MAPPING THE EPISTATIC NETWORK UNDERLYING MURINE
REPRODUCTIVE FATPAD VARIATION

Joseph P. Jarvis¹,² and James M. Cheverud¹

¹ Department of Anatomy & Neurobiology, Washington University School of Medicine, St. Louis, Missouri, 63110
² Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania, 19104
Running Title: Murine Fatpad Epistasis

Keywords: Genetic Architecture, Composite Interval Mapping, Pair-wise Interactions, Epistatic Positional Candidates

Address Correspondence to:
Joseph P. Jarvis Ph.D.
Tishkoff Lab - Dept of Genetics
University of Pennsylvania
422 Clinical Research Building
415 Curie Blvd
Philadelphia, PA 19104-6145
215-746-2322
E-mail: jarvisj@mail.med.upenn.edu
Genome-wide mapping analyses are now commonplace in many species and several networks of interacting loci have been reported. However, relatively few details regarding epistatic interactions and their contribution to complex trait variation in multicellular organisms are available and the identification of positional candidate loci for epistatic QTL (epiQTL) is hampered, especially in mammals, by the limited genetic resolution inherent in most study designs. Here we further investigate the genetic architecture of reproductive fatpad weight in mice using the F_{10} generation of the LG,SM Advanced Intercross (AI) line. We apply multiple mapping techniques including a single-locus model, locus-specific composite interval mapping, and tests for multiple QTL per chromosome to the twelve chromosomes known to harbor single locus QTL (slQTL) affecting obesity in this cross. We also perform a genome-wide scan for pairwise epistasis. Using this combination of approaches we detect 199 peaks spread over all 19 autosomes that potentially contribute to trait variation including all eight original F_{2} loci (Adip1-8), novel slQTL peaks on chromosomes 7 and 9, and several novel epistatic loci. Extensive epistasis is confirmed involving both slQTL confidence intervals as well as regions that show no significant additive or dominance effects. These results provide important new insights into mapping complex genetic architectures and the role of epistasis in complex trait variation.
INTRODUCTION

The development and elaboration of techniques such as interval mapping (Lander and Botstein 1989), composite interval mapping (Zeng 1994), and methods based on complex pedigree structures (Jannink et al. 2001) has produced an extensive repertoire for the statistical exploration of genotype-phenotype relationships, especially for single loci. Using these approaches genome-wide analyses have identified single-locus QTL (slQTL) underlying variance in characters as varied as agronomic traits and pest resistance in corn (Papst et al. 2004), life span in fruit flies (Wilson et al. 2006), alkylator induced cancer susceptibility in mice (Fenske et al. 2006), murine skeletal morphology (Kenney-Hunt et al. 2008), and an ever expanding list of human diseases and disorders including Age-Related Macular Degeneration (e.g. Klein et al. 2005), Type 2 diabetes (e.g. Sladek et al. 2007; Zeggini et al. 2008), and Crohn’s disease (e.g. Duerr et al. 2006).

In addition, several studies have successfully employed epistatic QTL (epiQTL) mapping strategies to describe multi-locus networks (e.g. Cheverud et al. 2001; Stylianou et al. 2006; Wentzell et al. 2007; Fawcett et al. 2008, Fawcett et al. 2010).

However, most mapping studies in model systems involve either F2 intercross populations or Recombinant Inbred (RI) strain panels (see also Hanlon et al. 2006). These populations harbor limited recombination and so tend to identify relatively large confidence intervals, complicating the physiological investigation of statistical results. Furthermore, while recombinant Inbred (RI) strain sets represent a four-fold expansion of the F2 recombination-based map, they require a minimum of 20 generations of brother-sister mating (Silver 1995) and the number of strains per set is usually low, especially in mammals. Conversely, the production of Advanced Intercross (AI) lines involves many
generations of outbreeding in a relatively large population. This preserves heterozygosity, retains many more recombinant gametes in the gene pool, decreases the average size of segregating linkage blocks, and increases mapping resolution (Haldane and Waddington 1931; Bartlett and Haldane 1935; Hanson 1959a; Hanson 1959b; Hanson 1959c; Hanson 1959d; Darvasi and Soller 1995; Rockman and Kruglyak 2008).

Specifically, the F_{10} generation of a murine AI line represents an approximately five-fold expansion of the F_2 map and thus an improvement in resolution over both F_2 intercross and RI line study designs.

