eQTL Tools
a collaboration in progress

Brian Yandell & Bioinformatics Team
Attie Lab, UW-Madison
1 jul 2009

experimental context

- B6 x BTBR obese mouse cross
  - model for diabetes and obesity
  - 500+ mice from intercross (F2)
  - collaboration with Rosetta/Merck
- genotypes
  - 5K SNP Affymetrix mouse chip
  - care in curating genotypes! (map version, errors, …)
- phenotypes
  - clinical phenotypes (>100 / mouse)
  - gene expression traits (>40,000 / mouse / tissue)
  - other molecular phenotypes

how does one filter traits?

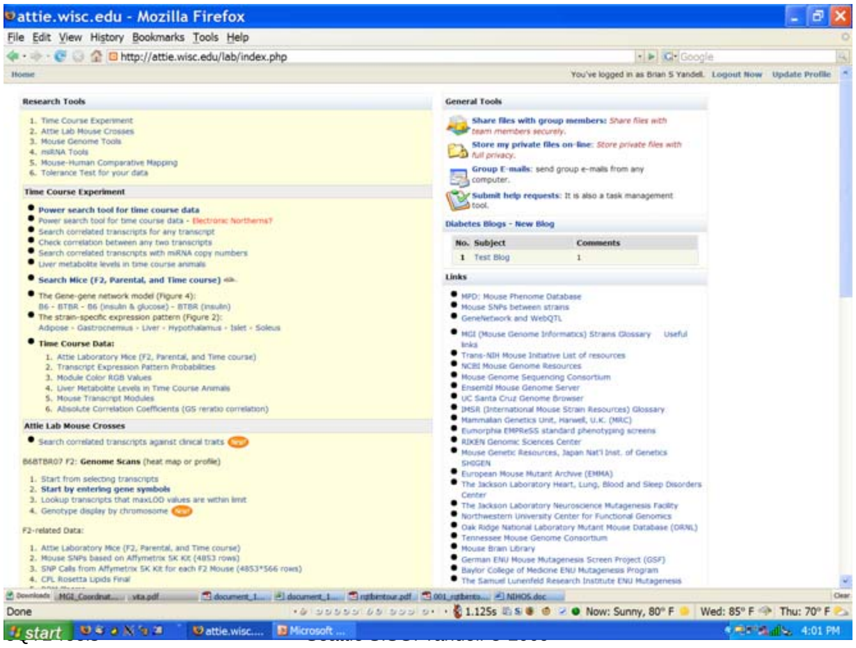
- want to reduce to “manageable” set
  - 10/100/1000: depends on needs/tools
  - How many can the biologist handle?
- how can we create such sets?
  - data-driven procedures
    - correlation-based modules
      - Zhang & Horvath 2005 *SAGMB*, Keller et al. 2008 *Genome Res*
      - Li et al. 2006 *Hum Mol Gen*
    - mapping-based focus on genome region
  - function-driven selection with database tools
    - GO, KEGG, etc
    - Incomplete knowledge leads to bias
  - random sample

why build Web eQTL tools?

- common storage/maintainence of data
  - one well-curated copy
  - central repository
  - reduce errors, ensure analysis on same data
- automate commonly used methods
  - biologist gets immediate feedback
  - statistician can focus on new methods
  - codify standard choices

how does one build tools?

- no one solution for all situations
- use existing tools wherever possible
  - new tools take time and care to build!
  - downloaded databases must be updated regularly
- human component is key
  - need informatics expertise
  - need continual dialog with biologists
- build bridges (interfaces) between tools
  - Web interface uses PHP
  - commands are created dynamically for R
- continually rethink & redesign organization



You've logged in as Brian S Yandell. Logout Now Update Profile

Research Tools

1. Time Course Experiment
2. Attie Lab Mouse Crosses
3. Mouse Genome Tools
4. miRNA Tools
5. Mouse-Human Comparative Mapping
6. Tolerance Test for your data

