Bringing Computational Biology to the High School Classroom to Narrow QTL

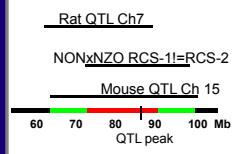
Deborah McGann, Randy Smith, Hyuna Yang, Sharon Tsaih, Keith Shockley, Gary Churchill

The Jackson Laboratory , Bar Harbor, Maine 04609 USA

Abstract

To narrow quantitative trait loci (QTL) associated with Obesity Related Diabetes (Diabesity), two areas of the mouse genome were interrogated using multiple applications of existing bioinformatics tools with a graded score assigned to each, narrowing the list of most likely candidates by a factor of ten, from 600 to less than 60. The techniques rely on phenotypic expression and sequence information where points are assigned for 1) identical QTL in homologous regions in other species, 2) Human Genome-Wide Association Study (GWAS), 3) known gene function via ontologies, 4) expression of mRNA in relevant tissue, 5) differential mRNA expression between strains. Additional points were assigned based on evaluation of 6) haplotype blocks, 7) Haplotype Association Mapping (HAM), and 8) potential functional change predicted by single nucleotide polymorphism (SNP). This extensive and objective approach for applying appropriate weight to diverse sources of evidence reduces the bias resulting from over reliance on existing knowledge of the genes in the region. Results implicate not only genes that have been associated with diabetes such as Recql4, Atxn10, and Ppara, but also others that deserve further consideration.

1) Comparative



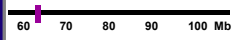
Overlapping Regions of Various Mammalian Diabetes QTL Mapped on *Mus musculus* Chromosome 15

1 pt. for region of Rat QTL
2 pt. in region different for RCS-1(+Dbsty) and RCS-2 (-Dbsty)

2) GWAS

ST3Gal1 Human Chromosome 8 in Scandinavian Population rs7016568 @ 134674046 bp

pValue
Initial Analysis 1.7×10^{-4}
MetaAnalysis 0.104



Human Genome-Wide Association Study (GWAS) genes in Dbsty4 Region

2 pt. (near)
1 pt. (within 1.0Mb)

Introduction

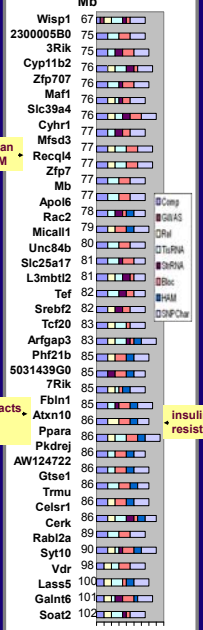
Analysis of a NZOxNON reciprocal backcross suggested the presence of QTL on eight different chromosomes contributing to body weight, plasma insulin, plasma glucose, BMI, and leptin (Reifsnyder, Churchill et al. 2000). While multiple interactions were identified between chromosomes and maternal environment, links from QTL to genes have been illusive. Outlined here is an approach to integrate multiple bioinformatic tools in an attempt to narrow the diabesity QTL (diabetes + obesity) on mouse chromosome 15:63-102 Mb.

References

Burgess-Herbert, S., K. Shockley, S. Sheehan, L. Bickerstaff, B. Harwood IV, Y. Shen, R. Li, G. Churchill, and B. Paigen. Comparative Genomics, Experimental Biology, and Bioinformatics Merge to Narrow a QTL for HDL Cholesterol on Mouse Chromosome 15 and Human Chromosome 8 (In preparation)

Reifsnyder, P. C., G. Churchill, et al. (2000). "Maternal environment and genotype interact to establish diabesity in mice." Genome Res 10(10): 1568-1578.

