Supplementary Materials for "Genetic fine-mapping and identification of candidate genes and variants for adiposity traits in outbred rats"

Keele *et al*

Contents

1 Animals

1.1 Housing

Rats were housed two per cage in micro-isolation cages in a conventional facility using autoclaved bedding (sani-chips from PJ Murphy). They had *ad libitum* access to autoclaved Teklad 5010 diet (Harlan Laboratories) and were provided reverse osmosis water chlorinated to 2-3 ppm.

2 Statistical genetic analysis

2.1 Modeling genetic effects on adiposity

All statistical genetic analyses described used the same general model (or approximations to it) for linking the genetics of a given rat to its measured phenotypic outcome. This was the linear mixed effect model (LMM)

$$f(y_i) = \text{covariates}_i + \text{QTL}_i(m) + u_i + \text{residual}_i, \tag{1}$$

where, in brief: $f(y_i)$ is the phenotype subject to a normalizing transformation, specifically, as a conservative measure to rein in high influence data points, we used the rank inverse normal transformation; covariates_i is a fixed effects term that includes variables representing time food deprived, order of tissue harvest, and dissector (notably, dissector significantly affected EpiFat and BMI_Tail_Base); $\text{QTL}_i(m)$ represents the effect of the quantitative trait locus (QTL) at genomic locus m, and is defined in more detail below; and residual_i models the remaining individual-to-individual variation as a normal deviate with variance σ^2. The u_i term is a random polygenic effect representing the effect of overall genetic relatedness, modeled as vector $\mathbf{u} = (u_i, \ldots, u_n)$ drawn from a multivariate normal with covariance matrix $\mathbf{G}\tau^2$, where τ^2 is unknown and $\mathbf{G}$ is the realized genetic relationship matrix, estimated as the pairwise distance in allelic dosages defined by the identity by descent (IBD) probabilities from founder haplotypes, standardized by allele frequency and averaged over loci across the genome, calculated using the `kinship.probs` function in the `DOQTL R` package (Gatti *et al.* 2014). The LMM in Eq 1 with $\text{QTL}_i(m)$ omitted is hereafter referred to as "the null model".

2.2 Heritability estimation

Narrow-sense heritability,

$$h^2 = \frac{\tau^2}{\tau^2 + \sigma^2} \times 100\% \,,$$

was estimated for each phenotype by fitting the null model as a Bayesian LMM using INLA (Rue *et al.* 2009; Holand *et al.* 2013), which gives a complete posterior distribution of h^2, along with point and interval estimates. Phenotypes were scaled to have a mean of 0 and standard deviation of 1, and a uniform prior on h^2 was obtained by setting priors on τ^{-2} and σ^{-2} to $\text{Ga}(1, 1)$, with other settings being default.

2.3 QTL mapping

QTL were identified by genome-wide association of imputed SNPs. This was performed in three steps. First, as in previous work (Solberg Woods *et al.* 2012), we obtained a probabilistic reconstruction of each rat's haplotype mosaic, that is, the configuration of inherited founder haplotypes that compose its genome, using a hidden Markov model (HMM), implemented in `R/qtl2geno` (Broman 2016), applied to the genotype data on HS rats and their founders. This HMM was used to calculate for each individual $i = 1, \ldots, n$, at each marker position $m = 1, \ldots 8218$, a vector of 36 descent probabilities, $\mathbf{p}_{im}$, containing the posterior probability of descent from each of the possible $\frac{8(8+1)}{2} = 36$ haplotype pairs (diplotypes). n, the sample size, varies between phenotypes, with

$n = 989$ for those irrespective of tissue harvest age, such as body weight, and $n = 743$ for those that include only individuals with tissue harvested at 17 weeks of age, such as RetroFat (two rats did not have RetroFat measurements, resulting in $n = 741$). Second, these descent probabilities were used to re-estimate the original SNP genotypes, that is, each $\mathbf{p}_{ij}$ was used to infer a 3-vector of imputed genotype probabilities $\mathbf{g}_{ij}$; these imputed genotypes, which, unlike their raw counterparts, were both complete and relatively robust to genotyping error, were carried forward into subsequent analyses. Third, at each SNP, we fitted the LMM in Eq 1, setting $\mathrm{QTL}_i(m) = \beta x_{mi}$ where x_{mi} is the expectation of the minor allele count (ie, the allele dosage) implied by $\mathbf{g}_{im}$, and β is a fixed effect; comparing the maximum likelihood (ML) fit of this model to that of the null model gave a likelihood ratio test and nominal p-value, reported as its negative base 10 logarithm, or logP. (Note that initially we used models testing the association between phenotype and haplotype descent, ie, $\mathbf{p}_{im}$, directly, as in the region-wide mapping of Solberg Woods *et al.* (2012), due to a combination of uncertainty in haplotype descent and strong imbalances in the estimated haplotype frequencies.)

Genome-wide significance thresholds for logP scores were estimated by parametric bootstrap samples from the fitted null (Valdar *et al.* 2009; Solberg Woods *et al.* 2010), with Bonferroni thresholds, which would be highly conservative due to the serial LD structure, calculated for comparison.

LD intervals for the detected QTL were defined by including neighboring markers that met a set level of LD, measured with the squared correlation coefficient r^2; we used $r^2 = 0.5$ to define intervals based on the plots of the SNP associations overlaid with LD information.

2.4 Fine-mapping through Group-LASSO with fractional resample model averaging

To prioritize SNP variants within the RetroFat chromosome 6 QTL interval, we used the multi-SNP modeling method LLARRMA-dawg (Sabourin *et al.* 2015), which we applied to the imputed SNP genotypes and a population structure-corrected version of the phenotype, namely the phenotypic residuals of the null model. LLARRMA-dawg uses a combination of variable selection and resampling to identify SNPs that have stable, independent associations with the phenotype. Each SNP receives a resample model inclusion probability (RMIP), an estimate of the probability it would be included in a parsimonious multi-SNP model applied to a resampling of the individuals. SNPs with high RMIPs thus represent stronger candidates, and the existence of multiple SNPs with a high RMIP is consistent with the presence of multiple independent signals.