Time Course Experiment

- **Power search tool for time course data**
- Power search tool for time course data - Electronic Northerns?
- Search correlated transcripts for any transcript
- Check correlation between any two transcripts
- Search correlated transcripts with miRNA copy numbers
- Liver metabolite levels in time course animals
- **Search Mice (F2, Parental, and Time course)**
- The Gene-gene network model (Figure 4):
  B6 - BTBR - B6 (insulin & glucose) - BTBR (insulin)
- The strain-specific expression pattern (Figure 2):
  Adipose - Gastrocnemius - Liver - Hypothalamus - Islet - Soleus
- **Time Course Data:**
  1. Attie Laboratory Mice (F2, Parental, and Time course)
  2. Transcript Expression Pattern Probabilities
  3. Module Color RGB Values
  4. Liver Metabolite Levels in Time Course Animals
  5. Mouse Transcript Modules
  6. Absolute Correlation Coefficients (GS reratio correlation)

Attie Lab Mouse Crosses

- Search correlated transcripts against clinical traits New!

B6BTBR07 F2: **Genome Scans** (heat map or profile)

1. Start from selecting transcripts
2. **Start by entering gene symbols**
3. Lookup transcripts that maxLOD values are within limit
4. Genotype display by chromosome New!

F2-related Data:

1. Attie Laboratory Mice (F2, Parental, and Time course)
2. Mouse SNPs based on Affymetrix 5K Kit (4853 rows)
3. SNP Calls from Affymetrix 5K Kit for each F2 Mouse (4853*566 rows)
4. CPL Rosetta Lipids Final

General Tools

Share files with group members: *Share files with team members securely.*

Store my private files on-line: *Store private files with full privacy.*

Group E-mails: send group e-mails from any computer.

Submit help requests: It is also a task management tool.

Diabetes Blogs - New Blog

No.	Subject	Comments
1	Test Blog	1

Links

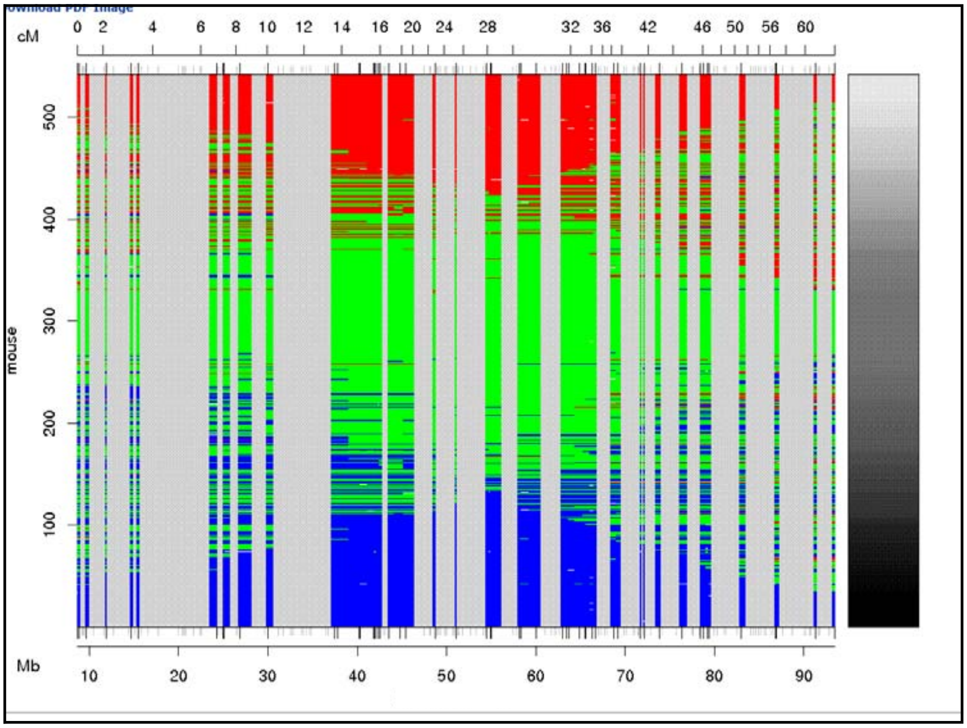
- MPD: Mouse Phenome Database
- Mouse SNPs between strains
- GeneNetwork and WebQTL
- MGI (Mouse Genome Informatics) Strains Glossary Useful links
- Trans-NIH Mouse Initiative List of resources
- NCBI Mouse Genome Resources
- Mouse Genome Sequencing Consortium
- Ensembl Mouse Genome Server
- UC Santa Cruz Genome Browser
- IMSR (International Mouse Strain Resources) Glossary
- Mammalian Genetics Unit, Harwell, U.K. (MRC)
- Eumorphia EMPReSS standard phenotyping screens
- RIKEN Genomic Sciences Center
- Mouse Genetic Resources, Japan Nat'l Inst. of Genetics SHIGEN
- European Mouse Mutant Archive (EMMA)
- The Jackson Laboratory Heart, Lung, Blood and Sleep Disorders Center
- The Jackson Laboratory Neuroscience Mutagenesis Facility
- Northwestern University Center for Functional Genomics
- Oak Ridge National Laboratory Mutant Mouse Database (ORNL)
- Tennessee Mouse Genome Consortium
- Mouse Brain Library
- German ENU Mouse Mutagenesis Screen Project (GSF)
- Baylor College of Medicine ENU Mutagenesis Program
- The Samuel Lunenfeld Research Institute ENU Mutagenesis

