Introduction to GWAS and QTL analysis

Lars Rönnegård

Swedish University of Agricultural Sciences

Dalarna University

NOVA course, August 2011
This is where I work

Uppsala
Borlänge
To find out more

Lars Rönnegård
lrn@du.se
www.larsronnegard.se
Outline

• Introduction to Genome-Wide Association Studies (GWAS)
• Analysis of Quantitative Trait Loci (QTL)
GWAS and QTL plots

Manhattan plot
Weedon et al. 2008, Nature Genetics 40, 575 - 583

QTL scan on chicken chromosome 1
(figure from Kerje et al. 2003, Animal Genetics)
Example: 3 individuals and 5 SNPs

<table>
<thead>
<tr>
<th>Individual 1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 cm</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>182 cm</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual 2</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 cm</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>182 cm</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual 3</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 cm</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>182 cm</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Linear model $y = \mu + x_j b + e$ with $y = \begin{pmatrix} 175 \\ 182 \\ 195 \end{pmatrix}$
GWAS - simplified example

Example: 3 individuals and 5 SNPs

Linear model $\mathbf{y} = \mu + \mathbf{x}_j \mathbf{b} + \mathbf{e}$ with $\mathbf{y} = \begin{pmatrix} 175 \\ 182 \\ 195 \end{pmatrix}$
Example: 3 individuals and 5 SNPs

Linear model $y = \mu + x_j b + e$ with $y = \begin{pmatrix} 175 \\ 182 \\ 195 \end{pmatrix}$ and $z_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$
GWAS - simplified example

Example: 3 individuals and 5 SNPs

<table>
<thead>
<tr>
<th>Individual 1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual 2</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>182 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Individual 3</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>195 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Linear model $y = \mu + x_j b + e$ with $y = \begin{pmatrix} 175 \\ 182 \\ 195 \end{pmatrix}$ and $x_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$

and $\hat{b} = 6.5$ ($P = 0.56$)
GWAS - simplified example

Example: 3 individuals and 5 SNPs

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>175 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>182 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>195 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Linear model $y = \mu + x_j b + e$ with $y = \begin{pmatrix} 175 \\ 182 \\ 195 \end{pmatrix}$ and $x_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$

Calculate P-value for each $b$ and plot $-\log_{10} P$
GWAS - simplified example

Example: 3 individuals and 5 SNPs

Linear model \( \mathbf{y} = \mu + \mathbf{x}_j \mathbf{b} + \mathbf{e} \) with
\[
\mathbf{y} = \begin{pmatrix} 175 \\ 182 \\ 195 \end{pmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}
\]

Calculate P-value for each \( \mathbf{b} \) and plot \(-\log_{10} P\)
GWAS plot

GWAS
QTL analysis
Significance testing
Assumptions in QTL analysis
Summary

GWAS plot
Manhattan plot

Example from: Weedon et al. 2008, Nature Genetics 40, 575 - 583
GWAS software

- **PLINK**
  - Purcell et al. 2007 PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81:559-575.
  - available at: http://pngu.mgh.harvard.edu/~purcell/plink/
  - open source C++

- **GenABEL**
  - R package available on CRAN
Different QTL methods

- Haley-Knott Regression
- Maximum Likelihood
  - Lander & Botstein 1989 Genetics 121:185–199
  - Kao 2000 Genetics 156: 855–865
- Bayesian Methods
- Variance Component Method
  - Goldgar 1990 Am J Hum Genet 47:957-967
Example: F2 -cross from Red Jungle fowl - Domestic White Leghorn hens. Body weight at 200 days of age
GWAS & QTL

QTL - simplified example

Jungle fowl

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ind. 1

Domestic hens

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ind. 2 ind. 3 ind. 4
QTL - simplified example

$F_1$ individuals
QTL - simplified example

$F_1$ individuals

"Crossing-over" during meiosis

Recombined egg/sperm cells
QTL - simplified example

$F_2$ individuals
QTL - simplified example

F₂ individuals

850g | 907g | 873g | 756g | 1066g

GWAS & QTL
QTL - simplified example
QTL - simplified example

<table>
<thead>
<tr>
<th></th>
<th>850g</th>
<th>907g</th>
<th>873g</th>
<th>756g</th>
<th>1066g</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Rönnegård GWAS & QTL
QTL - simplified example

