Salvaging an eQTL project

Identifying and correcting sample mix-ups

Karl W Broman

Department of Biostatistics & Medical Informatics
University of Wisconsin – Madison

www.biostat.wisc.edu/~kbroman
Human vs mouse
Backcross

P₁ \times P₂ → F₁

P₁ \times F₁ → BC
Intercross
Attie project

\(~500~\text{B6} \times \text{BTBR intercross mice, all ob/ob}\)

Genotypes at 2057 SNPs (Affymetrix arrays)

Gene expression in six tissues (Agilent arrays)
  adipose
  gastrocnemius muscle
  hypothalamus
  pancreatic islets
  kidney
  liver

Numerous clinical phenotypes
  (e.g., body weight, insulin and glucose levels)

Considerable brain power
  Alan Attie, Karl Broman, Aimee Teo Broman, Danielle Greenawalt, Mark Keller, Christina Kendziorski, Amit Kulkarni, Eric Schadt, Brian Yandell, . . .
Sex and the X chr

F₂ females: R/R or B/R
F₂ males: hemizygous B or R
Genotype mix-ups
Backcross
LOD curves

Map position (cM)

LOD score

D4Mit164

BB

BA
G to P

DNA (genotype) → phenotype
G to P

DNA (genotype) → mRNA → protein → phenotype
Sex and the X chr

F₂ females: R/R or B/R
F₂ males: hemizygous B or R
E vs G

expression of 499541

Genotype at rs13476158
Genotype at rs13476158
kNN classifier

expression of 499541

Genotype at rs13476158

BB BR RR
E vs G

Genotype at rs6244221

expression of 10004035488

expression of 518187

BB

BR

RR
Genotype at rs6244221

E vs G
E vs G

Genotype at rs13478402

expression of 502129

expression of 517583

BB

BR

RR
expression of 502129

expression of 517583

Genotype at rs13478402

E vs G
Basic scheme

expression traits

transcripts

observed eQTL genotypes

eQTL

mice
Basic scheme

expression traits

observed eQTL genotypes

transcripts

mice

inferred eQTL genotypes

eQTL

mice
Basic scheme

expression traits

mice

transcripts

observed eQTL genotypes

mice

eQTL

inferred eQTL genotypes

mice

eQTL
Basic scheme

expression traits

transcripts

mice

observed eQTL genotypes

eQTL

mice

inferred eQTL genotypes

mice
Decisions

Self vs best

- Fixable
- Not found
- Good
Genotype mix-ups
<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Plate 1631

1631
Plates 1632 and 1630
Plate 1630
E vs E

expression in islet
mice
transcripts

expression in liver
mice
transcripts
E vs E

expression in islet transcripts mice

expression in liver transcripts mice
E vs E

expression in islet

expression in liver

mice

mice

transcripts

transcripts

transcript 497973

islet expression

liver expression
E vs E

expression in islet

expression in liver

transcript 512831
E vs E

expression in islet transcripts mice
expression in liver transcripts mice

transcript 507042

islet expression
liver expression

mice

mice

transcripts

transcripts

−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
−2 −1 0 1

−2.0 −1.5 −1.0 −0.5 0.0 0.5 1.0
E vs E

expression in islet transcripts mice

expression in liver transcripts mice
E vs E

expression in islet

expression in liver

Mouse3280
E vs E expression in islet
expression in liver
mice
mice
transcripts
transcripts
Mouse3598

-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
-2
-1
0
1

−2
−1
0
1
islet expression
liver expression
E vs E

expression in islet transcripts mice

expression in liver transcripts mice

Mouse3599 liver vs Mouse3598 islet

Mouse3599 liver expr vs Mouse3598 islet expr
expression in islet transcripts mice
expression in liver transcripts mice

Mouse3598 liver vs Mouse3599 islet

Mouse3598 liver expr

Mouse3599 islet expr

E vs E
E vs E

Mouse3280

Mouse3280 islet

0.80

0.80

Mouse3280 liver

-0.18

-0.29

Mouse3281

Mouse3281 islet

0.85

Mouse3281 liver

-0.15

-0.09
E vs E

Mouse3295

-2 -1 0 1 2

0.90 0.97 0.37

Mouse3295 islet

Mouse3295 liver

Mouse3296

-2 -1 0 1 2

0.88 0.35 0.43

Mouse3296 islet

Mouse3296 liver
Insulin QTL

Chromosome LOD score

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

after
before

50
Summary

- Sample mix-ups happen
- With eQTL data, we can both identify and **correct** mix-ups
- There is great value in having expression on multiple tissues
- The general idea here has wide application for high-throughput data
- Very similar to **MixupMapper** (Westra et al., Bioinformatics 27:2104–2111, 2011)
  - Multiple tissues
  - Direct tissue-tissue comparisons
  - Predict genotype rather than expression phenotype
Acknowledgments

Alan Attie  
Mark Keller

Brian Yandell

Christina Kendziorski  
Aimee Teo Broman

Eric Schadt

Danielle Greenawalt  
Amit Kulkarni

Śaunak Sen

Biochemistry, UW–Madison

Statistics and Horticulture, UW–Madison

Biostatistics & Medical Informatics, UW–Madison

Pacific Biosciences of California

Merck & Co., Inc.

UCSF

NIH: R01 GM074244, R01 DK066369