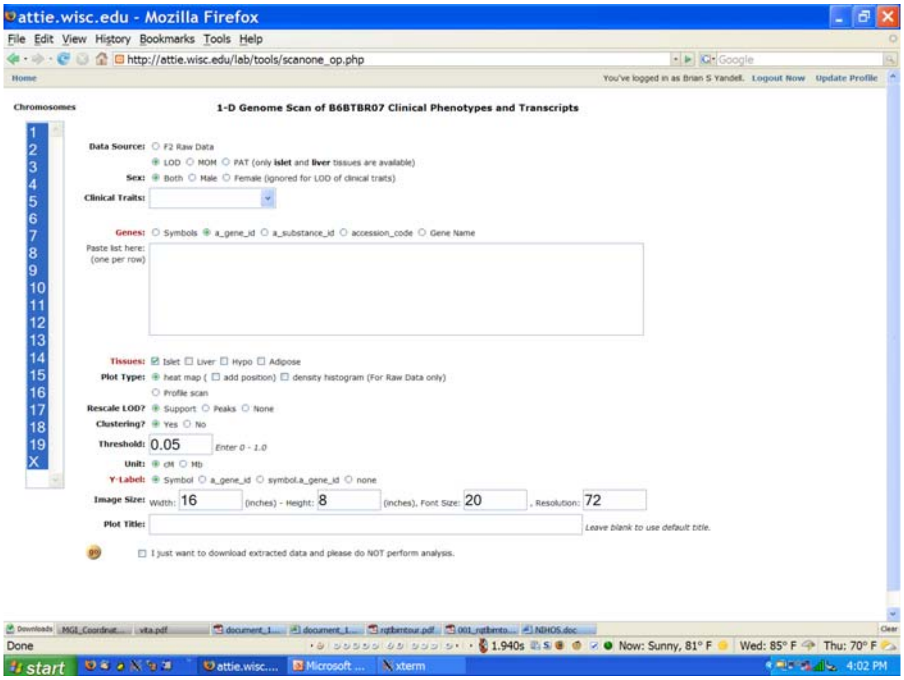
steps in using Web tools

- user enters data on Web page
- PHP tool interprets user data
- PHP builds R script
- R run on script
  - creates plots, summaries, warnings
- PHP grabs results & displays on page
- user examines, saves
- user modifies data and reruns

raw data or fancy results?