Obesity and related phenotypes are among the most intensively studied complex traits in mice and the LG,SM AI has proven particularly useful in the identification of adiposity QTLs. Previous work in this cross has characterized over 70 loci contributing to variance in fatpad weight, body weight and relevant organ weights (Cheverud et al. 1999; Cheverud et al. 2001; Cheverud et al. 2004; Fawcett et al. 2008). In addition, a recent study used the combined F_9 and F_{10} generations (Fawcett et al. 2010) to fine-map loci for a suite of obesity related characters and achieved an average confidence interval for fatpad loci of 4.14 Mb. These CI were subsequently tested for epistasis and extensive interaction was confirmed, though several direct effect loci identified in the F_2 and F_{2/3} generations failed to replicate and were thus not included. However, in a full genome-wide scan for pair-wise epistasis in the F_2 generation of this cross (Jarvis and Cheverud 2009) 38 fatpad loci that were not identified using a single locus mapping model show significant epistatic interactions. Consistent with results from other experimental systems (reviewed in Phillips 2008) this suggests that many biologically relevant loci are invisible to single locus scans. Thus, combining the increased genetic resolution of an F_{10} AI line
study, with the full range of single-locus and epistatic mapping strategies promises to produce novel insights into the contribution of epistatic interactions to variation in reproductive fatpad weight in mice. Furthermore, the accumulating data on positional candidate genes (e.g. Chehab 2008; Gat-Yablonski and Phillip 2008; Ichihara and Yamada 2008; Cheverud et al. 2009) provides the opportunity to explore functional hypotheses for identified loci and their interactions.

Utilizing the F10 generation of the LG,SM AI line (Cheverud et al. 2001) we further characterized the complex genetic architecture underlying murine reproductive fatpad weight. We first performed a slQTL scan on the original eight chromosomes harboring direct effect loci in the F2 generation (1, 6, 7, 8, 9, 12, 13, and 18) as well as the four shown to harbor slQTL in the combined F9-F10 population (3, 4, 10 and 16; Fawcett et al. 2010). Composite interval mapping and two-QTL tests were subsequently performed, the latter when multiple loci on a single chromosome were suspected. Finally, we carried out a full genome-wide scan for pair-wise epistasis. In order to identify the most meaningful set of loci to screen for candidate genes, marker genotypes representing slQTL and epiQTL that exceeded their appropriate thresholds were combined in linear models, first for each chromosome separately and ultimately the entire genetic system. Confidence intervals for peaks that remained significant in the full model were screened for positional candidate loci and potential physiological interactions via both the MGI database (www.informatics.jax.org/) and a literature search.

MATERIALS AND METHODS
Data: The population analyzed is the F₁₀ generation (N = 1298; 85 full-sib families; average litter size 8.45) of an Advanced Intercross (AI) line generated from an F₂ intercross of the inbred mouse strains SM/J and LG/J (Chai 1956; Chai 1956; Cheverud et al. 1996; Vaughn et al. 1999; Cheverud et al. 2001). The animal facility is maintained at a constant temperature of 21°C with 12-hour light-dark cycles. Animals were fed a standard rodent chow (PicoLab Rodent Chow 20 (#5053) with 12% of its energy from fat, 23% from protein, and 65% from carbohydrate) ad libitum and were weaned at 3 weeks of age. After weaning, animals were housed in single sex cages containing no more than 5 individuals.

Between the F₂ and F₁₀ generations the population was maintained at an effective size of approximately 300 with 75 mating pairs and no variance in family size. Mating between littermates was actively avoided. At greater than 13 weeks of age animals were sacrificed and necropsies performed. The reproductive fatpads of each animal were removed, combined and weighed on a digital scale to the nearest hundredth of a gram. Phenotypes were statistically corrected for age at necropsy, sex, litter size, and parity status (whether or not they were mated to produce the F₁₁) using multiple regression and the residuals used for further analysis. Genotypes for each individual were obtained at 1470 polymorphic SNPs across the genome by Illumina (San Diego, USA) GoldenGate Assay using DNA extracted from liver tissue obtained at necropsy. Inter-marker genotypes were imputed at 1 cM intervals using the equations of Haley and Knott (1992).

Mapping Analyses: A single locus QTL (sQTL) scan at all measured and imputed loci was first conducted on chromosomes 1, 3, 4, 6, 7, 8, 9, 10, 12, 13, 16, and 18 using the regression model:
\[ Y_i = \mu + a \cdot X_{ai} + d \cdot X_{di} + \text{error} \]  
\( (1) \)

where \( Y_i \) is the vector of corrected phenotypes, \( \mu \) is a constant, and \( X_{ai} \) and \( X_{di} \) are the vectors of genotype scores; \( a \) and \( d \) are the fitted additive and dominance regression coefficients respectively. The sums of squares for both model terms were pooled for significance testing. The results of the full genome-wide sQTL mapping in the combined F9-F10 generations were previously reported (Fawcett et al. 2010).