2.5 Estimating diplotype substitution effects at detected QTL

For detected QTL, the effect of substituting alternate diplotypes was estimated using the Diploffect model (Zhang *et al.* 2014), which can help identify interesting alleles of the candidate variants near the mapping signal. Although stability and power, along with the computational demands of a genome-wide analysis, led us to use SNP association for genetic mapping, these were no longer constraints for haplotype effect estimation at an identified QTL. Diploffect is a Bayesian hierarchical approach designed to work with probabilistically inferred haplotype descent, providing shrinkage that mitigates instability from low haplotype frequencies. In addition to the population structure effect in Eq 1, it models two genetic components at the QTL: additive (haplotype) effects, ie the effect of each dose of haplotype (eg WKY); and dominance deviations, those from the additive model for specific combinations of haplotype, (eg, WKY-ACI). Dominance deviations are typically less informed, but their inclusion stabilizes additive effect estimation. Both have their own variance parameters, τ^2_{add} and τ^2_{dom}, with QTL effect size recorded as the intraclass correlation coefficient

$$\rho_{\mathrm{QTL}} = \frac{\tau^2_{\mathrm{QTL}}}{\tau^2_{\mathrm{QTL}} + \tau^2 + \sigma^2} \ ,$$

where $\tau^2_{\text{QTL}} = \tau^2_{\text{add}} + \tau^2_{\text{dom}}$. The model was fitted using 200 importance samples from INLA (Rue *et al.* 2009; Holand *et al.* 2013), with phenotype transformations and variance component priors set as for heritability estimation above.

3 Analysis of RNA-Seq data

Total RNA was extracted from the livers of 398 of the HS rats using Trizol, followed by library preparation using Illumina's TruSeq Stranded mRNA library kit and sequencing on an Illumina HiSeq2500 (Illumina, Inc., San Diego, CA). BN reference genome sequence (genome build Rn6) and GTF files were obtained from Ensembl. RSEM (v1.3.0) `rsem-prepare-reference` function was used to extract the transcript sequences from the genome (Li & Dewey 2011) and to build Bowtie2 indices (Bowtie2 v2.2.8) (Langmead & Salzberg 2012). RSEM `rsem-calculate-expression` function was then used to execute Bowtie2 to align reads of each sample to the transcriptome prepared above and to compute transcript level and gene level expression abundance. Trim Galore (`http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/`) was used to perform quality-based trimming with a cutoff at Q=20. Seven animals were removed due to low number of input reads.

4 Mediation analysis of phenotype, expression, and QTL

Mediation analysis was used to identify genes with expression levels that mediate the relationship between QTL and physiological phenotype. Expression levels of genes contained within the LD-based physiological QTL intervals were assessed as potential candidates as full mediators (intermediates that completely explain the association between SNP and phenotype) and partial mediators (intermediates that explain some of the association between SNP and phenotype). Similar to Baron & Kenny (1986) and adapted for genetic data as in Battle *et al.* (2014), evidence of mediation was assessed by a series of association tests, presented as a series of steps below, evaluating the relationships between previously mapped phenotype QTL (X), some transformation of the expression of level of a candidate mediator gene $j = 1, \ldots, J$ (M), and some transformation of the phenotype (Y).

1. **Potential mediators:** The relationship, represented as an arrow, with directionality encoding causality, $X \rightarrow M$ is evaluated for all J candidate genes in the physiological QTL interval with non-zero expression in greater than 0.25 of the n rats by testing for the association between QTL and expression of gene j via the regression model

$$f(\text{gene.expression}_{ij}) = \text{mapped.QTL}_i + u_i + \text{residual}_i, \tag{2}$$

where briefly $f(\text{gene.expression}_{ij})$ is the expression level for gene j of rat i subject to some normalizing transformation, often a rank inverse normal transformation, mapped.QTL_i is the effect of the mapped QTL for rat i, and u_i and residual_i are respectively the polygenic and individual error terms as described in Eq 1. The maximum likelihood fit of the model in Eq 2 is compared with the null model (same as Eq 2 with mapped.QTL_i omitted) to produce a likelihood ratio statistic and corresponding p-value. The p-values are converted to q-values using the Benjamini-Hocheberg false discovery rate (FDR) method (Benjamini & Hochberg 1995). $X \rightarrow M$ for gene j is considered satisfied if $\text{q-value}_j < 0.1$. A lenient FDR controlling approach to multiple testing is used because the candidate set of genes is constrained to those local to the QTL interval, as well as the mediation analysis including further tests to satisfy mediator status. The set K ($K \leq J$) genes represent candidate mediators, and are also likely co-localizing eQTL to the QTL.

2. **Full mediators:** The relationship $X \perp\!\!\!\perp Y|M$ is representative of M being a full mediator of X on Y, suggesting that $X \rightarrow M \rightarrow Y$, specifically that X does not affect Y outside of through M. The support for this relationship in the data is evaluated by comparing the following regression models:

$$f(y_i) = \text{mapped.QTL}_i + f(\text{gene.expression}_{ij}) + u_i + \text{residual}_i, \tag{3}$$

and

$$f(y_i) = f(\text{gene.expression}_{ij}) + u_i + \text{residual}_i, \tag{4}$$

where Eq 3 is the alternative model and Eq 4 is the null model for a likelihood ratio test. The expression level of gene k is called a full mediator if p-value$_k > 0.05$, representing the situation in which the effect of QTL on the phenotype is fully explained by expression of gene k. After testing for all K candidate mediators, S ($0 \leq S \leq K$) full mediators are called.

3. **Partial mediators:** The relationship $M \rightarrow Y|X$ is representative of M being a partial mediator of X onto Y. To test the support for this relationship, Eq 3 for each candidate partial mediator t ($T = K - S$) is compared to

$$f(y_i) = \text{mapped.QTL}_i + u_i + \text{residual}_i, \tag{5}$$

producing a likelihood ratio statistic and p-value. The FDR controlling approach is used again to obtain corresponding q-values. If q-value$_t < 0.1$, expression of gene t is called a partial mediator of the relationship between the QTL and the phenotype. Gene t could also represent an independent effect on the phenotype from the QTL.