- raw data flexible but slow
  - LOD profiles for 100 (1000) traits?
- fancy results from sophisticated analysis
  - IM, MIM, BIM, MOM analysis
  - too complicated to put in biologists’ hands?
    - methods are unrefined, state-of-art, research tools
    - use of methods involved many subtle choices
  - batch computation over weeks
    - compute once, save, display many times





1-D Genome Scan of B6BTBR07 Clinical Phenotypes and Transcripts

Data Source: ○ F2 Raw Data
● LOD ○ MOM ○ PAT (only **islet** and **liver** tissues are available)

Sex: ● Both ○ Male ○ Female (ignored for LOD of clinical traits)

Clinical Traits:

Genes: ○ Symbols ● a_gene_id ○ a_substance_id ○ accession_code ○ Gene Name

Paste list here: (one per row)

Tissues: ☑ Islet ☐ Liver ☐ Hypo ☐ Adipose

Plot Type: ● heat map (☐ add position) ☐ density histogram (For Raw Data only)
○ Profile scan

Rescale LOD? ● Support ○ Peaks ○ None

Clustering? ● Yes ○ No

Threshold: 0.05 *Enter 0 - 1.0*

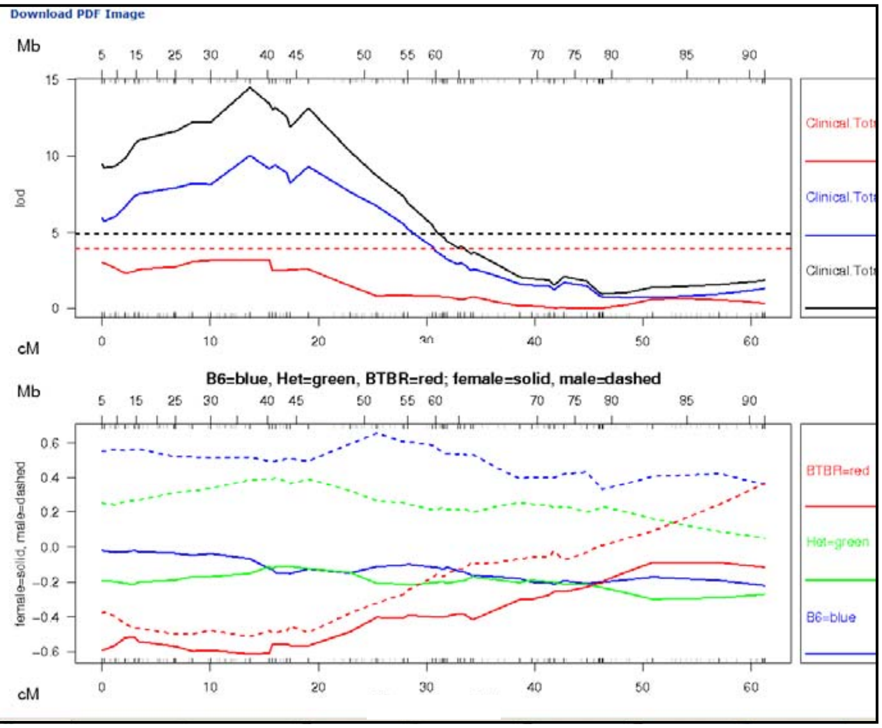
Unit: ● cM ○ Mb

Y-Label: ● Symbol ○ a_gene_id ○ symbol.a_gene_id ○ none

Image Size: Width: 16 (inches) - Height: 8 (inches), Font Size: 20 , Resolution: 72