Composite interval (CI) mapping (Zeng 1994) was applied to the identified, preliminary confidence intervals using the following model:

\[ Y_{ijk} = \mu + a \cdot X_{ai} + d \cdot X_{di} + \text{error} | X_{aj} X_{dj} X_{ak} X_{dk} \]  
\( (2) \)

In this case, \( X_{aj}, X_{dj}, X_{ak}, \) and \( X_{dk} \) represent vectors of genotype scores at loci greater than 20 F10 cM up- and down-stream of the confidence interval on whose effects the within-interval regressions were conditioned. This eliminates the effects of proximal and distal QTL on the same chromosome from being confounded with the target QTL. When multiple peaks on the same chromosome were suggested, the fit of all pair-wise two locus models were compared to the appropriate single locus case using a \( \chi^2 \) test with two degrees of freedom (\( \chi^2_{\text{crit}} = 2 \cdot \text{abs} \left[ \ln(1/p_{\text{one}}) - \ln(1/p_{\text{two}}) \right] \)), where \( p_{\text{one}} \) and \( p_{\text{two}} \) are p-values from the one and two locus models respectively (Sokal and Rohlf 1995).
Finally all genome-wide, between-chromosome, pair-wise combinations of measured and imputed autosomal loci were tested using the following epistatic mapping model:

\[ Y_{ij} = \mu + aa(X_{ai}X_{aj}) + ad(X_{ai}X_{dj}) + da(X_{di}X_{aj}) + dd(X_{di}X_{dj}) + \text{error} | X_{ai} X_{di} X_{aj} X_{dj} \]  

where \( aa, ad, da, \) and \( dd \) are the additive-by-additive, additive-by-dominance, dominance-by-additive, and dominance-by-dominance epistasis regression coefficients and \( X_{ai} X_{di} X_{aj} X_{dj} \) represent vectors of the corresponding additive and dominance genotypes at the two loci involved. The sums of squares and degrees of freedom for all four epistatic components were pooled for initial significance testing. The raw probability associated with each multiple regression for all mapping analyses above was transformed to a linear scale using the base 10 logarithm of the inverse of the probability of no epistasis (LPR = \( \log_{10}(1/p) \)) producing values comparable to LOD scores obtained through maximum likelihood analysis (Lander and Botstein 1989).

**Thresholds:** Interpretation of these analyses is complicated both by the large number of comparisons involved as well as the family structure present in the population. In order to account for these two issues simultaneously, simulations were performed using the known pedigree of all individuals between the F2 and F10 generations to generate a null distribution of expected effects from which the appropriate single-locus LPR threshold was determined (Fawcett et al. 2008, Norgard et al. 2009). Given a heritability of reproductive fatpad weight in the F10 of 0.47 (from sib-correlations) chromosome-specific thresholds for identifying novel sQTL ranged from 6.15
(chromosome 8) to 6.6 (chromosome 1). The experiment-wide threshold for novel slQTL detection was 7.34. For the purposes of replication, a corrected point-wise threshold (equivalent to p = 0.05) of 3.32 was applied for slQTL peaks within previously identified confidence intervals.

Following the method described in Fawcett et al. (2010), the analysis-wide epistasis threshold for the identification of novel interactions was calculated to be 8.33. The threshold for tests between a given slQTL and all other unlinked markers in the analysis was 6.06 and the analogous chromosome-specific thresholds ranged from 4.73 (chromosome 8) to 5.25 (chromosome 1). The corrected point-wise threshold for epistatic tests between two slQTL was 3.44. Tests involving slQTL are partially protected from multiple comparisons as they were identified with independent information.

Confidence Intervals: Due to the complexity of our mapping strategy, the conventional 1 LPR drop criterion was applied to define all reported confidence intervals. When multiple peaks, either slQTL, epiQTL or both, occurred in the same region, the most proximal and most distal 1 LPR drop was used to determine CI endpoints. Confidence intervals (CI) for slQTL peaks were also calculated for each location individually using the standard deviation of the simulated distribution of 1000 mapping iterations involving known effects on simulated chromosomes (Norgard et al. 2009). The two techniques yielded very similar CI for all slQTL though the simulation-based intervals were slightly smaller.

Linear Models: We constructed and evaluated separate chromosome-specific models using the linear model function in R (R Development Core Team) before
combining their results into a full model of the genetic system. This process began with terms representing each significant effect at all slQTL peaks identified by the single locus model (equation 1) and composite interval mapping (equation 2). For example, the chromosome 1 model (see Figure 1A) began with five slQTL terms representing the additive \((p = 0.00726)\) and dominance \((p = 0.0007)\) effects at 20.15 Mb, the additive \((p = 0.000268)\) and dominance \((p = 0.0383)\) effects at 70.77 Mb and the dominance effect \((p = 1.06 \times 10^{-06})\) at 134.82 Mb. The additive effect at 134.82 Mb was non-significant in the slQTL mapping model \((p = 0.868)\) and so was not included. Likewise, the chromosome 13 model (see Figure 1B) included two terms representing the additive effects at 53.54 Mb \((p = 3.05 \times 10^{-06})\) and 90.61 Mb \((p = 4.88 \times 10^{-05})\) respectively. In this case, neither dominance effect was significant in the slQTL mapping model \((p = 0.798\) and \(p = 0.634)\) and so both were excluded. When considered jointly, some individual terms (e.g. the dominance effect only at 70.77 Mb on chromosome 1) no longer remained significant \((p < 0.05)\) in Type I ANOVA tables (using the “anova” function). Such terms were removed. For those chromosomes not found to harbor slQTL, a similar process was performed beginning with all significant interactions.

