Genome-Wide Fine-Mapping Of Diabetic Traits

Brittany Baur
Marquette University

GENOME-WIDE FINE-MAPPNG OF DIABETIC TRAITS

by

Brittany Baur

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

May 2013

ABSTRACT
GENOME-WIDE FINE-MAPPING OF DIABETIC TRAITS

Brittany Baur

Marquette University, 2013

Type 2 diabetes results from both genes and the environment. Mapping genetic loci in animal models can help identify genes that are involved in type 2 diabetes to better understand the disease. Heterogeneous stock (HS) rats are derived from eight inbred founder strains and maintained in a breeding strategy that minimizes inbreeding. HS rats have a highly recombinant genome, which allows for rapid fine-mapping of complex traits genome-wide. However, this results in a complicated set of relationships between animals that is non-existent in traditional genetic mapping methods. To fine-map traits involved in type 2 diabetes, multiple diabetic phenotypes were collected in 1,038 HS male rats and these animals were genotyped using the Affymetrix 10K SNP array. Following ancestral haplotype reconstruction, a mixed modeling approach was used to identify genetic loci involved in two phenotypes suggestive of diabetes: fasting glucose and glucose area under the curve after a glucose tolerance test. Sibship was used as a random effect in the model to account for the complex family relationships. A genome-wide significant marker interval was detected on chromosome 11 for fasting glucose with a 95% confidence interval of 5.75 Mb. Genome-wide significant marker intervals were also detected on chromosomes 1,3, 10, and 13 for glucose area under the curve, with the average 95% confidence interval for these loci being only 3.15 Mb. A multilocus modeling technique involving resample model averaging was applied to the fasting glucose phenotype. This technique determines how frequently each locus is detected when resampling a portion of the original data-set, thus reducing potential false positives. Multilocus modeling results for fasting glucose coincided with the significant marker interval demonstrated in the mixed modeling approach. Both approaches are effective at detecting significant marker intervals that are expected to be involved in the phenotype of interest with a greater resolution over traditional methods

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

Type 2 diabetes is a serious condition, which over time can lead to heart disease, blindness, kidney damage and nerve damage. The prevalence of the disease is approximately 13% in the US and is predicted to more than double by 2050 (Cowie et al., 2009; Boyle et al., 2010). Studies of various racial and ethnic groups suggest a genetic component to the disease. Prevalence of type 2 diabetes is as high as 50% in Pima Indians, whereas prevalence is close to 5% in non-Hispanic whites in the Unites States (King & Rewers, 1993). Genome-wide association studies have revealed many genes involved in type 2 diabetes in humans (Morris et al., 2012). However, many of these genes only contribute a small percentage to the population variance. One way to find additional genes is to use animal models, such as heterogeneous stock rats.

Heterogeneous stock (HS) rats are derived from eight inbred founder strains, and then maintained in a breeding paradigm that minimizes inbreeding. This strategy creates a high degree of recombination events in the HS animals. In other words, the genetic material in each HS rat is a random mosaic of each of the founder strains. This random mosaic allows rapid identification of quantitative trait loci (QTLs) with greater resolution over traditional methods. Quantitative traits are traits that can be measured numerically, such as height and weight, or in this case diabetic traits such as fasting glucose. QTLs are regions in the genome that are found to be significantly associated with a trait of interest. A QTL involvement in the phenotype of interest can be attributed to underlying changes in DNA. QTLs are important because they allow for the identification of genetic regions involved in complex traits, thus leading to identification of possible candidate genes. Importantly, narrower QTLs reduce the number of possible candidate genes that underlie the locus.

Traditional mapping methods include the F2 intercross, in which two inbred strains are mated and then their offspring are mated. This provides some recombination events to map QTLs,

but the number of possible genes in the region is large. The difficulty of detecting the underlying genes in traditional crosses can be attributed to a number of factors. Importantly, the genetic heterogeneity in these crosses is not representative of the total genetic variance present in animal populations in the wild (Mott, Talbot, Turri, Collins & Flint, 2001). Furthermore, multiple genes affecting a trait can be present within the same region which will appear to be one large QTL (Mott et al., 2001). HS rats circumvent these potential problems by increasing the number of parents from two to eight, followed by outbreeding for many generations, allowing for increased genetic diversity than an F2 intercross (Mott et al., 2001). Both the phenotypic and genetic diversity in the HS rat allows the colony to be more representative of the total phenotypic and genetic diversity of animal populations in the wild. Additionally, recombination distances are only a few centiMorgans (Mott et al., 2001). This could potentially separate multiple QTLs involved in the phenotype of interest that are within close proximity, and provides a narrow region for the trait of interest, leading to a smaller number of candidate genes (Flint & Eskin, 2012).

Heterogeneous stock rats allow for greater mapping resolution, but results in a complicated set of relationships between animals; an issue not existent in traditional F2 crosses. Due to the complex correlation structure, many false positives could result during mapping if the family structure is not taken into account (Valdar, Holmes, Mott & Flint, 2009). All heterogeneous stock rats are related at some level, and siblings are likely to have similar phenotypes than non-siblings. To take into account this complex structure, multilocus modeling or a mixed modeling approach with family as a random intercept can be used. The mixed modeling approach has previously been used for a single region on rat chromosome 1 (Solberg Woods, Holl, Tschannen & Valdar, 2010; Solberg Woods et al., 2012). This study narrowed down a region, Niddm1, which was previously identified in F2 studies and implicated in various metabolic traits.

Using a mixed model with family as a random effect controls for the unequal relatedness among animals to a large extent, but does not model associations between loci directly. These associations do not necessarily have to do with relatedness, but can be due to unmodeled environmental effects or other heritable elements that cannot be captured by overall genetic relatedness. In such cases, it is better to use a multilocus modeling approach which models confounding associations between loci directly (Valdar et al., 2009). Multilocus modeling reduces the number of false positives compared to the mixed-modeling approach.

The present study used a genome-wide fine-mapping approach for glucose area under the curve (AUC) and fasting glucose, two phenotypes indicative of type 2 diabetes. Fasting glucose is a marker of type 2 diabetes in humans (Nichols, Hillier & Brown, 2008)**.** Glucose AUC is typically used in animal models to determine type 2 diabetes (Solberg Woods et al., 2010). The fasting glucose phenotype involved a basal sample of blood taken after 16 ± 1 hours of fasting. For glucose AUC, subsequent samples were taken 15, 30, 60 and 90 minutes after a 1 g/kg body weight glucose injection. Glucose AUC is the area under this curve. Over 1,000 animals were genotyped and phenotyped for use in this study. The average spacing between any two genotyped SNPs was 353 Kb.

