat and EpiFat, but not with body length or ParaFat. Genetic correlations for fasting glucose levels differ from the phenotypic correlations, with a negative correlation seen with body length_NoTail and a positive correlation with ParaFat and BMI_NoTail.

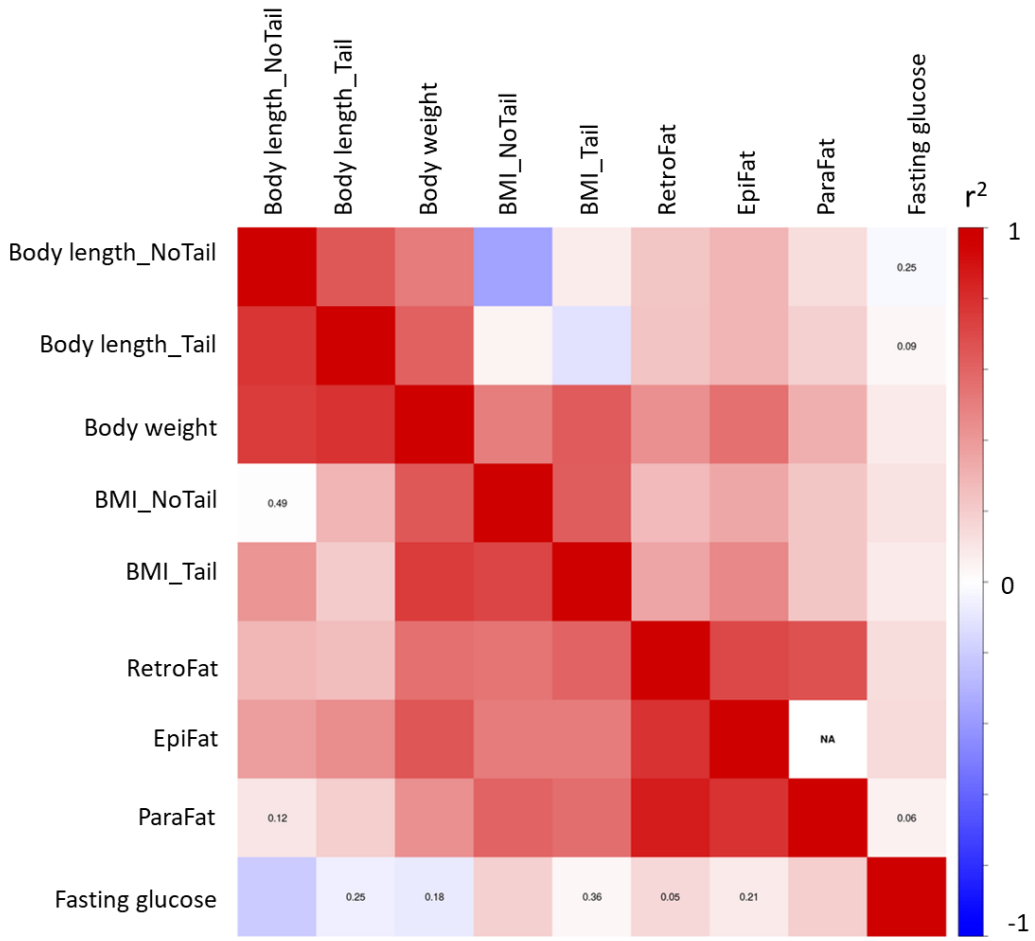


Figure 1. Genetic and phenotypic correlation between adiposity traits and fasting glucose. Phenotypic correlations are depicted in the upper part, genetic – in the lower part of the matrix. Number inside squares show p-value > 0.05.

Adiposity traits exhibit strong heritability

Heritability estimates for adiposity traits ranged from 0.26 $\pm$ 0.03 (BMI_NoTail) to 0.46 $\pm$ 0.03 (body weight; **Table 2**), where heritability estimates for fasting glucose were lower.

Table 2. SNP Heritability estimates

Trait	**Heritability ± SE**
body weight	0.46 ± 0.03
body length_Tail	0.36 ± 0.03
body length_NoTail	0.30 ± 0.03
BMI_Tail	0.31 ± 0.03
BMI_NoTail	0.26 ± 0.03
RetroFat	0.42 ± 0.03
EpiFat	0.37 ± 0.03
ParaFat	0.38 ± 0.04
Fasting Glucose	0.15 ± 0.03