4. **Consistency of effects:** The consistency of the signs of the effect of the relationships of X through the mediator M onto Y ($X \rightarrow M \rightarrow Y$) with X on Y ($X \rightarrow Y$) was checked for all called mediators. $X \overset{+}{\rightarrow} Y$ means that X causally increases Y, whereas $X \overset{-}{\rightarrow} Y$ means that X causally decreases Y. Consistent signs for $X \overset{+}{\rightarrow} Y$ would be $X \overset{+}{\rightarrow} M \overset{+}{\rightarrow} Y$ or $X \overset{-}{\rightarrow} M \overset{-}{\rightarrow} Y$. Similarly, for the $X \overset{-}{\rightarrow} Y$ relationship, consistent mediation relationships would be $X \overset{+}{\rightarrow} M \overset{-}{\rightarrow} Y$ or $X \overset{-}{\rightarrow} M \overset{+}{\rightarrow} Y$. Inconsistent signs, also referred to as paradoxical effects, occur when signs of the relationships to and from the mediator are not consistent with the sign of the relationship from X to Y, suggesting that M potentially acts as a suppressive mediator of the relationship $X \rightarrow Y$.

The validity of the causal inference from the mediation analysis depends on the underlying relationships following a directed acyclic graph (DAG). If cycles are present in the graph, the causal inference will likely not be valid. Cycles cannot exist with $X \rightarrow Y$ and $X \rightarrow M$ because the QTL genotype is essentially fixed and cannot be modulated by other quantities. Notably the assumption is made that $M \rightarrow Y$, and that $M \leftarrow Y$ does not occur, though it is plausible that a QTL could modulate a phenotype ($X \rightarrow Y$), which leads the phenotype to modulate expression of certain genes ($Y \rightarrow M$). These types of relationships would produce significant associations whose causal directionality would be misinterpreted by the mediation analysis, thus their inference is dependent on the assumption.

5 Mediation analysis results

Gene expression data from the liver was measured on 398 of the 989 HS rats (all in the cohort with tissue harvested at 17 weeks of life). The three QTL intervals (RetroFat chromosome 1 and chromosome 6 loci and body weight chromosome 4 locus) were evaluated with mediation analysis in an attempt to identify and prioritize possible candidates that could affect the phenotypes through their expression level variation.

5.1 Body weight chromosome 4 locus

The QTL interval for this locus contained 11 genes (**Table S3**). Three of these had liver expression measured. The main candidate *Grid2* was not sufficiently expressed (non-zero expression proportion < 0.25) in liver tissue. The expression levels of the other two genes (*Ccser1* and *LOC108350839*) were not significantly associated with the QTL ($X \to M$ was not satisfied).

5.2 RetroFat chromosome 1 locus

The QTL interval contained 15 genes (**Table S2**), of which 5 were contained in the expression data (*Emx2*, *Rab11fip2*, *Fam204a*, *Prlhr*, and *Cacul1*). *Emx2* and the primary candidate *Prlhr* were not sufficiently expressed in the liver. Similar to as in body weight, the remaining 3 genes' expression levels were not significantly associated with the QTL.

5.3 RetroFat chromosome 6 locus

The interval for this QTL is much wider than the previous intervals, and contains 130 genes (**Table S1**), of which 114 were measured in the liver expression data. Of the 114, 36 genes had non-zero expression below 0.25, leaving 78 genes for which to evaluate $X \to M$. 14 genes (**Table S4**) had a significant association (q-value < 0.1) between expression levels and the QTL. These 14 candidate mediators were then tested for evidence of being full mediators. *Krtcap3* was called a full mediator (p-value $= 0.15$). The remaining 13 were evaluated as partial mediators, resulting in 5 genes being selected (q-value < 0.1) (**Table S5**). As *Krtcap3* was a strong candidate as a full mediator, we replaced the QTL in the model of RetroFat with it. Each partial mediator was then individually included in a regression model of RetroFat with *Krtcap3* and compared to the null model with only *Krtcap3* (**Table S5**). Only *Slc30a3* remained significant, suggesting that it is the best candidate as an additional regulator of RetroFat, potentially separately from the QTL/*Krtcap3* signal.