This study rapidly identified several QTLs across the genome in a two-step approach. The first step involved estimating the expected proportions of founder haplotypes between each marker interval (Mott et al., 2001). In the second step, an appropriate null formula was found and then mixed model linear regression was used to identify significant QTLs. Mixed model linear regression is a type of regression model that allows the deterministic component to be a linear function of separate predictors made up of both fixed and random effects (Gelman & Hill, 2007). Using a fixed effect assumes the subjects in which measurements are drawn from are fixed. Using a random effect assumes that the measurements are drawn from a random sample of a larger population. A random effect was chosen for family because the differences between family members are of interest. Restricted estimate maximum likelihood (REML) was used for fitting

linear mixed models. A likelihood ratio test was performed between a model that omits locus-specific information and a model that contains the vector of haplotype proportions, which followed the full model. The additive model, which was not tested in this thesis, allows for increased phenotypic effect when the number of causal alleles increases. The full model follows the additive model, but allows for interactions between alleles, including dominance (Valdar et al., 2006). Additionally, a multilocus approach, which used non-parametric resampling of the population and forward selection to predict the QTL model, was applied to the fasting glucose phenotype (Valdar et al., 2009). This approach can reduce the number of false positives by detecting only true QTLs, and not peaks that may be confounded with the true QTL.

II. METHODS

QTLs were fine-mapped by first inferring the underlying haplotype structure. The HAPPY algorithm was applied, which uses hidden Markov models to estimate the expected proportions of founder haplotypes in each marker interval based on the observed genotypes (Mott et al., 2001). The genotypes of each HS rat chromosome are treated as observed symbols emitted from a chain of unobserved haplotype states. The vector of haplotype proportions, $g_i(m)$, for a marker interval *m* of rat *i* is then used in place of the raw genotype (Solberg Woods et al., 2010). Founder haplotypes characterize the rat's genome better than raw genotypes, as two strains exhibiting opposite effects at a QTL may share the same alleles at the genotyped marker, thus making the QTL impossible to detect. HAPPY uses flanking marker information and recombination distances to estimate the haplotype state, thus providing more information than obtained with a single marker, leading to gains in mapping power. Importantly, a HS rat's genome should be characterized as the founder haplotypes on marker intervals because it is highly unlikely that a raw, genotyped SNP is also a causal variant.

To run HAPPY, an alleles file and a data file are needed for each chromosome. The alleles file contains the position of each marker, as well as the probability that each allele is found in each of the eight founder strains. The data file contains the phenotype information, as well as each of the genotyped markers for all of the animals used in the study. The Python programming language was used to create these files from the original genotype and phenotype raw data files. Non-informative markers were excluded from the analysis. Non-informative markers are defined as markers where all of the founders are homozygous for the same allele. 1,963 uninformative SNPs were removed from the genome. This brought the total number of SNPs from 10,714 to 8,751. This served to reduce run-time without affecting the output.

Given the probabilities that each allele is found in each of the eight founder strains and the genotyping information for all of the HS animals, HAPPY can then estimate the ancestral

haplotypes for each individual. As a result of the HAPPY algorithm, the interval between two genotyped, observed SNPs is thus described as the estimated descent of founder haplotypes within that interval for each individual in the HS population (Mott et al., 2001). This is used for both the mixed modeling technique and the multilocus modeling technique to provide locus-specific information to the model. The vector of haplotype proportions is included in both of the models described below.

All identification of normalizing transformations were guided by the Box-Cox procedure. Any values three standard deviations from the mean were removed to moderate the effects of outliers. Only one outlier was removed for fasting glucose, and none for glucose AUC. Sibship was added to the model as a random effect to account for the complex correlation structure of the HS rat in mixed-modeling analysis. Environmental covariates were also taken into account, as they may explain some of the residual variance and lead to gains in the ability to detect true QTLs (Broman & Sen, 2009). Covariates were tested with an ANOVA using the R function lme from the package MASS (Venables & Ripley, 2002). Additionally, the log likelihood of the model with just sibship as a random effect was compared to the model with sibship plus the covariate.

After transformation to approximate residual normality, Valdar et al. (2009) defines the model for the effect of a genetic locus on the phenotype of rat *i* as:

$$y_i = \mu + \sum_{c \in C} \beta_c^T x_i(c) + \beta_m^T g_i(m) + s_{k[i]} + \varepsilon_i$$

In the above equation, $x_i(c)$ is defined as the value of covariate c of individual *i.* C is the set of all covariates (Solberg Woods et al., 2010). The set of pretreatment covariates is defined as: bleeder for fasting glucose and number of glucose injections and injector for glucose AUC. $g_i(m)$ is the vector of haplotype proportions, as described above, from the HAPPY algorithm for the full models (Solberg Woods et al., 2010). β is used generically to describe the predictor's effect. $s_{k[i]}$ is the effect of sibship *k* to which rat *i* belongs. $s_k \sim N(0, \sigma_s^2)$ and $\varepsilon_i \sim N(0, \sigma^2)$ are

random intercepts taken from a normal distribution (Solberg Woods et al., 2010). All models were fitted by restricted estimate maximum likelihood (REML). The nominal significance of association at each locus was calculated by a likelihood ratio test based on full likelihoods at the REML estimates of a model that omits locus-specific information against one that contains the vector of haplotype proportions, $g_i(m)$ (Solberg Woods et al., 2010). The equation above is a standard approach. Modeling the overall effects of genetic relatedness via one or more random intercepts was originally used in plant and animal breeding strategies prior to the HS rat (Valdar et al., 2009).

A genome-wide significance threshold was applied by repeating the genome scan on 200 sets of phenotypes generated by parametric bootstrap from the fitted null model (Solberg Woods et al., 2010). The maximum logPs of those scans were used to fit a generalized extreme value distribution. The upper 5% quantile was taken to be the 5% genome-wide significance threshold (Solberg Woods et al., 2010). To calculate the genome-wide significance threshold, the null scan results from all chromosomes were used simultaneously, therefore it inherently took into account comparisons between chromosomes. 95% confidence intervals were approximated using the 1.5-LOD (logarithm of odds) drop method. The 1.5-LOD drop is defined as the markers enclosing the interval that lie just outside of the 1.5-LOD drop (Solberg Woods et al., 2012). In cases where the LOD score drops below 1.5 and then comes back, a larger connected interval was used (Broman & Sen, 2009).

A model selection approach, as described by Valdar et al. (2009), was used to model confounding associations between loci directly. This approach distinguishes between true QTLs and confounding associations. The equation for the multilocus model is defined by Valdar et al. (2009) as:

$$y_i = \mu + \sum_{c \epsilon C} \beta_c^T \, x_i(c) + \sum_{m \epsilon M} \gamma_m \beta_m^T \, g_i(m) + \varepsilon_i$$

where M is the set of all genetic predictors and $\gamma_m \epsilon \{0,1\}$ is a variable indicating inclusion or exclusion of the predictor from the model (Valdar et al., 2009). In order to predict γ_m, non-parametric resampling was used, in which a proportion of the full data set was used to generate a new data set (Valdar et al., 2009). The QTL model was chosen using forward selection. When the logP of the partial F-test failed to exceed the 5% genome-wide significance threshold determined by permutation, forward selection was terminated (Valdar et al., 2009). Loci were excluded from further analysis if the logP of the partial F-test failed to exceed the 20% genome-wide significance threshold determined by permutation.