Next, individual coefficients from the epistatic mapping model \((aa, ad, da, dd;\) equation 3) at all peaks that exceeded their appropriate thresholds in the epiQTL scan were similarly examined to determine the type or types of interactions occurring. Terms representing all significant interactions were then added step-wise to each appropriate chromosome-specific model. Only epistatic terms that remained significant \((p < 0.05)\) in both Type I and Type II ANOVA tables, using the R functions “anova” and “Anova” (the latter from the package “car”) respectively and did not cause any established additive or
dominance effects to become non-significant (p < 0.05) were retained to define each final chromosome-specific model. These stringent criteria were established in order to obtain a tractable number of high-confidence CI to screen for positional candidates and physiological interactions.

Next, additive and dominance terms from all chromosome-specific models were combined and terms that became non-significant in either Type I or Type II ANOVA tables (or both) were culled to define the “slQTL system.” This model included 20 terms at 18 loci (15 additive and 5 dominance; bold in Supplemental Table 1). Epistasis terms significant in the chromosome-specific models were then added stepwise to the slQTL system as above to define the “full model.” In addition to the 20 marginal effect terms, this model includes 23 interaction involving 26 different epiQTL confidence intervals. Finally, since many epiQTL peaks occur at locations not represented in the slQTL system, the appropriate additive and dominance terms for each interaction were added to the full model to ensure that the identified epistatic contributions were not unduly biased upward by variance attributable to single locus effects. This had relatively little effect and resulted in the elimination of only 3 interactions, all of which are significant in Type I tests. The results from the full model are reported with these nominally significant terms noted in bold (Table 1, see below).

Candidate Genes: All CI for peaks identified in the full model were screened for plausible positional candidate genes and known interactions. This involved both queries of the MGI database for functional variants affecting adiposity as well as a broad literature search and was intended to generate meaningful and testable physiological hypotheses regarding the observed statistical associations.
RESULTS

**Replication and Identification:** Significant marginal effects, epistatic effects or both are observed in the F_{10} population on all eight chromosomes harboring the original Adip loci and three of the four additional chromosomes implicated in the combined F_{9}-F_{10} sQTL scan (Supplementary Figure 1). In the F_{10} alone, there were no significant sQTL on chromosome 16. Similar to the results of Fawcett et al. (2010), peak LPR scores from either the single locus scan or composite interval mapping at or near the confidence intervals of five Adip loci exceeded the experiment-wide threshold (7.34) even for novel QTL detection (Adip1: LPR = 9.2, Adip2: LPR = 8.9, Adip3: LPR = 8.3, Adip5: LPR = 9.6, and Adip8: LPR = 12.3). All three remaining F_{2} loci exceed the point-wise threshold (3.32) required for tests within previously defined confidence intervals (Adip4: LPR = 5.6, Adip6: LPR = 5.24; Adip7: LPR = 4.8). Additional sQTL on chromosomes 3, 4, and 10 also replicated. Interestingly, the chromosome 4 locus (Adip24, Fawcett et al. 2010; LPR = 12.65) roughly corresponds to two loci previously reported in the literature as Adip11 and Adip12 in a cross between C57BL/6J and DBA/2J (Keightley et al. 1996; Brockmann et al. 1998; Stylianou et al. 2006). Finally, composite interval mapping revealed novel loci on chromosomes 7 and 9 that both exceed their appropriate chromosome-specific thresholds of 6.36 and 6.38 respectively. A total of 22 potential marginal effect peaks were identified (Supplementary Table 1).

**epiQTL Mapping:** In the genome-wide scan for epistasis 177 peaks involving 217 interactions exceeded their appropriate significance thresholds and physically cluster into approximately 51 potential epiQTL (Supplementary Table 1). Additive-by-additive
interactions were the most common (98), Additive-by-Dominance or Dominance-by-Additive were the next most common (97) and Dominance-by-Dominance interactions were the most rare (22). Consistent with the results of Jarvis and Cheverud (2009) and several other studies (see Phillips 2008), many of these occurred at locations showing no significant marginal effects in this cross, though some occurred at locations significant in slQTL scans in other crosses (Table 1; Figure 1; Supplementary Table 1; Supplementary Figures 2-20).

**Linear Models:** In total, we identified 199 slQTL and epiQTL peaks that potentially contribute to trait variation. These cluster into roughly 73 confidence intervals showing a variety of combinations of additive, dominance and epistatic effects (Supplementary Table 1). In order to identify the most robust signals we systematically added vectors of genotype scores representing each into linear models and determined the set that is simultaneously significant in both Type I and Type II tests. We began by establishing a single locus model that contained all slQTL peaks that remain significant together. This slQTL system includes 20 marginal effect terms (15 additive and 5 dominance) shows an adjusted $R^2$ value of 0.2254 (F statistic = 18.64 on 20 and 1281 df). We next added epistatic peaks stepwise to generate a full model of the genetic system. This full model (Table 1) includes 23 additional interaction terms ($9 \text{aa}$, $10 \text{ad}/\text{da}$, and $4 \text{dd}$) involving 26 different epiQTL confidence intervals and shows an adjusted $R^2$ value of 0.3322 (F statistic = 15.71 on 43 and 1257 df). Using a chi-square goodness of fit test with 23 (43-20) degrees of freedom this represents a highly significant improvement in fit over the base slQTL model ($p < 10^{-25}$). Following the addition of all marginal terms involved in epistasis, three interaction terms become non-significant at the $p < 0.05$ level
in either the Type I or the Type II tables or both (bold terms in Table 1). Removing these interactions from the full model, its adjusted $R^2$ value is 0.3220 ($F$ statistic = 16.07 on 40 and 1260 df), which also represents a highly significant improvement in model fit ($p < 10^{-20}$).