Identification of multiple GWAS hits

We identified a total of 29 independent loci for eight adiposity traits, with four loci mapping to more than one trait (**Figure 2**). We identified two additional loci for fasting glucose. LD intervals were 0.2-9.2 Mb and contained anywhere from 1 to 96 genes. **Table 3** provides a summary of these findings; **Supplementary Table 3** contains all information within **Table 3** with additional information such as the strain distribution pattern of the founder strains as well as the full list of genes within each interval, and credible set analysis results. The genotype and phenotype data needed to reproduce these analyses have been deposited to Gene Network (www.genenetwork.org)

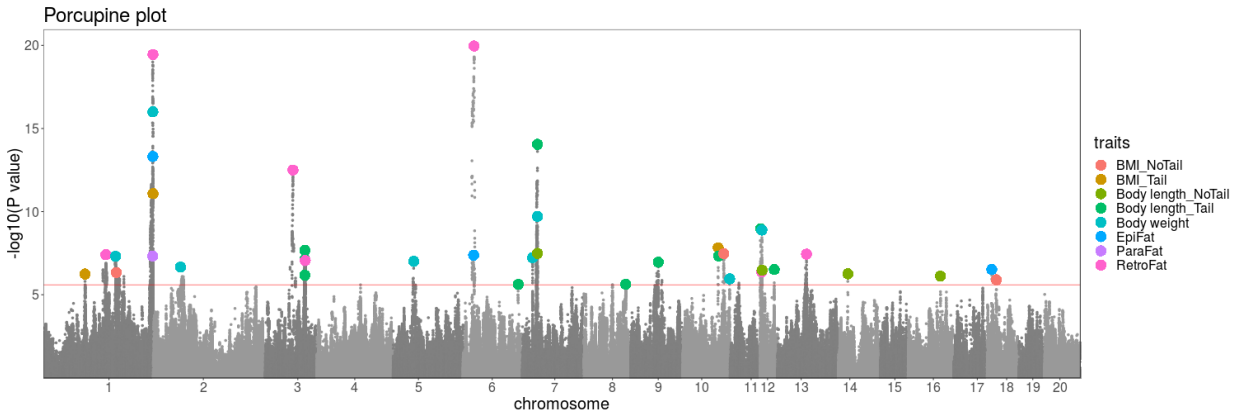


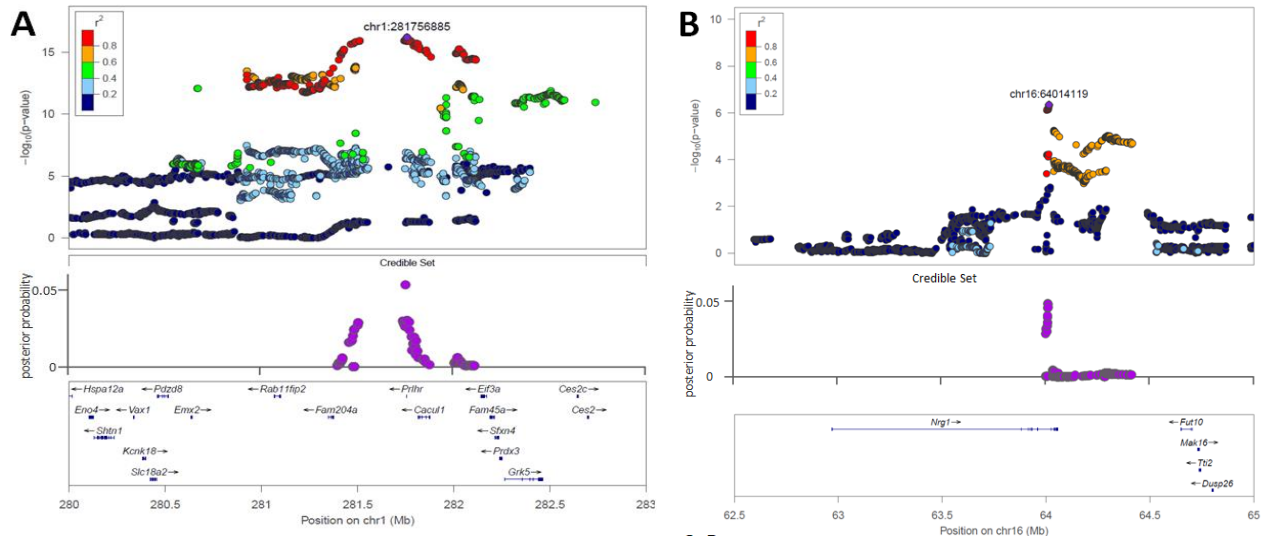
Figure 2. Combined Manhattan plots of GWAS data for eight adipose traits. Genome-wide association results from the GWA analysis. The chromosomal distribution of all the P-values (-log10 P-values) is shown, with top SNPs highlighted.