Applying model selection to the R resamples gives the following R x n(M) matrix:

$$\Gamma^T = [\gamma^{(1)}\, \gamma^{(2)}\, \ldots\, \gamma^{(R)}]$$

The number of times a genetic locus, m, is included in the multilocus model is given by the Resample Model Inclusion Probability (RMIP). A RMIP of .25 has been found to be true in 70% of cases (Valdar et al., 2006).

The models were fitted and covariates were analyzed using the statistical package R (R-Core-Development-Team, 2004) with the add-on packages lme4 (Bates, Maechler & Bolker, 2011) and MASS (Venables & Ripley, 2002). The HAPPY software was used for the haplotype reconstruction methods (Mott et al., 2001). The software BAGPIPE was used for the mixed model linear regression and significance thresholds (Valdar et al., 2009). Bagphenotype was used for the multilocus modeling (Valdar et al., 2009).

III. RESULTS

A. Transformation and Environmental Covariates

Fasting glucose was normalized with a square root transformation. After transformation, the correlation coefficient of the normal probability plot of the residuals was 0.99**.** Bleeder was added to the model as a covariate, as it increased the log likelihood from -395.2 to -379.3. Glucose AUC was normalized with the inverse square root transformation. After transformation, the correlation coefficient of the normal probability plot of the residuals was 0.99. Number of glucose injections (DblInj) had a significant p-value of 0.007 when an ANOVA was performed. Injector did not give a significant p-value, but did increase the log likelihood from 6588 to 6596. Injector and number of glucose injections were only relevant to the glucose AUC phenotype, since fasting glucose was measured before any injection. Bleeder had a larger effect on fasting glucose than glucose AUC possibly because of the lack of other stimuli in collecting the fasting glucose data. Figure 1 shows the plots for fasting glucose and glucose AUC versus family. These plots demonstrate a high degree of variability, both between and within family, for these traits. There are 273 unique families and between one to thirteen animals per family.

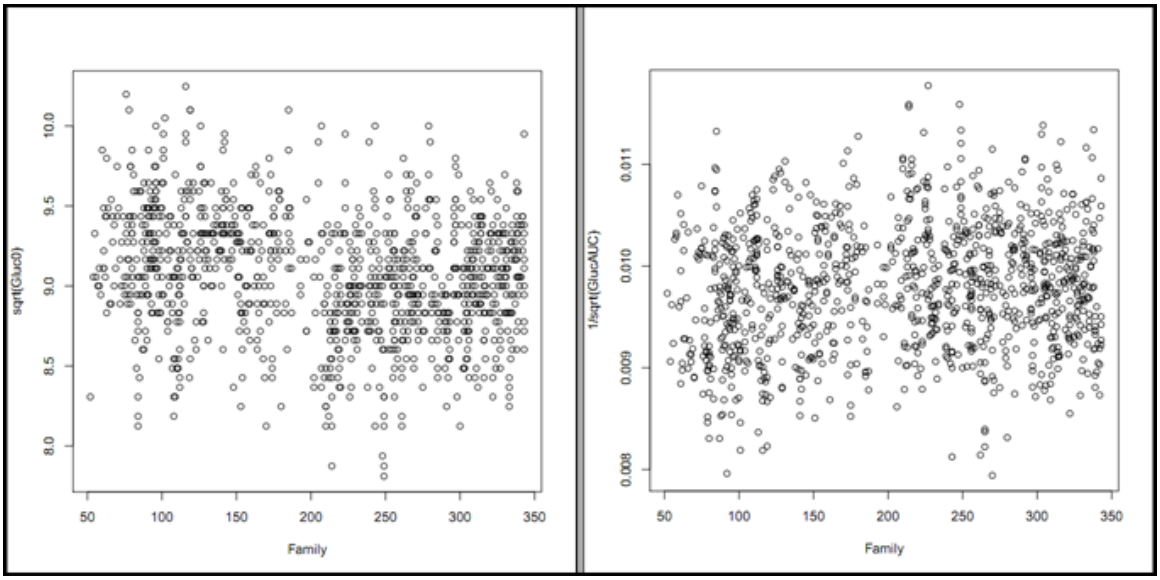

Figure 1: Fasting glucose(left) and glucose AUC(right) against family.

B. Fasting Glucose

Following ancestral haplotype reconstruction, mixed model linear regression was performed, using sibship as a random intercept. This approach largely takes into account the family structure of the HS rat. After performing a likelihood ratio test between the null model (Equation 1) and the model containing the vector of haplotype proportions for the full model (Equation 2), the -logP values were plotted against chromosomal position in centiMorgans (cM). cM was used because HAPPY uses recombination distances to estimate the haplotype variation. Therefore, -logP values indicate the significance of association between the haplotype variation along the chromosome and the phenotype of interest. Megabases (Mb) are reported for the purpose of identifying candidate genes, with one cM being approximately two Mb.

[1] sqrt(Gluc0) ~ (1|Family) + Bleeder
[2] sqrt(Gluc0) ~ (1]Family) + Bleeder + interval.full(THE.LOCUS)

This mixed-modeling approach demonstrated a broad, significant marker interval on chromosome 11 for the fasting glucose phenotype (Figure 2 & 3). The 5% genome-wide significance threshold for this phenotype was 5.258. The peak marker interval of this region was from 27.12 Mb to 27.26 Mb with a -logP of 6.8. When accounting for the multiple peaks in close proximity, the 1.5-LOD drop interval stretched from 21.76 Mb to 27.51 Mb (5.75 Mb).