**Positional Candidates:** While in-depth functional assays and other detailed molecular studies are required to sort out the biological basis of QTL and their interactions, examination of positional candidate genes in slQTL confidence intervals suggests testable physiological hypotheses for several observed statistical effects. In general, confidence intervals contain a variety of candidate loci including transcription factors, components of various signaling cascades (*e.g.* the *Wnt*, *Insulin*, and *Igf* signaling networks), neuro-endocrine hormones and their receptors, as well as genes directly implicated in glucose processing and metabolism. For example, the CI found at 6:133.92-142.67 Mb contains the promising candidate *Lrp6*, a low-density lipoprotein receptor-related protein that is thought to contribute to variation in a variety of metabolic risk factors in humans (Kahn et al. 2007; Mani et al. 2007) and *Cdkn1b*, a cyclin-dependent kinase inhibitor with known effects on pancreatic islet mass in diabetic mice (Uchida et al. 2005). Both *Lrp6* and *Cdkn1b* have differences in expression level in white fat ($p = 3.82 \times 10^{-12}$ and 0.013, respectively) and in the liver ($p = 1.62 \times 10^{-13}$ and $7.48 \times 10^{-8}$, respectively) between the two parental lines in this cross (Cheverud, unpublished results). The CI 18:58.77-80.76 Mb shows potential functional links to mammalian neurotransmitter signaling via *Htr4* (Gardner et al. 2008), as do 13:40.74-55.35 Mb via *Cplx2* (Brachya et al. 2006) and *Drd1a* (de Leeuw van Weenen et al. 2009). In addition, the region 6:114.73-121.97 Mb contains neuro-endocrine candidates *Adipor2*
(Yamauchi et al. 2007; Ziemke and Mantzoros 2010) and Ankrd26 (Bera et al. 2008), which also shows a significant difference in expression in liver between LG/J and SM/J (p = 0.0002; Cheverud, unpublished results). Together, these loci suggest a functionally similar genetic architecture to the emerging picture of Type 2 diabetes in humans (Doria et al. 2008).

There are also a number of strong candidate loci for observed epistatic interactions. The most striking involves the CIs 13:0-24.24 Mb and 1:118.37-138.01 Mb, which contain Inhba and Inhbb respectively. The proteins encoded by these loci are components of the Activin and Inhibin complexes which have wide-ranging effects on a variety of physiologic, homeostatic and metabolic processes including mammalian reproduction, inflammation and adipocyte differentiation (Woodruff and Mather 1995; Werner et al. 2006; Hirai et al. 2005). Interestingly, 13:0-24.24 Mb participates in five separate interactions that are significant in the full model (Table 1) and appears to interact with a region (9:68.10-95.10 Mb) containing an important receptor for serotonin (Htr1b).

Glutamate signaling and metabolism are also likely to underlie a portion of fatpad variation due to epistasis in this cross. The interacting epiQTL CI 1:42.41-52.71 Mb and 9:68.10-95.10 Mb contain the enzyme that catalyses the first reaction in the primary pathway for the renal catabolism of glutamine (Gls) and the first rate limiting enzyme in glutathione synthesis (Gclc) respectively. Gls also shows differential expression in white fat cells between the parental lines (p = 0.00097). Ghrelin and its associated pathways also appear as likely candidates. For example, 1:118.37-138.01 Mb contains Gpr39, a member of the ghrelin receptor family. This CI interacts with 6:133.92-142.67 Mb which harbors Pde3a, a locus known to be downstream of ghrelin signaling in platelets.
(Elbatarny et al. 2007) and which shows significant differences in gene expression in white fat between SM/J and LG/J (p=0.00018), and 12:73.42-89.12 Mb which contains Hif1a, whose protein product increases the expression of Vegf (Hoffmann et al. 2008). Interestingly, Vegfc shows a significant difference in expression in white fat between the parental lines (p = 0.001) and Vegfb shows differences in liver (p = 0.009). Ghrelin is also known to increase the expression of Vegf in human luteal cells (Tropea et al. 2007) and Vegf in turn, is thought to be an important regulator of adipogenesis and obesity (Cao 2007). A final interesting epiQTL CI is 12:108.99-120.28 Mb. It contains Dlk1, Meg3, and Rtl1, all three of which appear to participate in an interacting (and imprinted) network affecting growth in mice (Gabory et al. 2009).