6 Figures

6.1 Principal component analysis

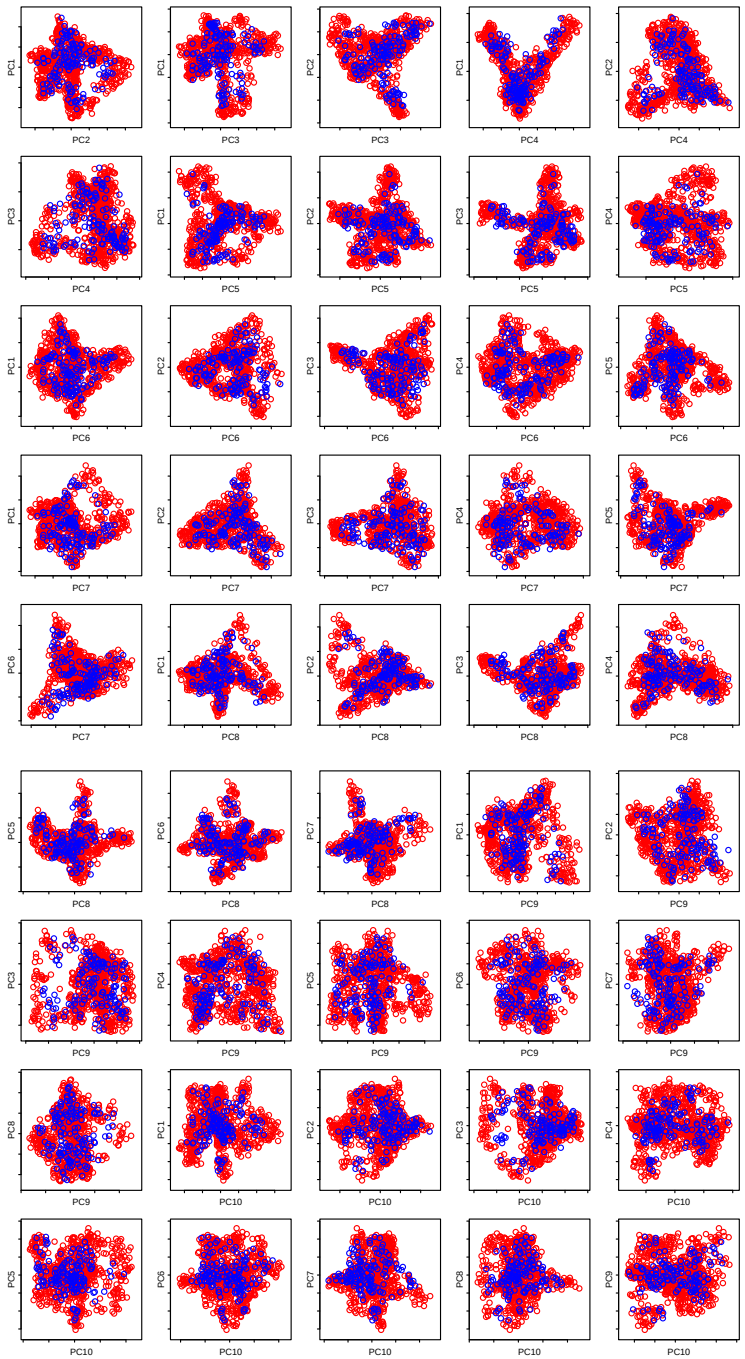


Figure S1: Principal component analysis for the first ten principal components of genotypes between two genotyping centers. Those genotyped at Hudson Alpha are plotted in red and those genotyped at Vanderbilt are plotted in blue. For all plots, red and blue points fall within the same general region indicating that there are no systematic differences in genotype between the two centers.

6.2 LLARRMA-dawg fine-mapping interval

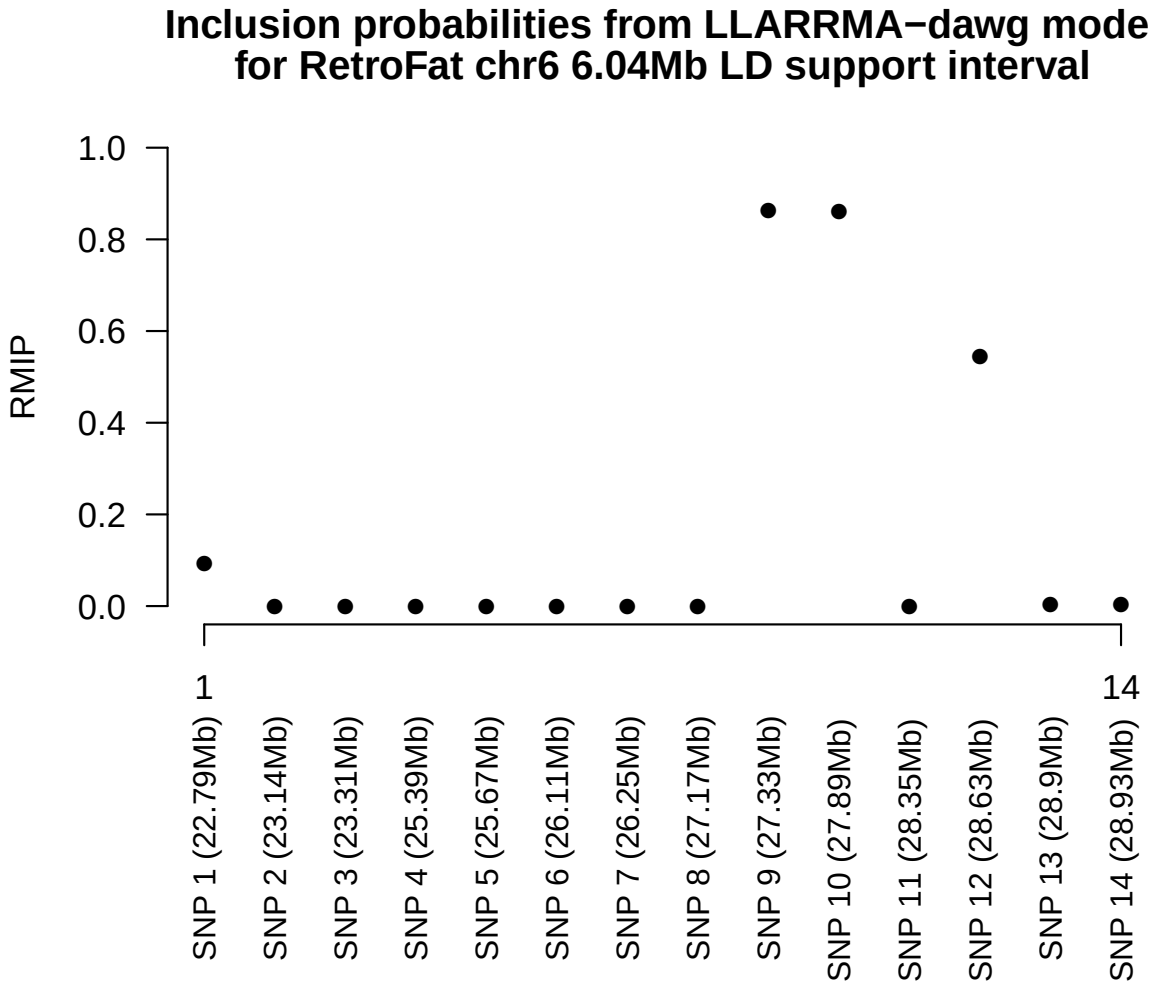


Figure S2: Fine-mapping of the chromosome 6 locus using LLARRMA-dawg reduced the LD support interval from 6.14 Mb to 1.46 Mb. LLARRMA-dawg jointly models and selects SNPs in a region, and returns probabilities corresponding to how often a SNP was included over many re-samples of the data (RMIP). Multiple SNPs with high RMIP suggests the potential for multiple independent signals beneath the QTL peak.