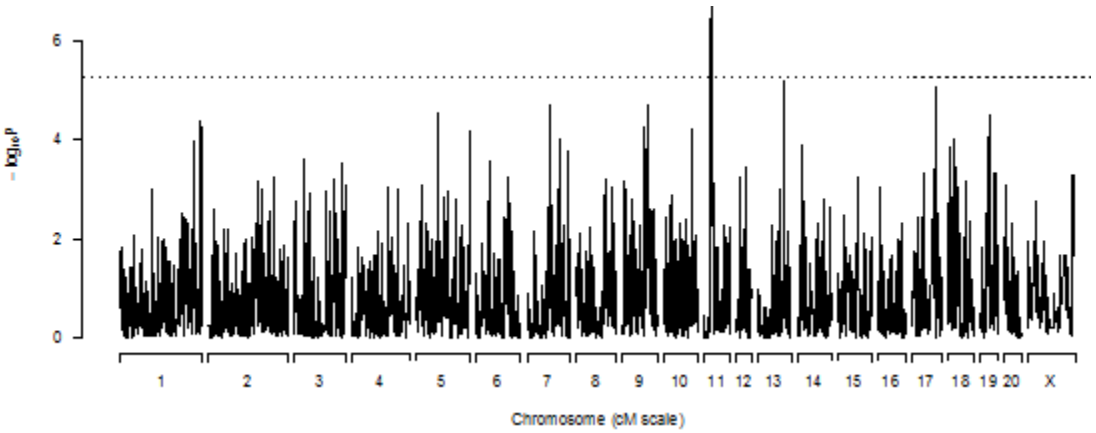

Figure 2: Genome-wide scan of fasting glucose

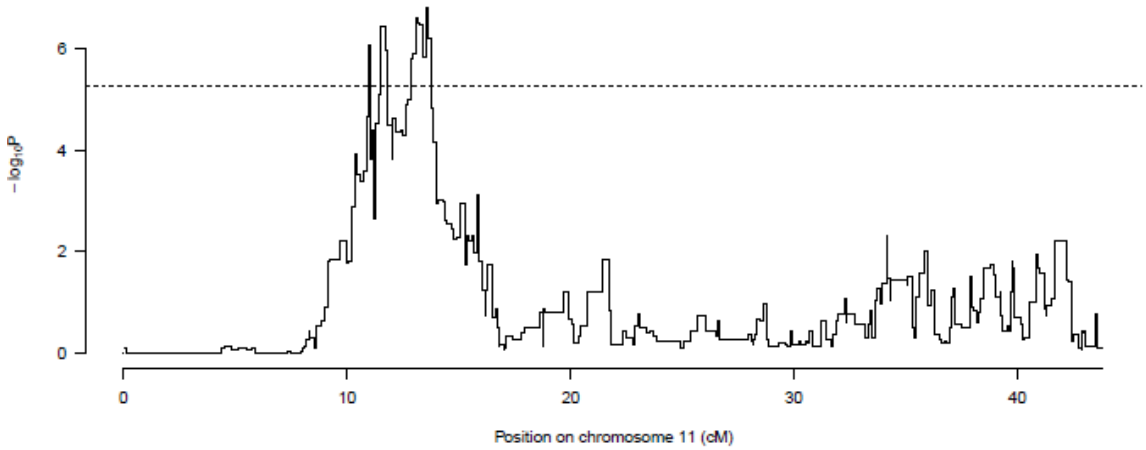

Figure 3: Significance of association of haplotype variation with fasting glucose on chr. 11

A multilocus approach was also applied to the fasting glucose phenotype. This model selection approach uses resample model averaging in order to predict those true QTLs that do not result from confounding associations between loci. The proportion of the full data set that was used to generate the new data set was 0.8. One thousand forward selections executed in order to predict the QTL model. Loci were excluded from analysis if the logP of the partial F-test failed to exceed 4.39, the 20% genome-wide significance threshold determined by permutation of the fitted null model. Forward selection was terminated if the logP of the partial F-test failed to exceed 5.00, the 5% genome-wide significance threshold determined by permutation of the fitted null model.

A similar peak in this approach corresponded with the peak above (Figure 4). The peak resample model inclusion probability was 0.151 and this was the highest peak in the genome. This peak interval ranged from 27.12 Mb to 27.25 Mb. The colors in Figure 4 vary with window radius size, and darker colors indicate a larger window size (Valdar et al., 2009). The window radius sizes chosen were 1 cM, 2 cM and 4 cM. The light green, for instance, indicates the probability that a marker would be chosen within 1 cM of any given point on the x axis. At the highest peak of 0.151, there is a 0.31 probability that a marker is chosen within a 1 cM window size, 0.47 within a 2 cM window and 0.624 within a 4 cM window.

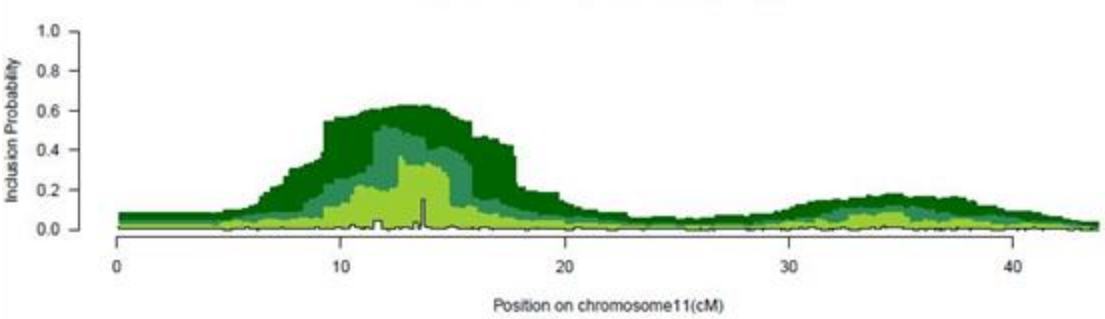

Figure 4: Resample model inclusion probabilities across chr. 11

C. Glucose Area Under the Curve

The mixed-modeling approach was applied to an additional phenotype, glucose AUC. A likelihood ratio test was performed between the null model (Equation 3) and the null model plus locus specific information (Equation 4).

[3] 1/sqrt(GTotalAUC)~(1|Family) + DblInj + injector
[4] 1/sqrt(GTotalAUC)~(1|Family) + DblInj + injector + interval.full(THE.LOCUS)

The 5% genome-wide significance threshold for the glucose AUC was 5.153. As seen in Figure 5, significant loci were detected on chromosomes 1, 3, 10 and 13.