DISCUSSION

While the family structure of an outbred population complicates some aspects of the mapping process, the F10 (and later) generations of advanced intercross lines hold an intrinsic advantage in mapping resolution over more conventional study designs. Here this advantage translated into a variety of results with important implications for mapping complex trait variation and new insights into the genetic architecture of murine fatpad weight.

The first and most striking result of this analysis from a mapping perspective is the relatively low level of overlap in the physical positions of slQTL and epiQTL peaks despite the analytical bias towards finding epistasis involving slQTLs due to their protected status with respect to multiple comparisons. Though slight discrepancies may be expected due to subtle patterns of linkage, larger map distances between peaks likely
indicate that multiple functional variants are present. Indeed, when both types are observed in close proximity, epistatic peaks tend not to line up well with their single-locus counterparts and epiQTL are frequently observed in regions showing no significant marginal effects at all (Figure 1; Table 1; Supplementary Table 1; Supplementary Figures 1-20). This supports the notion that a relatively large number of variable, functionally relevant loci exert their influence on complex trait variation primarily via epistatic interactions rather than through conventional additive and dominance effects. It is also interesting to note that some regions interact with multiple locations in the genome. For example, proximal chromosome 13 (13:0-24.24 Mb) shows five significant interactions in the full model including two with separate locations on chromosome 1. Identifying such repeated signals may be useful in developing significance thresholds that help ameliorate the penalties incurred by performing multiple comparisons. Such consistency may also help distinguish epiQTL at the center versus the edges of functional networks.

Next, in keeping with observations in congenic lines (e.g. Christians et al. 2006) as well as other recent sQTL mapping studies (Fawcett et al. 2010), F2 confidence intervals were frequently observed to divide into multiple significant sQTL (Figure 1, Supplementary Figure 1). Interestingly, we observe similar splitting of single-locus and epistatic signals. For example, at the proximal end of chromosome 1 (Figure 1A) marginal effect peaks observed in the F2, combined F2-3, and in an intercross between SM and NZO (obq7; Taylor et al. 2001) appear to resolve in our mapping population into three distinct peaks with two marginal effect loci flanking an epiQTL. This suggests that the original F2 and the subsequent F2-3 signals in this cross were composites of both single-locus and epistatic effects and that the boundaries of previously reported CI may
have been influenced by epistatic contributions to single-locus values. Thus, current estimates of the number of loci underlying trait variation are likely to be overly conservative and reported effect size estimates are potentially biased by the presence of multiple, closely linked functional elements. Interestingly, it also suggests that confidence intervals identified in other intercross experiments, especially those that share a parental strain, can be productively evaluated under \textit{a priori} epistatic hypotheses, which may also ease issues related to multiple testing. On this account, it is also striking that the epistatic network identified in Stylianou et al. (2006) as Chr4-\textit{Adip11} is centered on a region also identified here as contributing to the epistatic architecture of fatpad weight.

The results of composite interval mapping also suggest that adjacent slQTL and epiQTL impact the mapping process. For example, there is a dramatic and unexpected increase in significance (nearly 3 orders of magnitude) for the additive slQTL peak at 134.82 Mb on chromosome 1 when composite interval mapping was applied (Figure 1A). While this is the most dramatic example, such effects were repeatedly observed (Supplementary Figure 1) and on chromosomes 7 and 9 this resulted in the identification of two novel loci. Interestingly, this suggests that adjacent functional variants with opposite effects were fixed in the original parental lines during their production. Indeed, inspection of the regression coefficients from the full linear model shows that the epistatic peak closest to the slQTL signal at 134.82 Mb on chromosome 1 (DD with 12:73.42-89.12 Mb) and the marginal signal itself share a positive sign. However, the two slightly centromeric interactions involving the additive value on chromosome 1 (AA with 13:0-24.24 Mb and AD with 6:133.92-142.67 Mb) are both negative. Conditioning on these adjacent markers is indeed expected to enhance the signal of the neighboring
additive effect, consistent with our observations. Thus, comparing the results of conventional single-locus mapping model and composite interval mapping may be an indirect means of identifying neighboring functional variants. Further mapping in later generations of this Advanced Intercross will provide a great deal of additional information on the sign, magnitude and physiological basis for these observed effects as recombination is expected to further separate their statistical signatures.

**Conclusions:** The application of multiple mapping approaches, including an epistatic model, is a vital strategy for characterizing complex genetic architectures. Contrary to suggestions based on human GWAS findings, we found substantial numbers of pair-wise epistatic interactions involving many more loci than show single locus effects that account for an important portion of trait variation. This is likely due to the genetic structure of our experimental population where allele frequencies are intermediate; there are no rare alleles in our mapping system. This is critical since epistasis is known to produce predominantly additive and dominance variance when relatively rare alleles are involved (Cheverud and Routman, 1995; Cheverud, 2000).