6.3 Genes present in RetroFat chromosome 6 fine-mapping interval

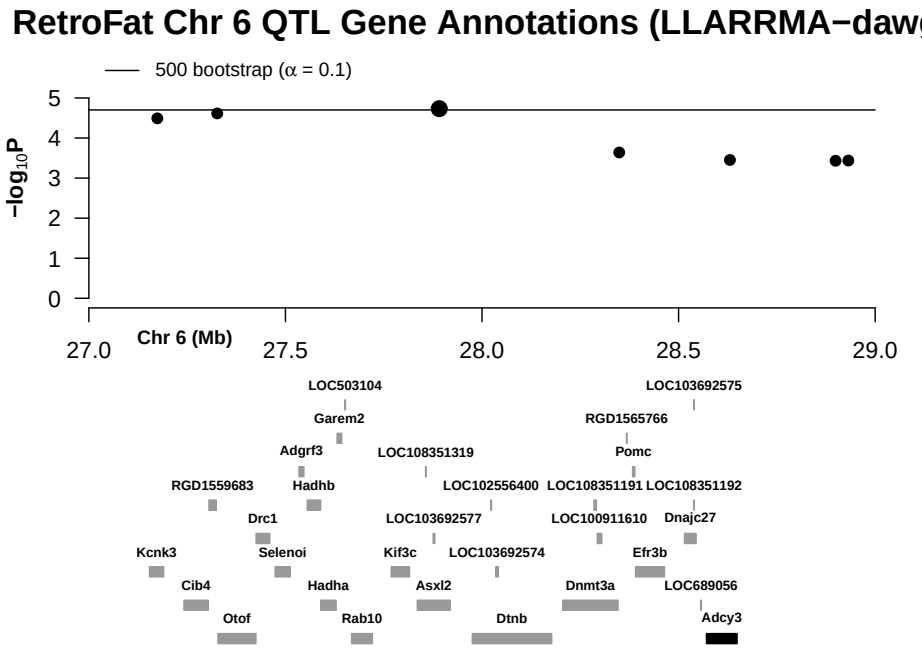
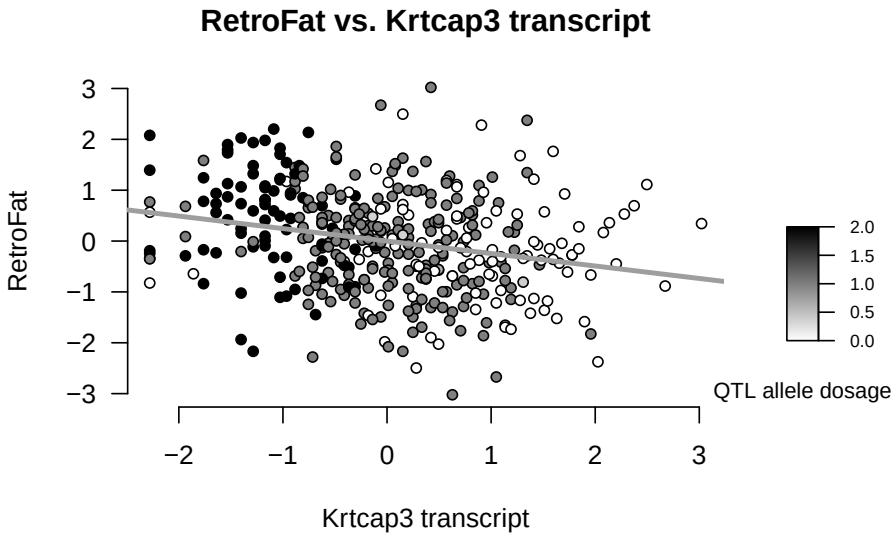
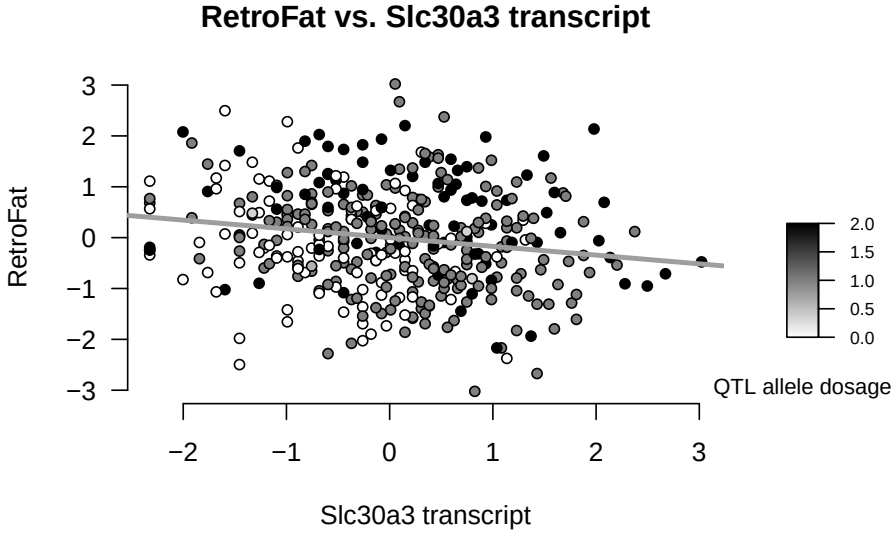


Figure S3: The SNP association (7 markers) present in the LLARRMA-dawg fine-mapping interval (**Figure S2**), including the annotations of the 30 genes local to the region. The candidate gene *Adcy3* is in bold, and possesses a non-synonymous WKY variant that is predicted to alter protein function (**Figure 3E**).

6.4 Candidate mediators: *Krtcap3* and *Slc30a3*



(a) *Krtcap3* expression is negatively correlated with RetroFatg (negative trend line). Genotype dosage of the QTL peak SNP is positively associated with RetroFat, and negtively associated with *Krtcap3* expression, which matches Figure 5.



(b) In contrast to *Krtcap3*, the peak SNP minor allele dosage is positive associated with *Slc30a3*, although its expression is negatively correlated with RetroFatg (negative trend line). The mediation path through *Slc30a3* is inconsistent with the QTL relationship with RetroFat, suggesting that *Slc30a3* may actually act in a suppressive manner with respect to the QTL effect.

Figure S4: Scatterplot of RetroFat and mediator expression levels, with data points colored by the peak SNP minor allele dosages at the QTL. RetroFat and expression levels are rank-inverse normal transformed.