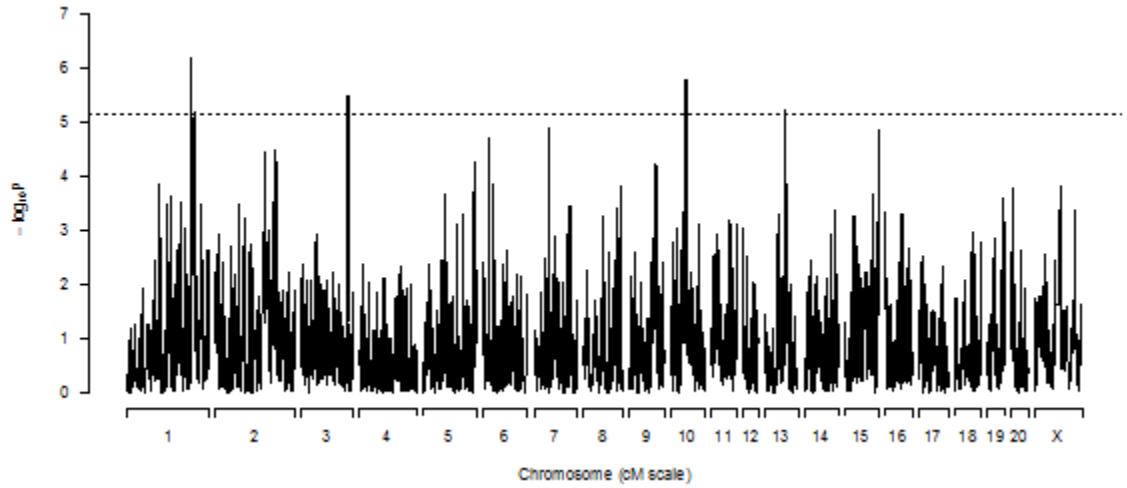

Figure 5: Genome-wide scan of glucose AUC

Due to the sharp drop in -logP and the numerous previously discovered QTLs in this region, chromosome 1 will be reported as having two separate peaks (Figure 6). The first peak

marker for this region was from 210.11 Mb to 210.31 Mb with a -logP of 6.2. The 1.5-LOD drop was from 209.12 Mb to 211.45 Mb (2.33 Mb). The second peak was 220.63 Mb to 220.87 Mb with a -logP of 5.2. The 1.5-LOD drop interval was from 220.21 Mb to 220.87 Mb (0.67 Mb).

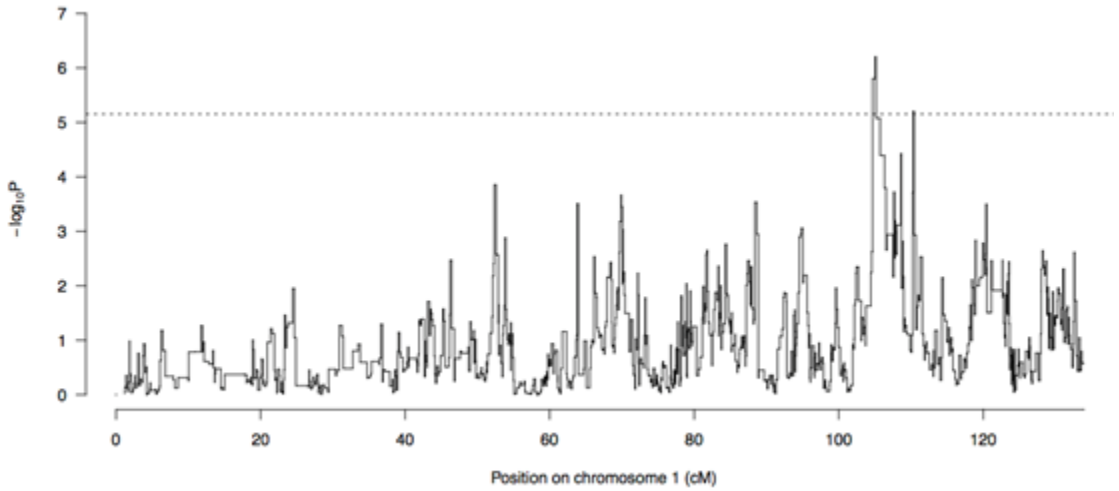

Figure 6: Significance of association of haplotype variation on with glucose AUC on chr. 1

Another marker interval that passed the genome-wide significance threshold is located on chromosome 3 (Figure 7). The peak marker interval for this region was 152.56 Mb to 153.13 Mb. The –logP for this interval was 5.49. The 1.5-LOD drop interval spanned from 150.66 Mb to 154.22 Mb (3.56 Mb).

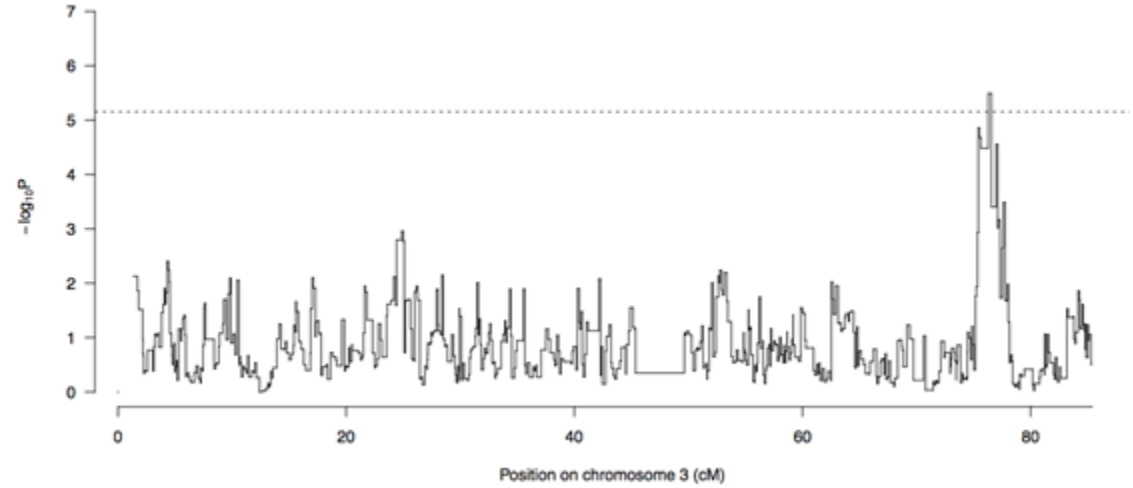

Figure 7: Significance of association of haplotype variation with glucose AUC on chr. 3

A larger region on chromosome 10 was also detected as a significant QTL (Figure 8). The peak marker interval for this QTL was from 49.78 Mb to 50.07 Mb, and had a –logP of 5.94. The 1.5 LOD-drop interval spanned from 49.36 Mb to 55.3 Mb (5.94 Mb).