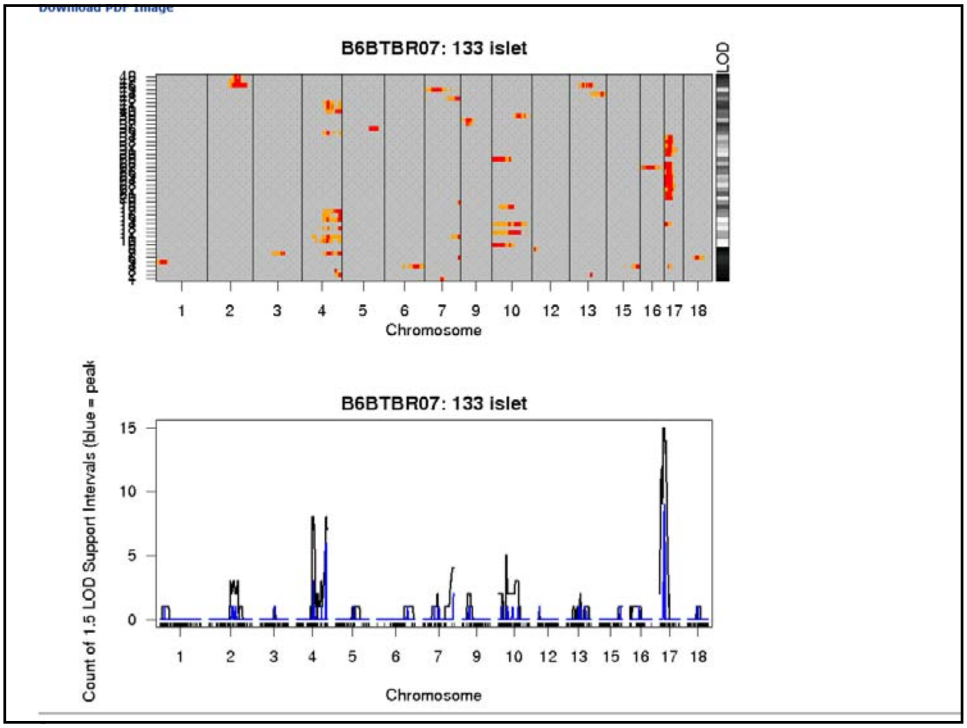
Plot Title: *Leave blank to use default title.*

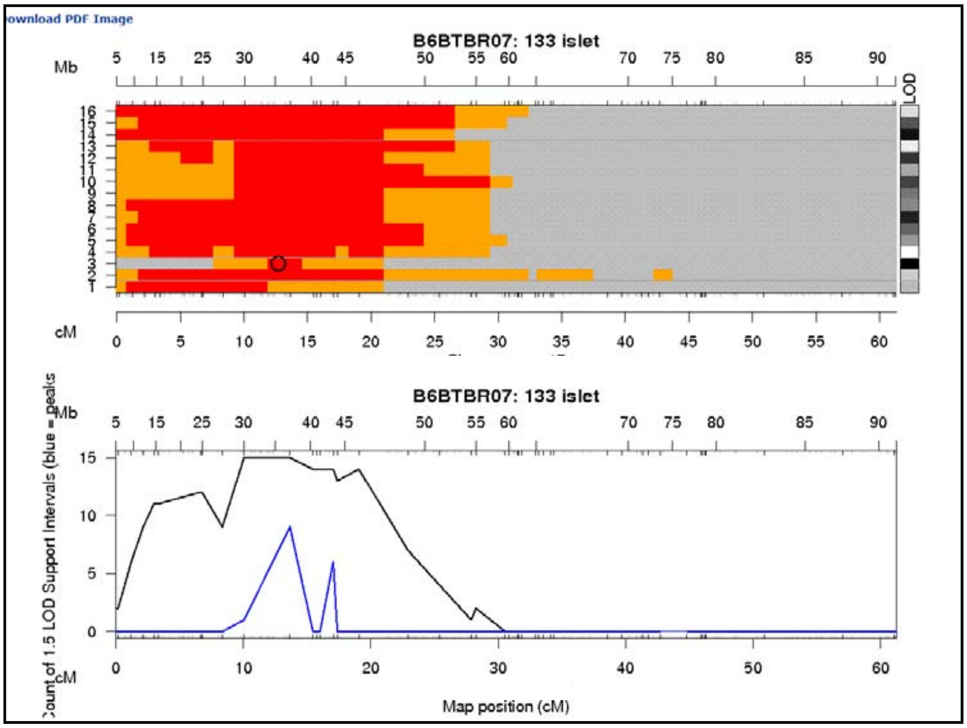
go ☐ I just want to download extracted data and please do NOT perform analysis.