Pleiotropic Loci: To determine if traits that mapped to the same location are pleiotropic, we considered the mean allele frequency (**MAF**), and the strain distribution pattern of the founder strains. In a few cases, we found loci that overlapped, but the MAF and strains distribution pattern were very different, suggesting different causal loci that happened to map to approximately the same genetic location. We also observed a few cases where the strains distribution pattern was similar but not identical (e.g. one strains' genotype was dissimilar between the two loci), we assumed that those situations were consistent with pleiotropy. Note that this approach is conceptually similar to the estimation of the phenotype associated with founder haplotypes in the diversity outcross, which was first introduced in Sevenson et al (Svenson et al. 2012), but would not perform well if there were more than two causal alleles. Using these criteria, we designated four loci as pleiotropic, which are highlighted in purple in Table 3. The chr 1: 281 Mb is the strongest, with –logP values that range from 7-20 for nearly all of the adiposity traits, including body weight, BMI_Tail, and all three fat pads (RetroFat, EpiFat and ParaFat). Although BMI_Tail maps to the pleitropic locus on chr 1: 281 Mb, BMI_NoTail does not. Interestingly there are other striking differences between body length with and without tail when analyzed on their own and when used to calculate BMI. In addition to the pleiotropic loci on chromosome 1, chr 6: ~27 Mb maps both RetroFat and EpiFat. Finally, chromosomes 3 and 7 map body weight, body length_Tail and body length_NoTail.
Nearby loci that were not considered pleotropic are highlighted in grey in Table 3. We identified two sets of QTLs that mapped to similar regions, and therefore might have appeared to be pleiotropic; however, after applying the criteria described above, we determined that they were not truly pleiotropic because the MAF and strain distribution pattern of the founder strains (see Supplementary Table 3) were not similar. These loci were chr 10: ~110 Mb for fasting glucose and body weight and chr 12: 6 Mb for body weight, RetroFat and body length_NoTail.

Table 3. **Summary of QTLs**

Trait	Peak Marker					LD Interval			Genes in the LD interval
	Position	**-logP**	**Ref Allele**	**Allele frequency**	**Effect size**	**start (bp)**	**stop (bp)**	**size (Mb)**	
BMI_Tail	chr1:106866154	6.05	G	0.29	0.15 ± 0.03	105,730,059	109,396,142	3.67	*7*
RetroFat	chr1:160530456	7.51	T	0.53	-0.14 + 0.02	157,254,290	162,857,247	5.60	*15*
Body Weight	chr1:185730317	7.58	C	0.74	0.16 ± 0.02	184,463,432	187,738,111	3.27	*6*
BMI_NoTail	chr1:187300775	6.72	A	0.66	0.15 ± 0.02	184,772,656	189,346,447	4.57	*27*
ParaFat	chr1:280924549	7.22	G	0.57	0.20 ± 0.03	280,876,316	282,114,080	1.24	*9*
Body Weight	chr1:281756885	16.21	C	0.58	0.21 ± 0.02	280,924,333	282,114,080	1.19	*Fam204a, Prlhr, Cacul1*
RetroFat	chr1:281777218	20.21	A	0.57	0.24 + 0.02	280,924,333	282,114,080	1.19	*9*
EpiFat	chr1:281802657	14.00	C	0.56	0.29 ± 0.03	280,924,333	282,736,277	1.81	*13*
BMI_Tail	chr1:282049439	11.44	C	0.55	0.18 ± 0.02	280,924,333	282,736,277	1.81	*13*
Body Weight	chr2:65816485	6.55	A	0.91	0.20 ± 0.03	62,570,942	71,814,490	9.24	*6*
RetroFat	chr3:95389621	13.12	A	0.44	-0.19 + 0.02	92,336,188	97,685,154	5.35	30
Body Length_NoTail	chr3:136021511	6.92	A	0.77	-0.16 ± 0.03	132,291,573	137,146,532	4.85	*10*
Body Length_Tail	chr3:136021511	5.97	A	0.77	-0.14 ± 0.03	132,291,573	137,146,532	4.85	*10*
Body Weight	chr3:136021511	7.34	A	0.77	-0.16 ± 0.02	132,291,573	137,146,532	4.85	*10*
RetroFat	chr3:137537161	8.21	G	0.44	-0.15 + 0.02	136,161,761	138,849,437	2.69	20
Body Weight	chr5:50933779	6.99	A	0.78	-0.16 ± 0.03	49,152,709	50,940,275	1.79	*12*
EpiFat	chr6:26266960	7.25	G	0.59	-0.20 ± 0.03	22,684,886	28,223,800	5.54	*55*
RetroFat	chr6:28148338	19.50	G	0.65	-0.26 + 0.02	25,954,450	28,752,109	2.80	*58*
Body Length_Tail	chr6:137745191	5.89	C	0.54	-0.12 ± 0.02	136,769,305	138,087,183	1.32	*24*
Body Weight	chr7:24886476	7.29	G	0.07	0.26 ± 0.04	24,869,890	25,213,445	0.34	*Tcp11l2, Nuak1, Ckap4*
Body Weight	chr7:36497588	9.77	C	0.14	-0.24 ± 0.03	34,119,928	36,522,260	2.40	*21*
Body Length_NoTail	chr7:36517726	7.88	T	0.12	-0.25 ± 0.04	34,119,928	36,522,260	2.40	*15*
Body Length_Tail	chr7:36517726	14.37	T	0.12	-0.33 ± 0.04	34,119,928	36,522,260	2.40	*21*
Body Length_Tail	chr8:118711320	5.72	A	0.67	0.13 ± 0.02	116,614,891	119,781,444	3.17	*96*