Here, the use of a combination of techniques was further enhanced by the improved genetic resolution offered by AI lines. While single locus scans remain the most tractable, pair-wise epistatic relationships can now be dissected in great detail as well and the identification of candidate loci for such interactions is possible. This is especially true for characters for which a large body of literature exists describing the mechanistic relationships among candidate genes and related pathologies. In such cases, incorporating *a priori* information regarding functional interactions can be used to help focus epistatic mapping studies and both ease the difficulties associated with multiple
comparisons and facilitate the physiological interpretation of statistical results. It is an exciting prospect that even more fine-scale mapping of these loci will be possible in later generations of the LG,SM AI line. Undoubtedly future analyses, coupled with the incorporation of sequence information from the parental lines, will aid in further refining the physiological hypotheses presented here for fatpad variation and greatly contribute to our understanding of the statistical signatures of complex genetic architectures.

ACKNOWLEDGEMENTS
The authors wish to acknowledge G. L. Fawcett and C. A. Lambert for offering insightful comments and suggestions on earlier drafts of this manuscript. This work was supported by a grant from the National Institutes of Health (DK-055736) and a Doctoral Dissertation Improvement Grant from the National Science Foundation (DEB-0608352).


Genome-Wide Association Study Identifies IL23R as an Inflammatory Bowel Disease Gene. Science. **314**: 1461-1463.


Fenske, T. S., C. McMahon, D. Edwin, J. C. Jarvis, J. M. Cheverud, M. Minn, V.


Hanson, W. D., 1959a  The theoretical distribution of lengths of parental gene blocks in the gametes of an F1 individual. Genetics. 44: 197-209.

Hanson, W. D., 1959b  Theoretical distribution of the initial linkage block lengths intact in the gametes of a population intermated for n generations. Genetics. 44: 839-846.

Hanson, W. D., 1959c  Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or selfing. Genetics. 44: 833-837.

Hanson, W. D., 1959d  The breakup of initial linkage blocks under selected mating systems. Genetics. 44: 857-868.


Prenzel, and P. V. Danenberg, 2008  High expression of HIF1a is a predictor of clinical outcome in patients with pancreatic ductal adenocarcinomas and correlated to PDGFA, VEGF, and bFGF. Neoplasia. 10(7): 674–679.


Kamei, Y., H. Ohizumi, Y. Fujitani, T. Nemoto, T. Tanaka, N. Takahashi, T. Kawada, M. Miyoshi, O. Ezaki, A Kakizuka, 2003  PPARγ coactivator 1β/ERR ligand 1 is an
ERR protein ligand, whose expression induces a high-energy expenditure and antagonizes obesity. PNAS. **100**: 12378-12383.


Papst, C., M. Bohn, H. F. Utz, A. E. Melchinger, D. Klein, J. Eder, 2004 QTL mapping for European corn borer resistance (Ostrinia nubilalis Hb.), agronomic and forage


Smith Richards, B. K., B. N. Belton, A. C. Poole, J. J. Mancuso, G. A. Churchill, R. Li, J.

Volaufova, A. Zuberi, and B. York, 2002 QTL analysis of self-selected


F. Payne, J. R. B. Perry, E. Pettersen, C. Platou, I. Prokopenko, L. Qi, L Qin, N.
W. Rayner, M. Rees, J. J. Roix, A. Sandbæk, B. Shields, M. Sjögren, V.
Steinthorsdottir, H. M. Stringham, A. J. Swift, G. Thorleifsson, U.
Watanabe, M. N. Weedon, C. J. Willer, Wellcome Trust Case Control
Consortium, T. Illig, K. Hveem, F. B. Hu, M. Laakso, K. Stefansson, O. Pedersen,
McCarthy, M. Boehnke, and D. Altshuler, 2008 Meta-analysis of genome-wide
association data and large-scale replication identifies additional susceptibility loci
Ziemke, F. and C. S. Mantzoros, 2010 Adiponectin in insulin resistance: lessons from
TABLE 1: Results from the full linear model of the epistatic network underlying murine reproductive fatpad weight in the LG,SM AI line. Chromosome, confidence intervals (Mb), peak locations (Mb), peak LPR scores, nearest SNP to the peak, effect type threshold and threshold value are all given for each term. The appropriate references for any a priori hypotheses are listed along with positional candidate loci for both sQTL and epiQTL. Bold terms are nominally significant (p > 0.05) when additive and dominance effects for all interactions are included in the model. References: 1Cheverud et al. 2001; 2Fawcett et al. 2008; 3Taylor and Phillips 1996; 4Taylor et al. 2001; 5Cheverud et al. 2004; 6Yi et al. 2006; 7Ishimori et al. 2004; 8Fawcett et al. 2010; 9Stylianou et al. 2006; 10Togawa et al. 2006; 11Brockmann et al. 2000; 12Warden et al. 1995; 13Keightley et al. 1998; 14Rosen et al. 2005; 15Kim et al. 2001; 16Yi et al. 2004; 17Corva et al. 2001; 18West et al. 1994; 19Horvat et al. 2000; 20Smith Richards et al. 2002; 21Mehrabian et al. 1998.