7 Tables

7.1 Genes in RetroFat chromosome 6 QTL interval

Table S1: **Genes in RetroFat chromosome 6 QTL interval.**

Gene Symbol	Gene Name	Start Location	Non-synonymous variants in WKY founder†	Polyphen prediction
Alk	ALK receptor tyrosine kinase	22696415		
LOC108351180	uncharacterized	22988727		
LOC108351181	uncharacterized	23205628		
Clip4	CAP-GLY domain containing linker protein family, member 4	23222020		
LOC103692578	basic proline-rich protein-like	23298261		
RGD1304963	similar to hypothetical protein MGC38716	23337507		
Togaram2	TOG array regulator of axonemal microtubules 2	23358762		
Wdr43*	WD repeat domain 43	23433532		
Trnac-gca30	transfer RNA cysteine (anticodon GCA) 30	23487063		
LOC102551341	tRNA (adenine(58)-N(1))-methyltransferase, mitochondrial-like	23487545		
Spdya	speedy/RINGO cell cycle regulator family member A	23493686	23495595	unknown
LOC102548558	protein tyrosine phosphatase type IVA 1-like	23493704		
Ppp1cb*	protein phosphatase 1 catalytic subunit beta	23548507		
LOC108351182	ALK tyrosine kinase receptor-like	23725713		
LOC298795	similar to 14-3-3 protein sigma	23757225		
LOC108351183	uncharacterized	23771355		
LOC103692567	uncharacterized	23885316		
LOC108351327	glyceraldehyde-3-phosphate dehydrogenase pseudogene	23936327		
LOC103692568	uncharacterized	23986197		
LOC102553396	uncharacterized	24064737		
Ypel5	yippee-like 5	24069351		
Lbh	limb bud and heart development	24154207		
LOC108351184	uncharacterized	24192828		
LOC108351185	uncharacterized	24256909		
LOC102547591	uncharacterized	24336223		
LOC100912066	uncharacterized	24342924		
Lclat1	lysocardiolipin acyltransferase 1	24377398		
LOC102547438	uncharacterized	24527464		
LOC685881	hypothetical protein	24562761		
Capn13	calpain 13	24579590		
LOC102554046	uncharacterized	24623564		
LOC102553955	uncharacterized	24657682		
Galnt14	polypeptide N-acetylgalactosaminyltransferase 14	24770308		
Ehd3	EH-domain containing 3	25076012		
LOC102554201	uncharacterized	25101552		
Xdh	xanthine dehydrogenase	25149570		
LOC100363233	splicing factor 3b, subunit 4-like	25226245		
Srd5a2	steroid 5 alpha-reductase 2	25279635		

Table S1: **Genes in RetroFat chromosome 6 QTL interval (continued).**

Gene Symbol	Gene Name	Start Location	Non-synonymous variants in WKY founder†	Polyphen prediction
Plb1	phospholipase B1	25375699		
LOC683819	hypothetical protein	25565221		
Fosl2	FOS like 2, AP-1 transcription factor subunit	25598936		
Babam2	BRISC and BRCA1 A complex member 2	25666654		
LOC103692569	uncharacterized	25885973		
Rbks	ribokinase	26051568	26072561 T to A	benign
Mrpl33	mitochondrial ribosomal protein L33	26130278		
LOC102548914	uncharacterized	26201017		
Slc4a1ap	solute carrier family 4 member 1 adaptor protein	26214083		
Supt7l	SPT7-like STAGA complex gamma subunit	26241672		
Gpn1*	GPN-loop GTPase 1	26255081		
RGD1560110	similar to RIKEN cDNA 4930548H24	26278440		
Zfp512	zinc finger protein 512	26284749		
LOC102556504	titin-like	26322470		
Gckr	glucokinase regulator	26355296		
LOC100910821	uncharacterized	26387284		
Ift172	intraflagellar transport 172	26390686		
LOC108351187	uncharacterized	26407404		
LOC108351186	60S ribosomal protein L37 pseudogene	26415619		
LOC103692570	dihydropyrimidinase-related protein 5-like	26423841		
Krtcap3*	keratinocyte associated protein 3	26485126		
Nrbp1	nuclear receptor binding protein 1	26486823		
Ppm1g	protein phosphatase, Mg2+/Mn2+ dependent, 1G	26517840		
Zfp513	zinc finger protein 513	26537707		
Snx17	sorting nexin 17	26541137		
Eif2b4	eukaryotic translation initiation factor 2B subunit delta	26546917		
Gtf3c2	general transcription factor IIIC subunit 2	26560601	26581578 T to C	unknown
Mpv17	MpV17 mitochondrial inner membrane protein	26585713		
Ucn	urocortin	26602144		
Trim54	tripartite motif-containing 54	26603364		
Dnajc5g	DnaJ heat shock protein family (Hsp40) member C5 gamma	26625526		
Slc30a3*	solute carrier family 30 member 3	26629752		
Cad	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	26657507		
Atraid*	all-trans retinoic acid-induced differentiation factor	26680628		
Slc5a6	solute carrier family 5 member 6	26685823		
Tcf23	transcription factor 23	26763159		