Results Dbsty4 (Chromosome 15)



3) Ontologies

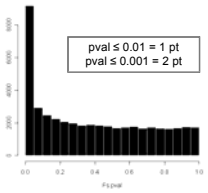
2 pts. Gene Ontology (GO)

biosynthetic calcium carbohydrate cholesterol fatty glucose glycolysis insulin lipid triglyceride zinc

3 pts. Mammalian Phenotypes (MP)

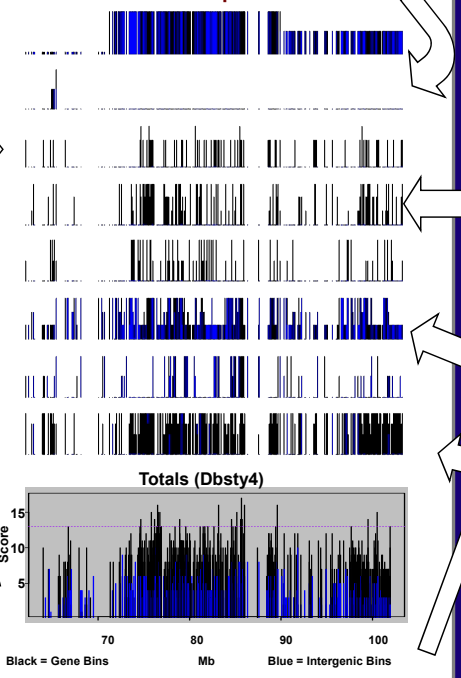
abnormal glucose homeostasis
increased circulating glucose levels
insulin resistance
abnormal glucose tolerance
abnormal adipocyte glucose uptake

5) Strain Expressions

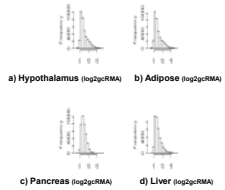


Histogram of Fs pval for ANOVA test of Strain:Diet Interaction in 3 replications of 10 strains of male mice fed either ordinary chow or high fat diet.

Example of Discrimination of Eight Techniques



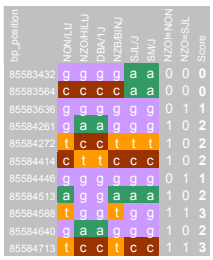
4) Tissue Expression



Histogram of $\log_2$ gcRMA as a measure of Gene Expression in C57B6 mice.

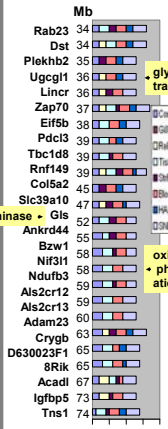
3x Median =1 pt Total Score = Sum-1

Figure 6 : Haplotypes



Maximum score for all SNPS within each bin was assigned to the bin

Results Dbsty1 (Chromosome 1)



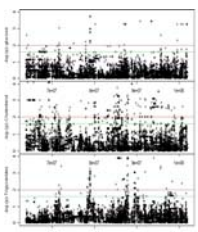
Conclusions

The results of this objective approach, simultaneously using multiple techniques, point toward genes already associated with glucose metabolism and others that have no known relationship. Both should be given further consideration as key contributory factors of diabesity.

The advantage of the first three, literature-based, techniques is that work has already been completed and reviewed. However, to focus too heavily on these three is to risk missing some as yet undiscovered gene function. The remaining five techniques, utilizing databases of evidence of gene action and impact of molecular differences, constitute vast untapped resources which may either support the literature or point in some alternate direction.

As a methodology, the two critical questions that remain are whether the techniques are sufficient and, equally important, whether or not the weighting of the techniques is appropriate.

7) HAM



Haplotype Association Mapping of 30 inbred male strains at 13-14 weeks

-logp 1.6 -2.0 = 1 pt. ≥2.0 = 2 pts. # groups > 4

-logp

8) SNP Characteristics

SNP Functional Class	Score	Count
Intronic	2	53985
3'UTR or 5'UTR	2	1707
Exon Coding Synonymous	3	1616
Exon Coding Non-Synonymous	4	537
Start Gained	4	22
Start Lost	4	24
Stop Gained	4	6

Evaluates impact potential within protein coding regions

Multiple Transcripts-Multiple SNPS in Gene-Use highest

Acknowledgments

Arthur Vining Davis Foundations
Howard Hughes Medical Institute (HHMI)
Bank of America/Balfour Foundation
National Institute of General Medical Sciences (NIGMS)
Center for Genome Dynamics (CGD)

Maine School of Science and Mathematics (MSSM)
North Carolina School of Science and Mathematics (NCSSM)

Gary Churchill's High School Class:
MSSM Students: Renee Symonds, Ryan Keating, and Chunan Liang
NCSSM Students: Elizabeth Deerhake, Cory Li, and Lisa Zhang