Body Length_Tail	chr9:65078205	6.81	C	0.76	0.16 ± 0.03	64,013,585	65,670,399	1.66	*20*
BMI_Tail	chr10:84080794	8.21	G	0.57	-0.16 ± 0.02	83,442,285	85,006,252	1.56	*42*
Body Length_Tail	chr10:85082795	7.37	C	0.72	0.17 ± 0.03	84,902,901	85,239,899	0.34	*10*
BMI_NoTail	chr10:96804258	7.66	T	0.48	-0.16 ± 0.02	96,561,667	98,097,621	1.54	*16*
Fasting Glucose	chr10:109944213	6.33	A	0.75	-0.18 ± 0.03	108,350,175	110,315,359	1.97	*49*
Body Weight	chr10:111010289	6.10	T	0.41	0.13 ± 0.02	110,664,101	111,022,246	0.36	*Tbcd, Znf750, B3gntl1, Metrnl*
Body Length_Tail	chr12:2199384	8.94	C	0.18	-0.22 ± 0.03	846,435	5,587,624	4.74	*50*
Body Weight	chr12:5738696	9.12	C	0.59	0.16 ± 0.02	4,938,015	6,078,451	1.14	*6*
RetroFat	chr12:5782829	7.12	T	0.73	0.14 + 0.02	455,837	6,259,634	5.80	59
Body Length_NoTail	chr12:6239515	6.26	A	0.52	0.13 ± 0.02	6,078,237	7,035,028	0.96	*9*
Body Length_Tail	chr12:43060205	6.27	G	0.75	-0.15 ± 0.02	42,024,283	43,190,584	1.17	*Oas1g, Tbx3, Tbx5*
RetroFat	chr13:55021887	7.16	C	0.67	0.14 + 0.02	54,910,667	56,487,438	1.58	*7*
Body Length_NoTail	chr14:26365986	6.54	A	0.73	-0.16 ± 0.03	25,129,356	26,478,660	1.35	*Epha5*
Fasting Glucose	chr14:86029588	6.27	T	0.27	-0.17 ± 0.03	82,473,632	87,575,807	5.10	*34*
Body Length_NoTail	chr16:64014119	6.35	G	0.55	0.13 ± 0.02	64,003,027	64,416,540	0.41	*Nrg1*
EpiFat	chr18:12674871	6.47	G	0.22	-0.22 ± 0.04	12,597,378	12,792,162	0.19	*Klhl14*
BMI_NoTail	chr18:25190274	5.78	T	0.04	-0.34 ± 0.07	23,785,532	25,418,376	1.63	*20*



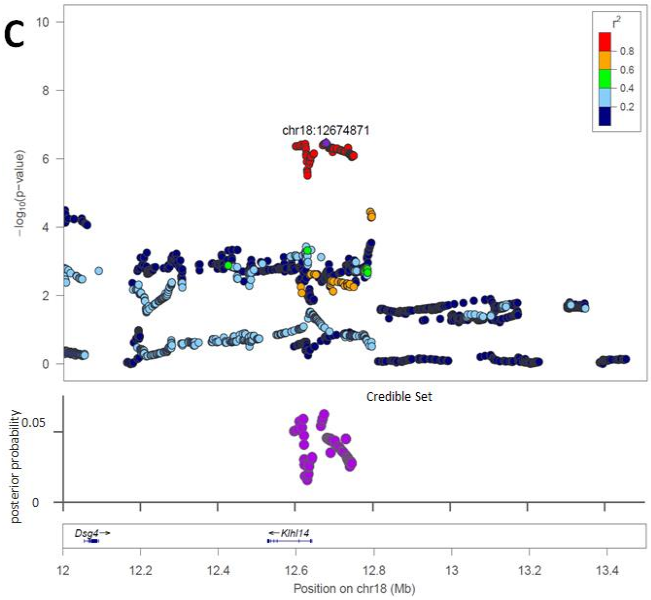


Figure 3. Regional association plots. X-axis shows chromosomal position. Top panel: Regional association plot. The SNPs with the lowest p-value ("top SNP") is shown in purple, and its chromosomal position is indicated on the plot. Correlation of each SNP with the top SNP is shown in color. Middle panel: Credible set track shows posterior probability for being causal for the SNPs which were identified as the smallest set of SNPs accounting for 99% of the posterior probability (see Methods). Bottom panel: Genes in the region, as annotated by the Refseq.