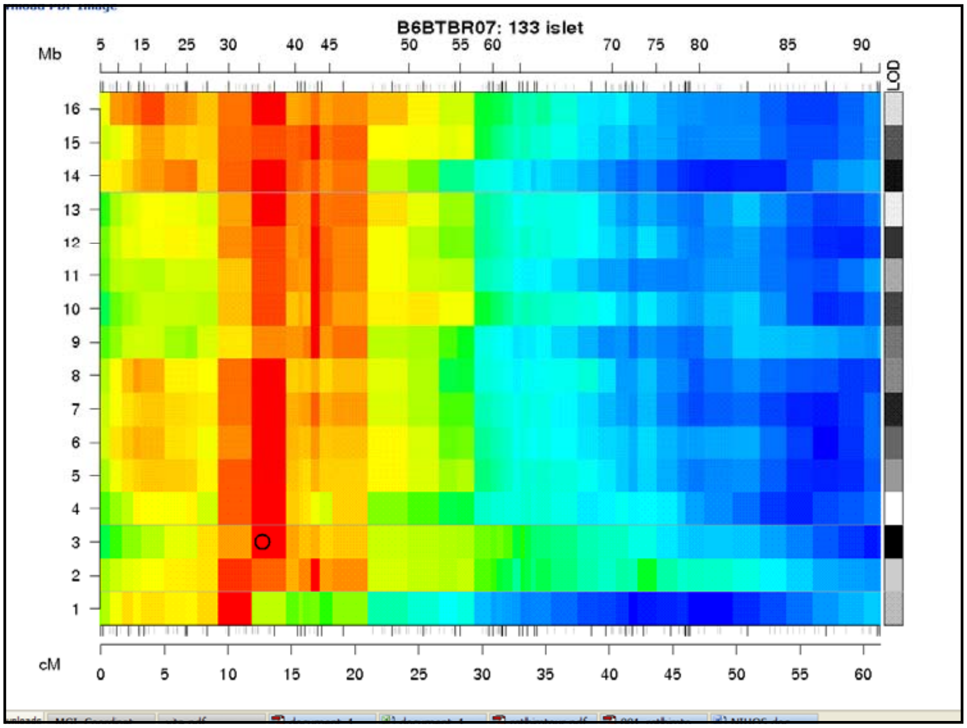


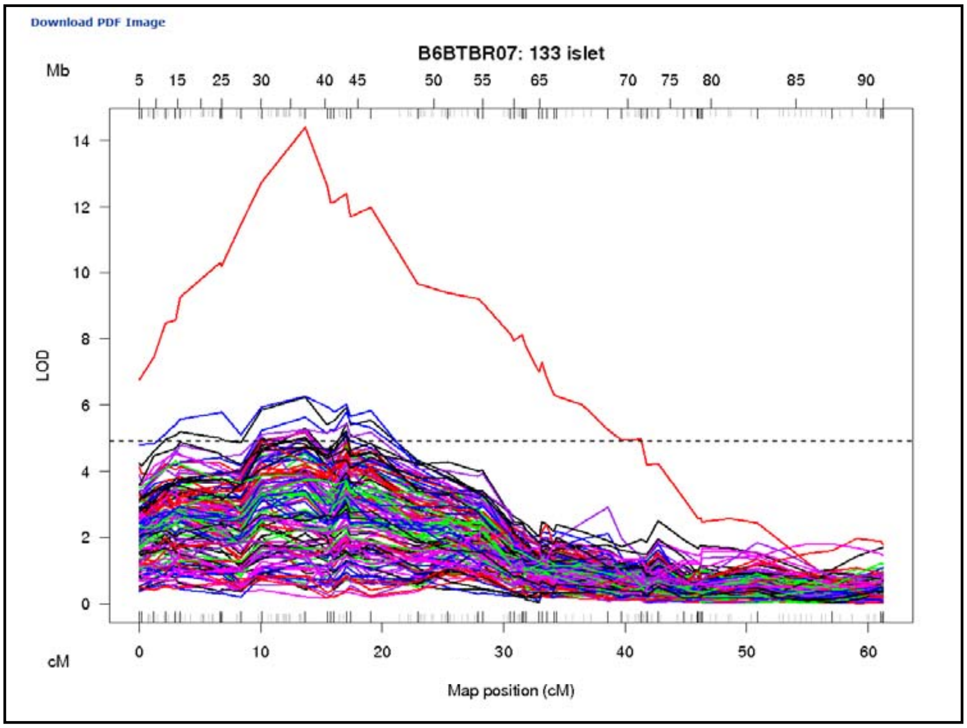
automated R script

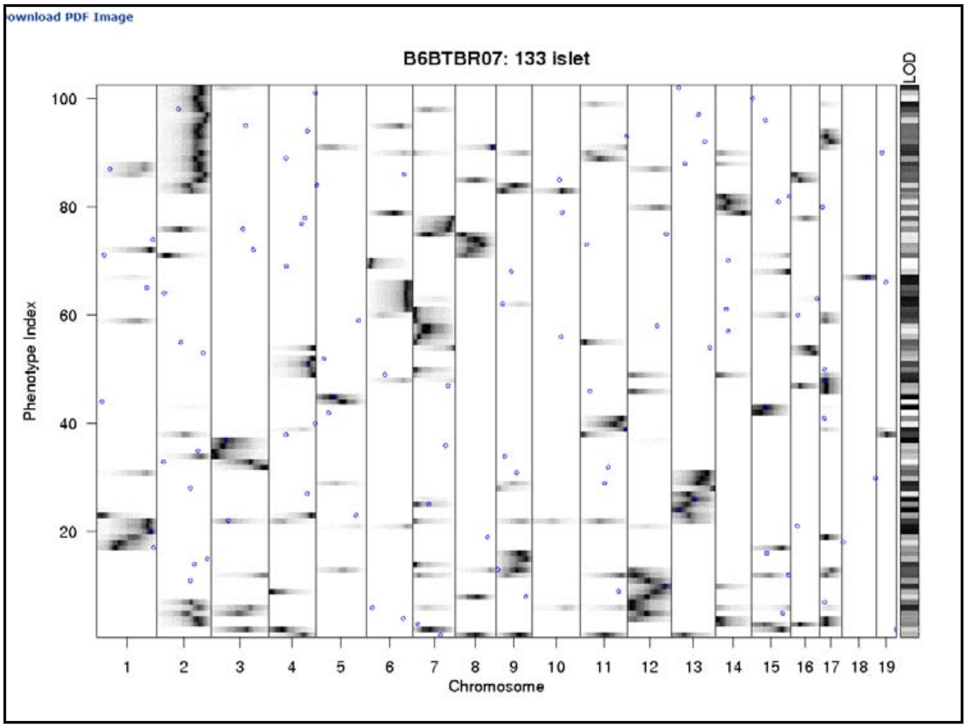
```
library('B6BTBR07')

out <- multtrait(cross.name='B6BTBR07',
  filename = 'scanone_1214952578.csv',
  category = 'islet', chr = c(17),
  threshold.level = 0.05, sex = 'both',)

sink('scanone_1214952578.txt')
print(summary(out))
sink()

bitmap('scanone_1214952578%03d.bmp',
  height = 12, width = 16, res = 72, pointsize = 20)
plot(out, use.cM = TRUE)
dev.off()
```