<table>
<thead>
<tr>
<th>y</th>
<th>X</th>
<th>850g</th>
<th>907g</th>
<th>873g</th>
<th>756g</th>
<th>1066g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Rönnegård
QTL - simplified example

<table>
<thead>
<tr>
<th>X</th>
<th>0</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>850g</td>
<td>907g</td>
<td>873g</td>
<td>756g</td>
<td>1066g</td>
</tr>
</tbody>
</table>

Rönnegård GWAS & QTL
QTL - simplified example

<table>
<thead>
<tr>
<th>y</th>
<th>850g</th>
<th>907g</th>
<th>873g</th>
<th>756g</th>
<th>1066g</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

GWAS & QTL
Example plot

QTL scan on chicken chromosome 1

(figure from Kerje et al. 2003, Animal Genetics)
QTL - simplified example

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>850g</td>
<td>907g</td>
<td>873g</td>
<td>756g</td>
<td>1066g</td>
</tr>
</tbody>
</table>

Rönnegård GWAS & QTL
QTL - simplified example

<table>
<thead>
<tr>
<th></th>
<th>850g</th>
<th>907g</th>
<th>873g</th>
<th>756g</th>
<th>1066g</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>1.3</td>
<td>1.0</td>
<td>1.1</td>
<td>0.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Rönnegård GWAS & QTL
QTL - simplified example

<table>
<thead>
<tr>
<th>y</th>
<th>850g</th>
<th>907g</th>
<th>873g</th>
<th>756g</th>
<th>1066g</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>1.3</td>
<td>1.0</td>
<td>1.1</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>0.9+0.4</td>
<td>0.1+0.9</td>
<td>0.2+0.9</td>
<td>0.2+0.4</td>
<td>0.1+0.9</td>
</tr>
</tbody>
</table>
Significance thresholds

Manhattan plot
Weedon et al. 2008, Nature Genetics 40, 575 - 583

QTL scan on chicken chromosome 1
(figure from Kerje et al. 2003, Animal Genetics)
Permutation testing

1. Permute the observations $y$ and call this new vector $\tilde{y}$
2. Fit the model $\tilde{y} = \mu + x_j b + e$ for all positions along the genome
3. Save the maximum value of the test statistic
4. Repeat 1000 times
5. Order the saved values from lowest to highest. The top 5% value gives the “5% genome-wide significance level”.

Different test statistics

- Wald test (standard error test)
- Likelihood ratio test
- Score test
Different test statistics
Assumptions in QTL analysis using the simple regression method

- Additive effects
- Markers are fixed within founder lines
- QTL are fixed within founder lines
- Two founder lines
- Normally distributed
- Independent observations
- Common residual variance for all observations
- Each position is tested separately. Would it be possible to test all positions in one go?
R/qtl and other related software in R

- **qtl**
  - QTL detection for inbred lines
  - available on CRAN
  - Broman & Sen 2009 A guide to QTL mapping with R/qtl. Springer

- **Other software using R/qtl**
  - **qtlbim**
    - Bayesian fitting and model selection using MCMC
    - available on CRAN
  - **wgaim**
    - interval mapping using random effects approach
    - available on CRAN
    - Taylor & Verbyla 2011 J Stat Soft 40(7)
  - **qtl.outbred**
    - for outbred crosses
    - available on https://r-forge.r-project.org/R/?group_id=844
    - Nelson, Shen & Carlborg 2011 BMC Reserach Notes 4:154
Other QTL software in R

- Experimental crosses having more than two founder lines
  - HAPPY
    - developed for complex crosses of inbred mice lines
    - available at: http://www.well.ox.ac.uk/happy/happyR.shtml
Summary

- The ideas behind QTL analysis and GWAS are very similar
  - Based on linear regression

- Differences
  - GWAS assumes dense markers (SNPs)
  - QTL studies (mainly) performed in experimental crosses and markers are usually sparse
  - GWAS tests at marker positions, whereas QTL analysis tests between markers

- Important concepts
  - permutation testing
  - test statistics
    - Wald Test, Likelihood Ratio Test, Score Test