A. Regional association plot for Body weight at Chromosome 1, 280 Mb to 283 Mb. **B.** Regional association plot for Body Length_NoTail at Chromosome 16, 62.5 Mb to 65 Mb. **C.** Regional association plot for Body Length_NoTail at Chromosome 18, 12 Mb to 13.5 Mb.

Candidate gene identification

The number of genes within the identified QTL ranges from 1 to 96 (**Table 3**). There are three regions that contain a single gene: *Epha5* within chr 14: 26 Mb for body length_NoTail, *Nrg1* within chr 16: 64 Mb for body length_NoTail and *Klh14 within* chr18: 12 Mb for EpiFat. Both *Epha5* and *Nrg1* are known to be involved in growth and development and are therefore logical

candidate genes. In contrast, very little is known about *Klhl14*, meaning that if this gene is responsible for the observed QTL it would offer novel biological insight.

Several other loci contained many genes. Within the chr 1: 282 Mb locus that is associated with multiple adiposity traits, however, we have previously identified a coding polymorphism in *Prlhr* as the likely causal variant for Retrofat (Keele et al. 2018) and it is likely that this gene is driving the other traits that map to this highly pleiotropic locus. In addition, previous work has identified multiple genes, including *Adcy3, Krtcap3* and *Slc30a3*, as candidate genes within chr 6: 27 Mb locus for RetroFat and EpiFat. Representative LocusZoom plots for select loci are shown in **Figure 3**, locus zoom plots for all other loci are in **Supplemental Figure 1.**

Exponential increase in the number of QTL detected with increased sample size

In an effort to assess the importance of sample size for QTL discovery, we examined the number of QTLs identified in sub-samples of the full data-set. We found that the number of QTL detected exponentially increases with larger sample size (**Figure 4**). The sub-sampling analysis was conducted in three traits with low, medium and high chip heritabilities: ~0.15 for fasting glucose, ~0.3 for BMI and ~0.45 for body weight, using sample sizes of 500, 1,000, 1,500 for all three traits and including samples sizes of 2,000, and 2,500 for BMI and body weight. We note the largest increase in number of QTL detected for body weight, the trait with the highest heritability, with more than a ten-fold increase in detected QTL when the sample size is increased from 500 to 2500. Similar trends are seen for both BMI and fasting glucose.

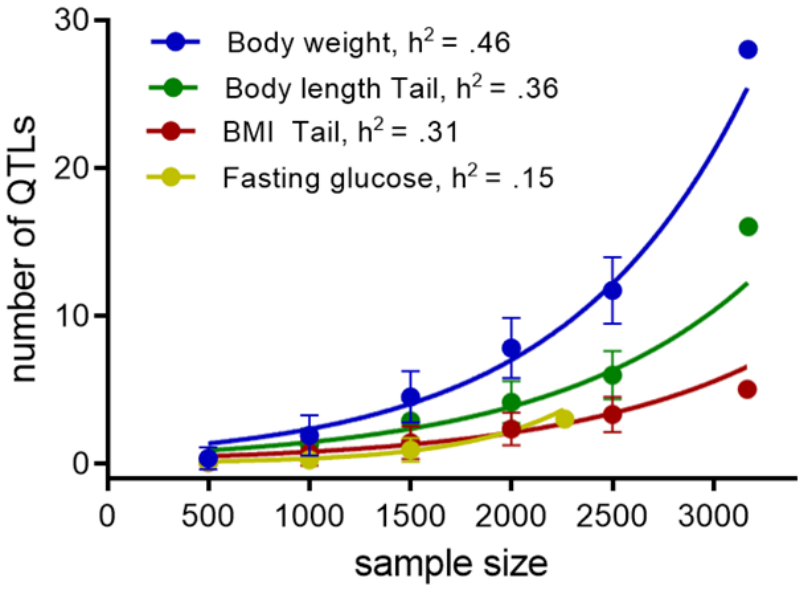


Figure 4. Number of detected QTLs increases with the increase of sample size. Each dot is an average number of QTLs obtained in 100 GWAS, each performed on a randomly selected subset of the actual dataset. Error bars indicate standard deviation. This simulation was performed on four traits with different heritability: body weight (h2 = 0.46 + 0.03), body length_Tail (h2 = 0.36 + 0.03), BMI_Tail (h2 = 0.31 + 0.03) and fasting glucose (h2 = 0.15 + 0.03).