Table S1: **Genes in RetroFat chromosome 6 QTL interval (continued).**

Gene Symbol	Gene Name	Start Location	Non-synonymous variants in WKY founder†	Polyphen prediction
Prr30	proline rich 30	26780352		
Preb	prolactin regulatory element binding	26784088	26786379 A to G	benign
Abhd1	abhydrolase domain containing 1	26787807		
Cgref1	cell growth regulator with EF hand domain 1	26797126		
Khk	ketohexokinase	26810577		
Emilin1	elastin microfibril interfacer 1	26821249		
LOC103692571	uncharacterized	26833107		
Ost4	oligosaccharyltransferase complex subunit 4, non-catalytic	26836216		
Agbl5	ATP/GTP binding protein-like 5	26837299		
Trnaa-agc6	transfer RNA alanine (anticodon AGC) 6	26856068		
Trnay-gua	transfer RNA tyrosine (anticodon GUA)	26856459		
Trnay-gua3	transfer RNA tyrosine (anticodon GUA) 3	26856459		
Tmem214	transmembrane protein 214	26867638		
Mapre3	microtubule-associated protein, RP/EB family, member 3	26878738		
LOC108351190	uncharacterized	26890051		
LOC108351189	uncharacterized	26918219		
LOC108351188	60S ribosomal protein L37 pseudogene	26931127		
Dpysl5	dihydropyrimidinase-like 5	26939696		
LOC103692572	uncharacterized	27069013		
Cenpa	centromere protein A	27072259		
Slc35f6	solute carrier family 35, member F6	27095144		
LOC103692573	uncharacterized	27139210		
Kcnk3	potassium two pore domain channel subfamily K member 3	27154274		
Cib4	calcium and integrin binding family member 4	27241804		
RGD1559683	similar to RIKEN cDNA 1700001C02	27305402		
Otof	otoferlin	27328343		
Drc1	dynein regulatory complex subunit 1	27425237	27428501 G to A	benign
Selenoi	selenoprotein I	27473748		
Adgrf3	adhesion G protein-coupled receptor F3	27534525		
Hadhb	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), beta subunit	27555408		
Hadha	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit	27589840		
Garem2	GRB2 associated regulator of MAPK1 subtype 2	27631364		
LOC503104	similar to retinoblastoma binding protein 4	27651115		
Rab10	RAB10, member RAS oncogene family	27668387		

Table S1: **Genes in RetroFat chromosome 6 QTL interval (continued).**

Gene Symbol	Gene Name	Start Location	Non-synonymous variants in WKY founder†	Polyphen prediction
Kif3c	kinesin family member 3C	27768943		
Asxl2	additional sex combs like 2, transcriptional regulator	27835346		
LOC108351319	28S ribosomal protein S21, mitochondrial pseudogene	27856408		
LOC103692577	RNA pseudouridylate synthase domain-containing protein 4 pseudogene	27875868		
Dtnb	dystrobrevin, beta	27975302	28004664 A to G	benign
LOC102556400	transcription factor BTF3-like	28022498		
LOC103692574	uncharacterized	28034953		
Dnmt3a	DNA methyltransferase 3 alpha	28205375		
LOC108351191	60S ribosomal protein L37 pseudogene	28284681		
LOC100911610	dihydropyrimidinase-related protein 5-like	28293250		
RGD1565766	hypothetical gene supported by BC088468; NM_001009712	28367389		
Pomc	proopiomelanocortin	28382937		
Efr3b	EFR3 homolog B	28390541		
Dnajc27	DnaJ heat shock protein family (Hsp40) member C27	28515054		
LOC108351192	cytochrome c oxidase subunit 7B, mitochondrial pseudogene	28539158		
LOC103692575	cytochrome c oxidase subunit 7B, mitochondrial pseudogene	28539172		
LOC689056	similar to general transcription factor IIH, polypeptide 5	28556618		
Adcy3	adenylate cyclase 3	28570941	28572363 A to C	damaging
Cenpo	centromere protein O	28648804		
Ptrhd1	peptidyl-tRNA hydrolase domain containing 1	28663602		
Ncoa1	nuclear receptor coactivator 1	28677563		
LOC103692576	uncharacterized	28812571		

Genes in bold are found within the most likely region of the QTL based on multi-SNP fine-mapping analysis.

*Full or partial mediators of RetroFat called by mediation analysis.

†RetroFat chromosome 6 haplotype effects: WKY has decreased fat pad weight (Figure 3D).

7.2 Genes in RetroFat chromosome 1 QTL interval

Table S2: **Genes in RetroFat chromosome 1 QTL interval.**

Gene Symbol	Gene Name	Start location	Gene Function (UniProt)	Non-synonymous variants in founders with the haplotype effect†
LOC103691392	uncharacterized	280573647	Unknown	
Emx2	empty spiracles homeobox 2	280633938	Transcription factor which acts to generate the boundary between the roof and archipallium in the developing brain.	
LOC108349711	uncharacterized	280653842	Unknown	
LOC502394	hypothetical	280753676	Unknown	
LOC102555781	uncharacterized	280796426	Unknown	
LOC108349712	uncharacterized	280934585	Unknown	
Rab11fip2	RAB11 family interacting protein 2	281065346	A Rab11 effector binding preferentially phosphatidylinositol 3,4,5-trisphosphate (PtdInsP3) and phosphatidic acid (PA) and acting in the regulation of the transport of vesicles from the endosomal recycling compartment (ERC) to the plasma membrane. Involved in insulin granule exocytosis. Also involved in receptor-mediated endocytosis and membrane trafficking of recycling endosomes, probably originating from clathrin-coated vesicles.	
LOC102556164	uncharacterized	281227923	Unknown	
LOC102556108	uncharacterized	281289720	Unknown	
LOC102556023	acyl carrier protein, mitochondrial-like	281304776	Unknown	
Fam204a	family with sequence similarity 204, member A	281343692	Unknown	

Table S2: **Genes in RetroFat chromosome 1 QTL interval (continued).**

Gene Symbol	Gene Name	Start location	Gene Function (UniProt)	Non-synonymous variants in founders with the haplotype effect†
LOC103691393	uncharacterized	281395030	Unknown	
LOC108349713	uncharacterized	281397476	Unknown	
Prlhr	prolactin releasing hormone receptor	281754472	Receptor for prolactin-releasing peptide (PrRP). Implicated in lactation, **regulation of food intake** and pain-signal processing.	281755911 C to T translation start site in BUF and WKY
Cacul1	CDK2-associated, cullin domain 1	281814226	Cell cycle associated protein capable of promoting cell proliferation through the activation of CDK2 at the G1/S phase transition.	

The gene in bold (Prlhr) is the most likely candidate in the region.

†RetroFat chromosome 1 haplotype effects: BUF, MR, WKY haplotypes lead to increased fat pad weight (Figure 6D).