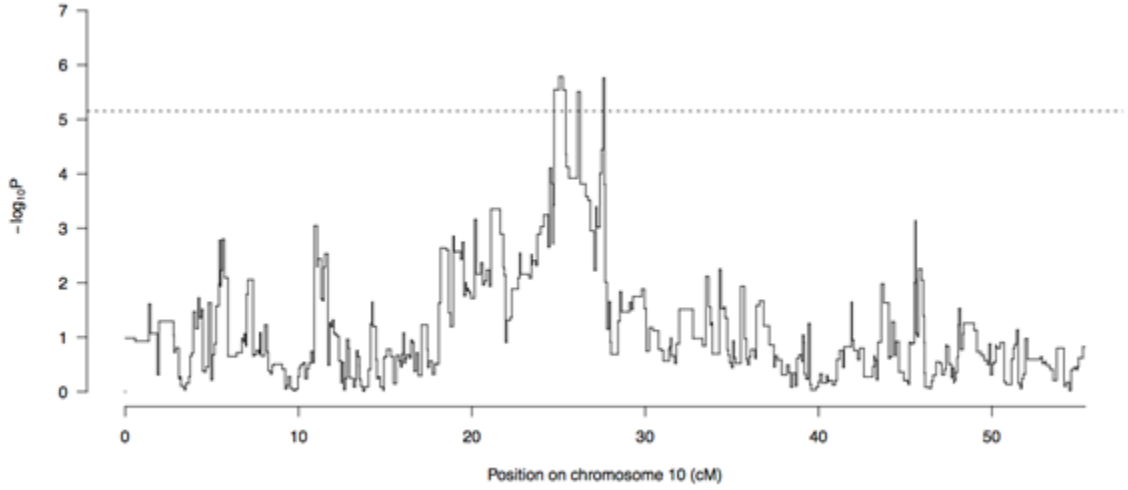

Figure 8: Significance of association of haplotype variation with glucose AUC on chr. 10

Finally, a significant marker interval was detected on chromosome 13 (Figure 9). This thin peak spanned from 68.98 Mb to 69.38 Mb, with a –logP of 5.24. The 1.5-LOD-drop interval ranged from 68.82 Mb to 69.62 Mb (.8 Mb). The significant intervals for the mixed-modeling analysis are summarized in the table below.

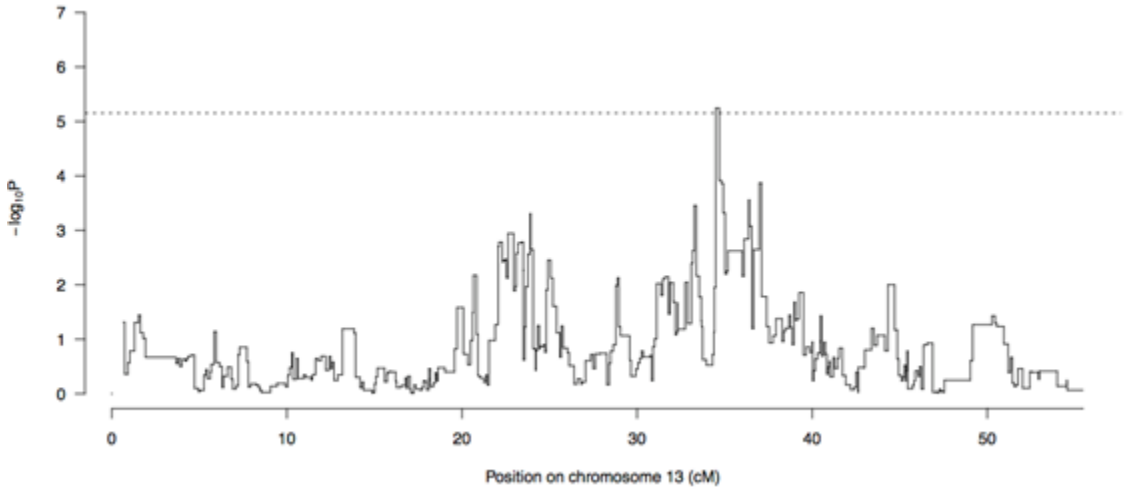

Figure 9: Significance of association of haplotype variation with glucose AUC on chr. 13

Phenotype	Chr	Peak Marker Interval	-logP	1.5-LOD drop (size in Mb)
Fasting Glucose	11	27.12 - 27.26	6.80	21.76 - 27.51 (5.75)
Glucose AUC	1	210.11 - 210.31	6.20	209.12 - 211.45 (2.33)
Glucose AUC	1	220.63 - 220.87	5.20	220.21 - 220.87 (0.67)
Glucose AUC	3	152.56 - 153.13	5.49	150.66 - 154.22 (3.56)
Glucose AUC	10	49.78 - 50.07	5.94	49.36 - 55.3 (5.94)
Glucose AUC	13	68.98 - 69.38	5.24	68.82 - 69.62 (0.8)

Table 1: Summary of genome-wide significant QTLs detected in linear mixed-model regression

IV. DISCUSSION

All significant marker intervals detected in this study, except for the region on chromosome 13, overlap with at least one previously discovered QTL involved in a diabetic phenotype. Many of the previously known QTLs were determined by more traditional methods (such as F2 intercross or backcrosses) and therefore span a much larger range then the peaks detected in this study. Narrowing down the known QTLs is important because it reduces the number of possible candidate genes in the region.

According to the Rat Genome Database (www.rgd.mcw.edu), the region on chromosome 11 for fasting overlaps with Iddm17, which is involved in non-fasting glucose levels (Yokoi et al., 1997). Iddm17 spans from 19.46 Mb to 87.76 Mb, a 68.3 Mb region. Assuming that the same QTL is being observed in this study, the present marker interval narrows the region for Iddm17 to only 5.75 Mb. There are 35 known genes within this region. Of interest are Adamts1 and Adamts5, two genes that fall within the 1.5-LOD drop range. These genes are potential candidates because Adamts9 has been identified in a human genome-wide association study, where it was implicated in type 2 diabetes (Zeggini et al., 2008). Adamts proteins are metalloproteases that degrade aggrecan, one of the major components of the extracellular matrix in cartilage (Mitani et al., 2006). Adamts1 inhibits capillary sprouting and causes angiopathy, a common problem in diabetic patients (Hohberg et al., 2011).

The significant marker interval on chromosome 11 for fasting glucose also overlaps with a known QTL, Niddm51 (Watanabe et al., 1999), a QTL identified in the OLETF rat for body fat. Although this is a different phenotype from fasting glucose, many diabetic phenotypes are inter-related (Scott et al., 2012). Therefore, this QTL could have the same underlying gene. However, there is a possibility that separate genes within the same region are governing these two phenotypes. The base pair positions for Niddm51 are 22.21 Mb to 43.16 Mb, over a 20 Mb

region. Niddm51 was discovered by using a backcross method, in which a hybrid is crossed with a parent.

The significant marker interval on chromosome 1 for glucose AUC from 209.12 Mb to 211.45 Mb overlaps with 8 previously discovered QTLs involved in diabetic phenotypes (Solberg Woods et al., 2010). This marker interval does not overlap, but is close to the glucose AUC QTL that was fine-mapped in a similar HS population in Solberg Woods et al., 2010. The 1.5-LOD drop region reported here is only 2.33 Mb. According to the Rat Genome Database, there are 69 known genes in this region. The second significant marker interval from 220.21 to 220.87 contains a gene, Gcnt1, that has been implicated in type 2 diabetes (Nishio et al., 1995). This enzyme may be responsible for the increased deposition of glycoconjugates and abnormal heart function found in the hearts of diabetic rats (Nishio et al., 1995).