Discussion

The current work is the largest GWAS ever conducted in rodents. This study measured multiple adiposity traits collected at three different sites at multiple ages. Traits measured include body weight, body length (with and without tail), BMI (calculated with and without tail), fat pad weights and fasting glucose. We identified 31 loci, several of which correspond to narrow regions that contained logical candidate genes. We replicated previously identified loci and identified many additional novel loci. The large number of significant associations, the small regions implicated and the replication of previously reported loci despite age, diet and other environmental differences all underscore the power of HS rats for GWAS.

With the exception of fasting glucose, all traits demonstrate strong phenotypic and genetic correlations indicating a common or at least overlapping genetic basis. The pleiotropic loci that we identified further underscore the common genetic basis of these traits. The strongest pleiotropic locus identified falls on chr 1: 282 Mb and maps five of the adiposity traits. Interestingly this locus had previously been identified only for RetroFat in a study using only 742 HS rats (Keele et al. 2018). The locus on chr. 6: 37 Mb maps both RetroFat and EpiFat and was also previously identified using fewer HS rats (Keele et al. 2018). We further identify novel pleiotropic loci for body weight and body length (with and without tail) on chromosomes 3 and 7. Interestingly, our data demonstrate that not all loci that map multiple traits should be considered pleiotropic. We identified two regions that map more than one trait, but because both the mean allele frequency and the founder strain distribution patterns do not match between the traits, these loci are likely driven by different genes and/or variants.

One of the strengths of using a highly recombinant outbred strain such as the HS is the ability to map to relatively small regions of the genome, greatly narrowing the number of potential candidate genes that may drive the QTL. In the current work, three of the identified loci contain only a single gene. For example, *Eph receptor A5 (Epha5)* is the only gene that fall within the chr 14 locus for body length_NoTail. Eph receptors make up the largest sub-family of the receptor protein tyrosine kinases and are known to be involved in embryonic development, cell migration and axon guidance. Although *Epha5* has not previously been associated with body weight or height in human GWAS, *Epha5* knock-out mice have increased body weight relative to wild-type mice (Mamiya et al. 2008), suggesting a potential role for this gene at the QTL. Further work is needed to demonstrate how *Epha5* may be involved in body length of HS rats. In addition, *Neureglin1 (Nrg1)* is the only gene that falls within the chr. 16 locus for body

length_NoTail. *Nrg1* mediates cell-cell signaling and plays a critical role in growth and development of multiple organ systems. ICV injection of *Nrg1* leads to increased food intake and weight gain in rodents (Ennequin et al. 2015) and a recent study has demonstrated positive metabolic effect of *Nrg1* on multiple metabolic parameters, including body weight (Zhang et al. 2018). In addition, *Nrg1* has been associated with BMI in a Korean population (Lee et al. 2016), making *Nrg1* a highly plausible gene within this region. Finally, *Kelch like family member 14 (Klhl14)* is the only gene that falls within the Epifat locus at chr 18. This gene localizes to the endoplasmic reticulum and has not previously been associated with adiposity traits in human or rat GWAS, suggesting that this finding could offer a novel biological insight. Regions containing a single gene are attractive, but we note that these regions may also contain regulatory variants for neighboring genes or un-annotated genes or transcripts that underlie the association with these traits.

In addition to identification of multiple novel loci, the current work replicates two previously identified QTL for fat pad weight in HS rats (Keele et al. 2018): the pleiotropic locus on chr 1: 282 Mb and the locus for RetroFat and EpiFat on chr 6: 28 Mb. In previous work, these loci both mapped to only RetroFat, likely because the previous study employed only 742 HS rats and was therefore significantly underpowered relative to the current study. That study employed multiple genetic and statistical tools to identify candidate genes and likely causal variants that underlie these loci. Within the chromosome 1 locus, a highly conserved, likely damaging variant was identified within the prolactin releasing hormone receptor (*Prlhr*) gene, which is known to play a role in feeding behavior. ICV administration of prolactin into the hypothalamus decreases food intake (Lawrence et al. 2000), and *Prlhr* knock-out mice exhibiting increased food intake, body weight and fat pad weight (Gu et al. 2004). The variant falls within the start codon of the gene and leads to removal of the first 65 amino acids of the protein (Keele et al. 2018). Previous work also identified multiple genes within the chromosome 6 locus that could impact fat pad weight: *Adcy3, Krtcap3* and *Slc30a3*. Our previous work identified a highly conserved, potentially damaging variant in *Adcy3* that is located within the transmembrane pocket that likely alters that way that ADCY3 interacts with other proteins. In addition, using gene expression data from HS liver, we identified *Krtcap3* and *Slc30a3* as likely mediators of fat pad weight at this locus. Although *Adcy3* has been found to play a role in both complex and monogenic forms of obesity (Nordman et al. 2008; Speliotes et al. 2010; Grarup et al. 2018; Saeed et al. 2018) and has recently been shown to interact with MC4R within the hypothalamic neuron to influence weight gain (Siljee et al. 2018), neither *Krtcap3 nor Slc30a3* have previously been associated with adiposity in human studies. This work demonstrates the utility of the HS rat model of

