

SUPPLEMENT - RESEARCH DESIGN AND METHODS

Animals

The outbred NIH heterogeneous stock (HS) rat used in this study came from a colony maintained at The Medical College of Wisconsin (NMcwi:HS; RGD_2314009). A total of 1519 male rats from 323 families were weaned at 21 days of age, housed 2 per cage in micro-isolation cages in a conventional facility using autoclaved bedding (sani-chips from PJ Murphy). They were given ad libitum access to autoclaved Teklad 5010 diet (Harlan Laboratories) and were provided reverse osmosis water chlorinated to 2-3 ppm. We ran animals through the experimental pipeline below using approximately 12 animals/batch, with a new batch being saved weekly.

Phenotypes

Glucose and insulin phenotypes

We calculated GluAUC, InsAUC, and QUICKI, a measure of insulin sensitivity, as follows:

$$GluAUC = \sum \frac{(SG_t + SG_{t+1}) \times \Delta min}{2}$$

Where serum glucose (SG) was measured at t=0, 15, 30, 60, 90 min. InsAUC was calculated using the same formula, where serum insulin replaced SG in the equation.

We calculated QUICKI [17] as $[\log(Ins0) + \log(Gluc0)]^{-1}$.

Phenotype transformation

Prior to statistical genetic analysis, tissue weights were adjusted to account for body weight by dividing by sacBW. Each phenotype was transformed to approximate normality, as summarized in Supplement Table 1. CHOL and TRIG were log transformed while all other traits except GluAUC (see explanation below) were rank-inverse normal transformed (RINT) as defined below:

$$RINT(y_i) = \Phi^{-1} \left(\frac{\text{rank}(y_i) - 0.5}{n} \right)$$

where $\Phi^{-1}()$ is the inverse cumulative distribution function of the standard normal distribution, y_i is the original phenotype and n is the number of observations. We found that RINT for GluAUC introduced more extreme values than were originally present, which

might generate false positive associations. Therefore, we applied a more conservative RINT transformation (TRINT), where data are mapped to a normal distribution with tails truncated at the θ percentile:

$$\text{TRINT}(y_i) = \Phi^{-1} \left(\theta + \frac{[\text{rank}(y_i) - 1](1 - 2\theta)}{n - 1} \right)$$

$\theta = 0.01$ was chosen for GluAUC as it showed reasonable looking values that appear normal without outliers. This setting limits the most extreme points to approximately 2.33 standard deviation from the mean.

RNA-Seq

RNA from liver and adipose tissue was extracted using Trizol. Rats for RNAseq were chosen to maximize genetic diversity (e.g., no more than 1 or 2 rats per family), while still encompassing the full phenotypic spectrum of fat pad weight. Liver tissue RNA-Seq was run by the Genomics Core at the Medical College of Wisconsin, while adipose tissue RNA-Seq was run by the Genomics Core at Wake Forest School of Medicine. For liver, poly-A libraries were prepared on the Illumina NeoPrep platform using the TruSeq Stranded mRNA Library prep Kit for NeoPrep (Illumina cat# NP-202-1001). For adipose tissue, Ribo depletion libraries were prepared using the Illumina® TruSeq Stranded Total RNA with Ribo-Zero Gold Preparation kit (Illumina Inc.).

Libraries were run on an Illumina HiSeq 2500 to obtain 37bp paired-end reads for liver and 75bp single-end reads for adipose tissue. We used STAR (v2.6.1a) [18] to align reads to the reference Rn6.0, PICARD (v2.5.0) to remove PCR (polymerase chain reaction) duplicates and featureCount in R package Rsubread to compute gene level expression counts which was later normalised by sequencing depth, gene length and RNA composition using DESeq2 R package (v1.24.0). We excluded very lowly expressed genes with average reads per sample < 1. The normalised expression of 18,358 genes from adipose and 16,796 genes from liver were then RINT-transformed for eQTL mapping and further analyses.

Low coverage genome sequencing and genotype imputation

All HS rats were Illumina-sequenced using 143bp paired-end reads low coverage sequenced at mean coverage of 0.24x at BGI, followed by imputation using STITCH. The STITCH algorithm uses a FAST-PHASE type approach, optimised for low-coverage sequence data.

DNA from the original eight founder strains were previously Illumina-sequenced at high coverage (24-28X) [14] with 143bp paired-end reads which were used to provide a haplotype reference panel. 4,865,047 imputed SNPs were retained after quality control (imputation info score > 0.4 and Hardy-Weinberg Equilibrium p-value > 10^{-6}). Imputed genotypes were then compared with a subset of 989 rats genotyped with a 10K Rat SNP array [10] - in which 3,253 SNPs were in common. The concordance and the correlation R^2 was 0.98 and 0.91, respectively, indicating the imputed genotypes were generally accurate. For genetic mapping we removed SNPs in high linkage disequilibrium (LD) with R^2 threshold of 0.95 using Plink version 1.9, resulting in a tagging set of 125,611 SNPs available for genetic mapping.

Statistical analyses

Founder haplotype dosage

Each HS rat chromosome is a mosaic of the eight founder haplotype reference panel, which STITCH represents as founder dosages to account for uncertainty in the haplotype assignment, either because of identity by state between founders, or ambiguities caused by heterozygosity or genotyping error. The STITCH haplotype dosages, which are initialized to be the known founder haplotypes, are dynamically updated during imputation and therefore need not be identical to the original founders. Therefore, we re-computed founder haplotype dosages from the imputed 122K tagging SNP genotypes plus the original founder haplotype reference panel, using R/qtI2.

Kinship matrices and estimation of heritability and genetic correlations

To account for the unequal relatedness between HS rats, we estimated the kinship matrix. This was done in two ways based on the application needed. The first is a SNP-based additive kinship matrix \mathbf{K}_{SNP} computed from the tagging SNPs, which was used to estimate heritability and genetic correlations and for genetic mapping of the expression traits. Heritability and genetic correlations were estimated in bi-variate linear mix model using Model 1 below. The second kinship \mathbf{K}_{HAP} used for genetic mapping of physiological traits was estimated in R/qtI2 using the founder haplotype dosages computed above and was calculated either for all chromosomes (used for SNP-based mapping) or each chromosome

separately by omitting haplotype data from that chromosome (“leave-one-chromosome-out”) (used for haplotype-based mapping).