The significant marker interval on chromosome 3 for glucose area under the cruve overlaps with Niddm39 (Ogino et al., 1999). This QTL has been implicated in non-fasting blood glucose levels. Niddm39 spans a large region, with positions approximately 93.89 Mb to 161.98 Mb (68.12 Mb). Hnf4a is a candidate gene for Niddm39, as evidenced by a study involving SNPs in the P2 promoter region of this gene (Borroso et al., 2008). The region identified in the current study is 3.56 Mb and contains 23 known genes. Interestingly, hnf4a lies just outside of this region and could therefore be a candidate gene for this QTL.

The significant marker interval on chromosome 10 for glucose area curve also overlaps with a known QTL, Gluco60, and was detected using the Goto-Kakizaki rat, a model of early onset type 2 diabetes (Nobrega, Woods, Fleming & Jacob, 2009). The position for Gluc60 is 14.72 Mb to 154.167 Mb, or 139.45 Mb. The interval reported here is much smaller than this region, at only 5.94 Mb, and containing 49 genes.

The narrow peak on chromosome 13 for glucose AUC does not overlap with any previously discovered diabetic QTL. It is possible that this QTL was missed by traditional methods because of the small interval size. There are only 7 genes in this region. Interestingly,

Cacna1e is among these genes. Cacna1e is a voltage-gated ion channel involved in the regulation of insulin secretion in a cellular response to glucose stimulus (Matsuda, Saegusa, Zong, Noda & Tanabe, 2001). This gene has been associated with hyperglycemia (Matsuda et al., 2001), which is strongly associated with the glucose AUC phenotype and is therefore a potential candidate gene for this QTL.

As expected, the multilocus modeling results correspond with the mixed modeling results, indicating the chromosome 11 QTL for fasting glucose, is likely a true QTL. The peak marker interval for both the RMA and the mixed-modeling is the same (27.12 Mb to 26.26 Mb). The mixed-modeling approach is more readily understandable, less computationally expensive, easier to use and largely takes into account the uneven distribution of phenotypes among siblings. However, the multilocus approach models the loci directly and therefore reduces the number of false positives by taking into account confounding associations. There could be a number of unmodeled environmental effects or effects that are not captured well by genetic relatedness. The multilocus modeling technique compensates for these issues by modeling loci directly. Whenever possible, a peak detected with mixed-model linear regression should be complimented with a model selection approach.

Both approaches are successful at detecting QTLs across the genome for a variety of complex traits (Valdar et al., 2006). The QTLs are detected at a narrower interval than in traditional mapping methods. Traditional mapping methods work well for quickly determining larger loci that may play a role in the phenotype of interest without the complications that arise from a complex family structure. However, HS rats can narrow these previously known QTLs to a few megabases with the methods described above. These narrower intervals lead to rapid identification of candidate genes in this study, which would not be possible with an F2 intercross. This demonstrates the usefulness of having narrower intervals for QTLs.

Future work should be geared towards implementing an approach that takes into account epistatic interactions between loci, which was not done here. Furthermore, additional traits will be

analyzed, including insulin phenotypes, and the additive model will be used. The 1.5-LOD drop is limited as a method of estimation. It is possible for the causal gene to lie just outside of the identified confidence interval. Therefore, bootstrapping should be performed to provide additional confidence to the region (Solberg Woods et al., 2012). Finally, follow up expression and sequence analysis of key candidate genes from this study will aid to confirm or dismiss the gene's involvement in type 2 diabetes.

V. BIBLIOGRAPHY

Bates D, Maechler M, Bolker B. lme4: linear mixed-effects models using S4 classes. R package version 0.99375-50, 2011, http://CRAN.R-project.org/package=lme4.

Borroso I, Laun J, Wheeler E, Whittaker P, Wasson J, Zeggini E, Weedon MN, Hunt S,Venkatesh R, Frayling TM, Delgado M, Neuman RJ, Zhao J, Sherva R, Glaser B, Walker M, Hitman G, McCarthy MI, Hattersley AT, Permutt MA, Wareham NJ, Deloukas P. Population-specific risk of type 2 diabetes conferred by HNF4A P2 promoter variants: a lesson in replication studies. *Diabetes* 57:3161-3165, 2008.

Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality and prediabetes prevalence. *Popul Health Metr* 8:29, 2012.

Broman KW, Sen S. *A guide to QTL mapping with R/qtl*. New York: Springer, 2009.

Cowie CC, Rust KF , Ford ES, Eberhardt MS, Byrd-Holt DD, Li C, Williams DE, Gregg EW,Bainbridge KE, Saydah SH, Geiss LS. Full accounting of diabetes and pre-diabetes in the U.S. population in 1988-1994 and 2005-2006. *Diabetes Care* 32: 287-294, 2009.

Doria A, Patti ME, Kahn CR. The emerging genetic architecture of type 2 diabetes. *Cell Metab* 8:186-200, 2008.

Flint J, Eskin E. Genome-wide association studies in mice. *Genetics* 13:807- 817, 2012.

Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models*. New York:Cambridge University Press, 2007.

Hohberg M, Knöchel J, Hoffmann CJ, Chlench S, Wunderlich W, Alter A, Maroski J, Vonderwülbecke BJ, Da Silva-Azevedo L, Knudsen R, Lehmann R, Fiedorowicz K, Bongrazio M, Nitsche B, Hoepfner M, Styp-Rekowska B, Pries AR, Zakrzewicz A. Expression of ADAMTS1 in endothelial cells is induced by shear stress and suppressed in sprouting capillaries. *J Cell Physiol* 226:360-361, 2011.

Johnsen AK, Valdar W, Golden L, Ortiz-Lopez A, Hitzemann R, Flint J, Mathis D, Benoist C. Genome-wide and species-wide dissection of the genetics of arthritis severity in heterogeneous stock mice. *Arthritis & Rheumatism* 63:2630-2640, 2011.

King H, Rewers M. Global estimates for prevalence of diabetes mellitus and glucose tolerance in adults. *Diabetes Care* 16:157-177, 1993.

Matsuda Y, Saegusa H, Zong S, Noda T, Tanabe T. Mice lacking Ca(v)2.3 (alpha1E) calcium channel exhibit hyperglycemia. *Bichem Biophys Res Commun* 289:791-795, 2001.

Mitani H, Takahashi I, Onodera K, Bae JW, Sato T, Takahashi N, Sasano Y, Igarashi K. Comparison of age-dependent expression of aggrecan and ADAMTSs in mandibular condylar cartilage, tibial growth plate, and articular cartilage in rats. *Histochem Cell Biol* 126:371-380, 2006.

Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A, Prokopenko I, Kang HM, Dina C, Esko T, Fraser RM, Kanoni S, Kumar A, Lagou V, Langenberg C, Luan J, Lindgren CM, Müller-Nurasyid M, Pechlivanis S, Rayner NW, Scott LJ, Wiltshire S, Yengo L, Kinnunen L, Rossin EJ, Raychaudhuri S, Johnson AD, Dimas AS, Loos RJ, Vedantam S, Chen H, Florez JC, Fox C, Liu CT, Rybin D, Couper DJ, Kao WH, Li M, Cornelis MC, Kraft P, Sun Q, van Dam RM, Stringham HM, Chines PS, Fischer K, Fontanillas P, Holmen OL, Hunt SE, Jackson AU, Kong A, Lawrence R, Meyer J, Perry JR, Platou CG, Potter S, Rehnberg E, Robertson N, Sivapalaratnam S, Stančáková A, Stirrups K, Thorleifsson G, Tikkanen E, Wood AR, Almgren P, Atalay M, Benediktsson R, Bonnycastle LL, Burtt N, Carey J, Charpentier G, Crenshaw AT, Doney AS, Dorkhan M, Edkins S, Emilsson V, Eury E, Forsen T, Gertow K, Gigante B, Grant GB, Groves CJ, Guiducci C, Herder C, Hreidarsson AB, Hui J, James A, Jonsson A, Rathmann W, Klopp N, Kravic J, Krjutškov K, Langford C, Leander K, Lindholm E, Lobbens S, Männistö S, Mirza G, Mühleisen TW, Musk B, Parkin M, Rallidis L, Saramies J, Sennblad B, Shah S, Sigurðsson G, Silveira A, Steinbach G, Thorand B, Trakalo J, Veglia F, Wennauer R, Winckler W, Zabaneh D, Campbell H, van Duijn C, Uitterlinden AG, Hofman A, Sijbrands E, Abecasis GR, Owen KR, Zeggini E, Trip MD, Forouhi NG, Syvänen AC, Eriksson JG, Peltonen L, Nöthen MM, Balkau B, Palmer CN, Lyssenko V, Tuomi T, Isomaa B, Hunter DJ, Qi L, Shuldiner AR, Roden M, Barroso I, Wilsgaard T, Beilby J, Hovingh K, Price JF, Wilson JF, Rauramaa R, Lakka TA, Lind L, Dedoussis G, Njølstad I, Pedersen NL, Khaw KT, Wareham NJ, Keinanen-Kiukaanniemi SM, Saaristo TE, Korpi-Hyövälti E, Saltevo J, Laakso M, Kuusisto J, Metspalu A, Collins FS, Mohlke KL, Bergman RN, Tuomilehto J, Boehm BO, Gieger C, Hveem K, Cauchi S, Froguel P, Baldassarre D, Tremoli E, Humphries SE, Saleheen D, Danesh J, Ingelsson E, Ripatti S, Salomaa V, Erbel R, Jöckel KH, Moebus S, Peters A, Illig T, de Faire U, Hamsten A, Morris AD, Donnelly PJ, Frayling TM, Hattersley AT, Boerwinkle E, Melander O, Kathiresan S, Nilsson PM, Deloukas P, Thorsteinsdottir U, Groop LC, Stefansson K, Hu F, Pankow JS, Dupuis J, Meigs JB, Altshuler D, Boehnke M, McCarthy MI. Large-scale association analysis provides insight into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 44:981-990, 2012.

Mott R, Talbot CJ, Turri MG, Collins AC and Flint J. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci* 97:12640-12654, 2001.

Nichols GA, Hillier TA, and Brown JB. Normal fasting plasma glucose and risk of type 2 diabetes diagnosis. *Am J Med* 121: 519-524, 2008.

Nishio YW, Buczek-Thomas CE, Rulfs JA, Koya J, Aiello D, Feener LP, Miller EP, Dennis JR, King GL. Identification and characterization of a gene regulating enzymatic glycosylation which is induced by diabetes and hyperglycemia in rat cardiac tissue. *J Clin Invest* 96:1759-1767, 1995.

Nobrega MA, Woods LC, Fleming S, Jacob HJ. Distinct genetic regulation of progression of diabetes and renal disease in the Goto-Kakizaki rat. *Physiol Genomics* 39:38-46, 2009.

Ogino T, Moralejo DH, Toide K, Wei S, Wei K, Yamada T, Mizuno A, Matsumoto K, Shima K. Identification of possible quantitative trait loci responsible for hyperglycemia after 70% pancreatectomy using a spontaneously diabetogenic rat. *Genet Res* 73:29-3, 1999.

RGD ID 631841. 634341, 2316949, 2140, Rat Genome Database Web Site, Medical College of Wisconsin, Milwaukee, Wisconsin. World Wide Web (http://rgd.mcw.edu). 03/2013.

Scott RA et al. Large-scale association analyses indentify new genetic loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 44:991-1005, 2012.

Solberg Woods LC, Holl K, M Tschannen, Valdar W. Fine-mapping a locus for glucose tolerance using heterogeneous stock rats. *Physiol Genomics* 41:102-108, 2010.

Solberg Woods LC, Holl K, Oreper D, Xie Y, Tsaih S, Valdar W. Fine-mapping diabetes traits, including insulin resistance, in heterogeneous stock rats. *Physiol Genomics* 44:1013-1026, 2012.

Valdar W, Holmes CC, Mott R, Flint J. Mapping in structured populations by resample model averaging. *Genet* 182:1263-1277, 2009.

Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JNP, Mott R, Flint J. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* 38:879-887, 2006.

Venables WN, Ripley BD. *Modern Applied Statistics.* New York: Springer, 2002.

Vissher PM, Thompson R, Haley CS. Confidence intervals in QTL mapping by bootstrapping. *Genet* 143:1013-1020, 1996.

Watanabe TK, Okuno S, Oga K, Mizoguchi-Miyakita A, Tsuji A, Yamasaki Y, Hishigaki H, Kanemoto N, Takagi T, Takahashi E, Irie Y, Nakamura Y, Tanigami A. Genetic dissection of "OLETF," a rat model for non-insulin-dependent diabete mellitus: quantitative trait locus analysis of (OLETF x BN) x OLETF. *Genomics* 58:233-239, 1999.

Yokoi N, Kanazawa M, Kitada K, Tanaka A, Kanazawa Y, Suda S, Ito H, Serikawa T, Komeda K. A non-HMC locus essential for autoimmune type 1 diabetes in the Komeda diabetes-prone rat. *J Clin Invest* 100: 2015-2021, 1997.

Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Boström KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jørgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjögren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638-645, 2008.

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of genome-wide association signal in UK samples reveals risk loci for type 2 diabetes. *Science* 316:1335-1341, 2007.

Zimmet PZ, Shaw JE, Sicree RA. Global estimates of the prevalence of diabetes for 2012 and 2030. *Diabetes Res and Clin Prac* 87:4-14, 2010.