FIGURE 1: Mapping results of significant terms from the full model of reproductive fatpad weight in the LG,SM AI line for chromosomes 1 (A) and 13 (B). Results from the single-locus model are given as connected grey dots, composite interval mapping as smooth black lines and epistatic interactions by other connected shapes. Confidence intervals from previous analyses are represented by horizontal bars below the QTL plot.
<table>
<thead>
<tr>
<th>Chr</th>
<th>C1</th>
<th>Begin (Mb)</th>
<th>End (Mb)</th>
<th>Peak 1 (Mb)</th>
<th>Peak 2 (Mb)</th>
<th>Peak 3 (Mb)</th>
<th>Genes (Mb)</th>
<th>Goodness Type</th>
<th>Threshold</th>
<th>Reported Adjause (q-value)</th>
<th>QTL Reference(s)</th>
<th>Candidate(s) CI 1</th>
<th>Candidate(s) CI 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>16.69</td>
<td>21.24</td>
<td>9.94</td>
<td>11.31</td>
<td>9.94</td>
<td>NA</td>
<td>Adip, Obq</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>65.79</td>
<td>78.85</td>
<td>6.76</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Adip, Obq</td>
<td>6.60</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>118.37</td>
<td>138.01</td>
<td>9.17</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Obq1, Gen1</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>20.36</td>
<td>27.82</td>
<td>5.56</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Gen1, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>6.71</td>
<td>11.82</td>
<td>10.83</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>79.28</td>
<td>98.33</td>
<td>11.87</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>114.73</td>
<td>121.07</td>
<td>5.01</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Obq1, Gen1</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>133.82</td>
<td>162.67</td>
<td>9.89</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>30.18</td>
<td>44.44</td>
<td>3.72</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>50.68</td>
<td>77.37</td>
<td>6.85</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>73.33</td>
<td>93.57</td>
<td>11.37</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>112.53</td>
<td>143.20</td>
<td>10.24</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>64.84</td>
<td>84.65</td>
<td>8.79</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>6.71</td>
<td>79.28</td>
<td>11.87</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>118.37</td>
<td>125.50</td>
<td>11.08</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>118.37</td>
<td>125.50</td>
<td>11.08</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>118.37</td>
<td>125.50</td>
<td>11.08</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>118.37</td>
<td>125.50</td>
<td>11.08</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>118.37</td>
<td>125.50</td>
<td>11.08</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>New</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>15.41</td>
<td>19.05</td>
<td>13.06</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>30.53</td>
<td>34.61</td>
<td>11.30</td>
<td>125.00</td>
<td>125.50</td>
<td>125.70</td>
<td>Adip2, Obq2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>125.88</td>
<td>139.02</td>
<td>13.81</td>
<td>125.20</td>
<td>135.70</td>
<td>136.70</td>
<td>Adip2, Obq2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
<td>143.17</td>
<td>154.78</td>
<td>13.81</td>
<td>125.20</td>
<td>135.70</td>
<td>136.70</td>
<td>Adip2, Obq2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>33.46</td>
<td>44.84</td>
<td>11.30</td>
<td>125.00</td>
<td>125.50</td>
<td>125.70</td>
<td>Adip1, Obq2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>55.02</td>
<td>74.18</td>
<td>7.02</td>
<td>106.47</td>
<td>106.97</td>
<td>107.05</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>132.63</td>
<td>142.07</td>
<td>5.01</td>
<td>106.47</td>
<td>106.97</td>
<td>107.05</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>124.83</td>
<td>129.12</td>
<td>9.07</td>
<td>106.47</td>
<td>106.97</td>
<td>107.05</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>20.24</td>
<td>29.76</td>
<td>13.12</td>
<td>110.66</td>
<td>110.66</td>
<td>110.66</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>108.99</td>
<td>120.28</td>
<td>13.11</td>
<td>110.66</td>
<td>110.66</td>
<td>110.66</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>132.63</td>
<td>142.07</td>
<td>5.01</td>
<td>106.47</td>
<td>106.97</td>
<td>107.05</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>172.61</td>
<td>189.05</td>
<td>13.12</td>
<td>110.66</td>
<td>110.66</td>
<td>110.66</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>40.74</td>
<td>55.35</td>
<td>8.06</td>
<td>9.89</td>
<td>9.89</td>
<td>9.89</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>143.52</td>
<td>154.77</td>
<td>13.12</td>
<td>9.89</td>
<td>9.89</td>
<td>9.89</td>
<td>Lrfn3, Gen2</td>
<td>7.33</td>
<td>adjP6308392</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
A) SlQTL Mapping

- CI Mapping
- AD with 6:133.92−142.67
- DD with 9:68.10−95.10
- AA with 9:104.05−118.18
- DD with 12:73.42−89.12
- AA with 13:0−24.24

B) F2

- F2−F3
- RI
- Other: obq7

- CI Mapping
- AD with 1:118.37−138.01
- AA with 1:174.21−189.05
- AD with 4:143.52−154.77
- AA with 6:134.92−NA
- AA with 9:68.10−95.10
- AA with 12:73.42−89.12
- AD with 18:24.19−56.21

Chromosome 1 (Mb)

Chromosome 13 (Mb)