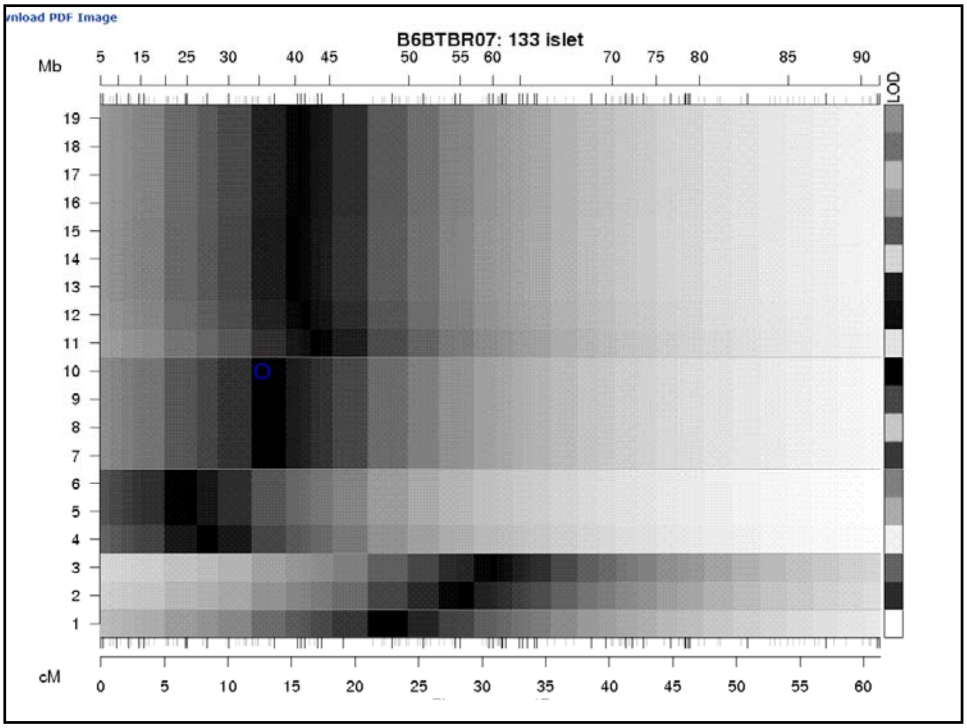

B6BTBR07: 133 islet
LOD
1 2 3 4 5 6 7 9 10 12 13 15 16 17 18
Chromosome
B6BTBR07: 133 islet
Count of 1.5 LOD Support Intervals (blue = peak
0
5
10
15
Chromosome

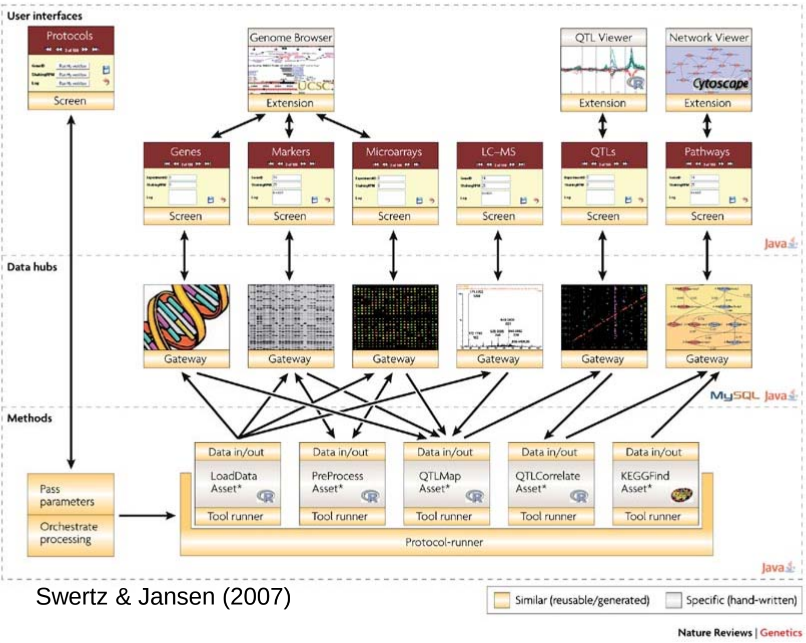

ownload PDF Image
B6BTBR07: 133 islet
Mb
5 15 25 30 40 45 50 55 60 70 75 80 85 90
LOD
16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
cM
0 5 10 15 20 25 30 35 40 45 50 55 60
B6BTBR07: 133 islet
Count of 1.5 LOD Support Intervals (blue = peaks
Mb
0
5
10
15
cM
0 10 20 30 40 50 60
Map position (cM)


B6BTBR07: 133 islet
Mb
5 15 25 30 40 45 50 55 60 70 75 80 85 90
LOD
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
cM
0 5 10 15 20 25 30 35 40 45 50 55 60


Download PDF Image
B6BTBR07: 133 islet
Mb
5 15 25 30 40 45 50 55 65 70 75 80 85 90
LOD
14
12
10
8
6
4
2
0
cM
0 10 20 30 40 50 60
Map position (cM)


ownload PDF Image
B6BTBR07: 133 islet
LOD
Phenotype Index
100
80
60
40
20
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Chromosome


nload PDF Image
B6BTBR07: 133 islet
Mb
5
15
25
30
40
45
50
55
60
70
75
80
85
90
LOD
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
cM
0
5
10
15
20
25
30
35
40
45
50
55
60



Swertz & Jansen (2007)

eÇ