identifying both novel and known genes involved in human adiposity as well as the power of HS rats to uncover multiple causal genes and likely variants underlying a single complex locus. Future work will employ RNAseq in multiple tissues including brain, liver and adipose, to identify potential transcripts whose abundance is influence by these same loci (that is, to identify eQTLs that co-map with and share the same strain distribution pattern with, these phenotypic QTLs).

This is the largest GWAS ever performed HS rats and has identified a similarly large number of significant loci. We subsampled this dataset to demonstrate the well-understood role of sample size on QTL identification. We find an exponential increase in the number of QTL identified with increasing sample size, particularly for body weight, a trait with high heritability. This indicates that many previous studies are likely underpowered and argues for increased sample sizes in future genetic studies of complex traits in outbred rodents. Similar observations in human genetics (Visscher et al. 2012; Sullivan et al. 2018) suggest an initial exponential growth of discoveries, followed by a lineal phase when increasing sample size produces a linear increase in the number of significant loci. Ongoing studies by the NIDA-funded Center for GWAS in outbred rats (P50DA037844; see www.ratgenes.org) will allow us to increase the sample size well above 5,000 rats in the future, which should produce even more significant loci for these traits.

Although this is one of the most well powered genetic studies in a rodent model, there are limitations. Despite the fact that adiposity measures were collected from both males and females, we found that the current study is underpowered to confidently identify sex-specific loci. We are continuing to collect adiposity traits in additional animals and plan to use large sample sizes to identify sex-specific QTLs. A second limitation is that we do not currently have expression QTL information, which would have been helpful for identifying alleles that might alter these traits by modulating gene expression. Because these rats had essentially no opportunities for exercise and also did not have the ability to select different foods, our study would not be expected to identify some of the loci important for human obesity, which likely influence propensity to exercise and food choice, both of which have obvious implications for obesity. Finally, our study examined these traits at only one time point, which differed across the three study sites, which prevented us from exploring the temporal specificity of these QTLs, and may have diluted our power if different loci are important for these traits at the different ages measured in our study.

In conclusion, the current study is the largest GWAS using HS rats ever performed. We replicated previously identified loci from a smaller GWAS and identified numerous novel loci for multiple adiposity traits. Three of these loci contain only a single gene, indicating these are the likely causal genes at these regions. Several other loci contain only a few genes, which simplifies the identification of candidate genes. This work demonstrates the power of HS rats for fine-mapping and gene identification of adiposity traits as well as demonstrates the importance of large sample sizes for genetic dissection of complex traits.

Acknowledgements

P50 DA037844 (AAP), R01 DK106386 (LSW)

References

Baud A, Hermsen R, Guryev V, Stridh P, Graham D, McBride MW, Foroud T, Calderari S, Diez M, Ockinger J et al. 2013. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat Genet* **45**: 767-775.

Cheng R, Palmer AA. 2013. A simulation study of permutation, bootstrap, and gene dropping for assessing statistical significance in the case of unequal relatedness. *Genetics* **193**: 1015-1018.

Cheng R, Parker CC, Abney M, Palmer AA. 2013. Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. *G3 (Bethesda)* **3**: 1861-1867.

Ennequin G, Boisseau N, Caillaud K, Chavanelle V, Etienne M, Li X, Montaurier C, Sirvent P. 2015. Neuregulin 1 affects leptin levels, food intake and weight gain in normal-weight, but not obese, db/db mice. *Diabetes Metab* **41**: 168-172.

Gonzales NM, Seo J, Hernandez-Cordero AI, St. Pierre CL, Gregory JS, Distler MG, Abney M, Canzar S, Lionikas A, Palmer AA. 2018. Genome wide association study of behavioral, physiological and gene expression traits in a multigenerational mouse intercross. *BioRxiv* **in press**.