7.3 Genes in body weight chromosome 4 QTL interval

Table S3: **Genes in body weight chromosome 4 QTL interval.**

Gene Symbol	Gene Name	Start location	Gene Function (UniProt)	Non-synonymous variants in founders with the haplotype effect†
Ccser1	coiled-coil serine-rich protein 1	91235885	Unknown, has been associated with cocaine	None
LOC103692146	uncharacterized	91601766	Unknown	None
LOC108350840	uncharacterized	91959690	Unknown	None
LOC108350839	high mobility group protein B1-like	92431517	Unknown	None
LOC103692148	developmental pluripotency-associated protein 2 pseudogene	92443293	Unknown	None
LOC103692149	axoneme-associated protein mst101(2)-like	92501663	Unknown	None
LOC103692147	glutamate receptor ionotropic, delta-2-like	93012791	Unknown	None
Hint1-ps1	histidine triad nucleotide binding protein 1, pseudogene 1	93405665	Pseudogene, likely not functional	None
LOC103692150	thyrotropin receptor pseudogene	93447412	Unknown	None
LOC108350813	Ig kappa chain V-II region 26-10-like	93857773	Unknown	None
Grid2	glutamate ionotropic receptor delta type subunit 2	94068112	Receptor for glutamate. L-glutamate acts as an excitatory neurotransmitter at many synapses in the central nervous system.	None

The gene in bold (Grid2) is the most likely candidate in the region.

†Body Weight chromosome 4 haplotype effects: ACI, BUF, F344 and MR haplotypes lead to decreased body weight while BN haplotype leads to increased body weight (Figure 7D).

7.4 Potential mediators in RetroFat chromosome 6 QTL interval

Table S4: **Genes in RetroFat chromosome 6 QTL interval that support X → M relationship.**

Gene Symbol	Gene Name	Start location	q-value
RGD1304963	similar to hypothetical protein MGC38716	23337507	4.79E-05
Wdr43	WD repeat domain 43	23433532	7.82E-02
Ppp1cb	protein phosphatase 1 catalytic subunit beta	23548507	7.46E-03
Galnt14	polypeptide N-acetylgalactosaminyltransferase 14	24770308	3.09E-02
Rbks	ribokinase	26051568	3.13E-11
Gpn1	GPN-loop GTPase 1	26255081	3.38E-04
Krtcap3	keratinocyte associated protein 3	26485126	3.30E-41
Slc30a3	solute carrier family 30 member 3	26629752	1.19E-07
Atraid	all-trans retinoic acid-induced differentiation factor	26680628	7.46E-03
Dpysl5	dihydropyrimidinase-like 5	26939696	9.66E-14
Hadha	hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit	27589840	3.96E-03

7.5 Candidate mediators of RetroFat chromosome 6 QTL

Table S5: **Genes in RetroFat chromosome 6 QTL interval that mediation analysis supports as candidate mediators of the effect of QTL on RetroFat.**

Gene Symbol	Gene Name	Start location	Full mediation p-value	Joint with *Krtcap3* p-value	Consistency with QTL effect
Wdr43	WD repeat domain 43	23433532	1.05E-04	0.47	Inconsistent
Ppp1cb	protein phosphatase 1 catalytic subunit beta	23548507	7.36E-05	0.28	Inconsistent
Gpn1	GPN-loop GTPase 1	26255081	1.10E-03	0.48	Consistent
Krtcap3	keratinocyte associated protein 3	26485126	0.15	.	Consistent
Slc30a3	solute carrier family 30 member 3	26629752	8.40E-07	2.36E-03	Inconsistent
Atraid	all-trans retinoic acid-induced differentiation factor	26680628	6.09E-05	0.89	Inconsistent

Genes in bold are called as mediators.

References

Baron, Reuben M., & Kenny, David A. 1986. The Moderator-Mediator Variable Distinction in Social The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, **51**(6), 1173–1182.

Battle, Alexis, Mostafavi, Sara, Zhu, Xiaowei, Potash, James B., Weissman, Myrna M., McCormick, Courtney, Haudenschild, Christian D., Beckman, Kenneth B., Shi, Jianxin, Mei, Rui, Urban, Alexander E., Montgomery, Stephen B., Levinson, Douglas F., & Koller, Daphne. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Research*, **24**(1), 14–24.

Benjamini, Yoav, & Hochberg, Yosef. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B*, **57**(1), 289–300.

Broman, Karl W. 2016. *qtl2geno: Treatment of Marker Genotypes for QTL Experiments*. R package version 0.4-23.

Gatti, D. M., Svenson, K. L., Shabalin, A., Wu, L.-Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., Palmer, A., Chesler, E. J., Broman, K. W., & Churchill, G. A. 2014. Quantitative Trait Locus Mapping Methods for Diversity Outbred Mice. *G3*, **4**(9), 1623–1633.

Holand, Anna Marie, Steinsland, Ingelin, Martino, Sara, & Jensen, Henrik. 2013. Animal models and integrated nested Laplace approximations. *G3*, **3**(8), 1241–51.

Langmead, Ben, & Salzberg, Steven L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**(4), 357–9.

Li, Bo, & Dewey, Colin N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**(1), 323.

Rue, Håvard, Martino, Sara, & Nicolas, Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, **71**(2), 319–392.

Sabourin, Jeremy, Nobel, Andrew B., & Valdar, William. 2015. Fine-Mapping Additive and Dominant SNP Effects Using Group-LASSO and Fractional Resample Model Averaging. *Genetic Epidemiology*, **39**(2), 77–88.

Solberg Woods, Leah C, Holl, Katie, Tschannen, Michael, & Valdar, William. 2010. Fine-mapping a locus for glucose tolerance using heterogeneous stock rats. *Physiological genomics*, **41**(1), 102–8.

Solberg Woods, Leah C, Holl, Katie L, Oreper, Daniel, Xie, Yuying, Tsaih, Shirng-Wern, & Valdar, William. 2012. Fine-mapping diabetes-related traits, including insulin resistance, in heterogeneous stock rats. *Physiological genomics*, **44**(21), 1013–26.

Valdar, William, Holmes, Christopher C, Mott, Richard, & Flint, Jonathan. 2009. Mapping in structured populations by resample model averaging. *Genetics*, **182**(4), 1263–77.

Zhang, Zhaojun, Wang, Wei, & Valdar, William. 2014. Bayesian modeling of haplotype effects in multiparent populations. *Genetics*, **198**(1), 139–56.