Grarup N, Moltke I, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, Mahendran Y, Jorsboe E, Larsen CVL, Dahl-Petersen IK et al. 2018. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nat Genet* **50**: 172-174.

Gu W, Geddes BJ, Zhang C, Foley KP, Stricker-Krongrad A. 2004. The prolactin-releasing peptide receptor (GPR10) regulates body weight homeostasis in mice. *J Mol Neurosci* **22**: 93-103.

Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. 2018. Trends in Obesity and Severe Obesity Prevalence in US Youth and Adults by Sex and Age, 2007-2008 to 2015-2016. *JAMA* **319**: 1723-1725.

Hansen C, Spuhler K. 1984. Development of the National Institutes of Health genetically heterogeneous rat stock. *Alcohol Clin Exp Res* **8**: 477-479.

Keele GR, Prokop JW, He H, Holl K, Littrell J, Deal A, Francic S, Cui L, Gatti DM, Broman KW et al. 2018. Genetic Fine-Mapping and Identification of Candidate Genes and Variants for Adiposity Traits in Outbred Rats. *Obesity (Silver Spring)* **26**: 213-222.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**: 356.

Lawrence CB, Celsi F, Brennand J, Luckman SM. 2000. Alternative role for prolactin-releasing peptide in the regulation of food intake. *Nature neuroscience* **3**: 645-646.

Lee M, Kwon DY, Kim MS, Choi CR, Park MY, Kim AJ. 2016. Genome-wide association study for the interaction between BMR and BMI in obese Korean women including overweight. *Nutr Res Pract* **10**: 115-124.

Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. 2012. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**: 2540-2542.

Loos RJ. 2018. The genetics of adiposity. *Curr Opin Genet Dev* **50**: 86-95.

Maes HH, Neale MC, Eaves LJ. 1997. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**: 325-351.

Mamiya PC, Hennesy Z, Zhou R, Wagner GC. 2008. Changes in attack behavior and activity in EphA5 knockout mice. *Brain Res* **1205**: 91-99.

Nordman S, Abulaiti A, Hilding A, Langberg EC, Humphreys K, Ostenson CG, Efendic S, Gu HF. 2008. Genetic variation of the adenylyl cyclase 3 (AC3) locus and its influence on type 2 diabetes and obesity susceptibility in Swedish men. *Int J Obes (Lond)* **32**: 407-412.

Parker CC, Gopalakrishnan S, Carbonetto P, Gonzales NM, Leung E, Park YJ, Aryee E, Davis J, Blizard DA, Ackert-Bicknell CL et al. 2016. Genome-wide association study of

behavioral, physiological and gene expression traits in outbred CFW mice. *Nat Genet* **48**: 919-926.

Saeed S, Bonnefond A, Tamanini F, Mirza MU, Manzoor J, Janjua QM, Din SM, Gaitan J, Milochau A, Durand E et al. 2018. Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nat Genet* **50**: 175-179.

Siljee JE, Wang Y, Bernard AA, Ersoy BA, Zhang S, Marley A, Von Zastrow M, Reiter JF, Vaisse C. 2018. Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. *Nat Genet* **50**: 180-185.

Solberg Woods LC. 2014. QTL mapping in outbred populations: successes and challenges. *Physiol Genomics* **46**: 81-90.

Speliotes EK Willer CJ Berndt SI Monda KL Thorleifsson G Jackson AU Allen HL Lindgren CM Luan Ja Magi R et al. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**: 937-948.

Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Borglum AD, Breen G, Cichon S, Edenberg HJ, Faraone SV, Gelernter J et al. 2018. Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry* **175**: 15-27.

Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, Palmer AA, McMillan L, Churchill GA. 2012. High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics* **190**: 437-447.

Tsaih SW, Holl K, Jia S, Kaldunski M, Tschannen M, He H, Andrae JW, Li SH, Stoddard A, Wiederhold A et al. 2014. Identification of a novel gene for diabetic traits in rats, mice, and humans. *Genetics* **198**: 17-29.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am J Hum Genet* **90**: 7-24.

Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, Zhang J, Dunwell JM, Xu S, Zhang YM. 2016. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep* **6**: 19444.

Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. 2011. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet* **378**: 815-825.

Wellcome Trust Case Control C, Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, Howson JM, Auton A, Myers S et al. 2012. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat Genet* **44**: 1294-1301.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**: 76-82.

Zhang P, Kuang H, He Y, Idiga SO, Li S, Chen Z, Yang Z, Cai X, Zhang K, Potthoff MJ et al. 2018. NRG1-Fc improves metabolic health via dual hepatic and central action. *JCI Insight* **3**.