Regulatory variation

- What do trait-associated variants do?
- Genetic changes to:
  - Coding sequence **
  - Gene expression levels
  - Splice isomer levels
  - Methylation patterns
  - Chromatin accessibility
  - Transcription factor binding kinetics
  - Cell signaling
  - Protein-protein interactions

Within a population

- Damerval et al 1994
- 42/72 protein levels differ in maize
- 2D electrophoresis, eyeball spot quantitation
- Problems:
  - genome coverage
  - quantitation
  - post-translational modifications
- Solution: use expression levels instead!
Genetics of gene expression (eQTL)

- **cis-eQTL**
  - The position of the eQTL maps near the physical position of the gene.
  - Promoter polymorphism?
  - Insertion/Deletion?
  - Methylation, chromatin conformation?

- **trans-eQTL**
  - The position of the eQTL does not map near the physical position of the gene.
  - Regulator?
  - Direct or indirect?

Modified from Cheung and Spielman 2009 Nat Gen
Yeast

- Brem et al. Science 2002
- Linkage in 40 offspring of lab x wild strain cross
- 1528/6215 DE between parents
- 570 map in cross
  - multiple QTLs
  - 32% of 570 have cis linkage
- 262 not DE in parents also map

**trans** hotspots

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of messages</th>
<th>Common function</th>
<th>Linkage bin</th>
<th>Putative regulator</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>Budding daughter</td>
<td>II:550000</td>
<td>C1773</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>Cell separation</td>
<td>III:90000</td>
<td>LEU2</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>Leucine biosynthesis</td>
<td>III:190000</td>
<td>MMT</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>Ubiquitination</td>
<td>V:110000</td>
<td>LAM3</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>Heme, fatty acid</td>
<td>X:670000</td>
<td>HAP3</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>Subcellular</td>
<td>X:1830000</td>
<td>SRE3</td>
</tr>
<tr>
<td>7</td>
<td>94</td>
<td>Encoded helicases</td>
<td>XIV:490000</td>
<td>Unknown</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>Mn2/4-dependent</td>
<td>XV:170000</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Brem et al. Science 2002

Mammals I

- F2 mice on atherogenic diet
- Expression arrays; WG linkage

Yvert et al. Nat Genet 2003

Mammals II

- No major trans loci in humans
  - Cheung et al. Nature 2003
  - Monks et al. AJHG 2004

Mammals III

Chesler et al. Nat Genet 2005
WHERE ARE THE TRANS eQTLs?

Issues with trans mapping

- **Power**
  - Genome-wide significance is $5 \times 10^{-8}$
  - Multiple testing on ~20K genes
  - Sample sizes clearly inadequate
- **Data structure**
  - Bias corrections deflate variance
  - Non-normal distributions
- **Sample sizes**
  - Far too small

But...

- Assume that trans eQTLs affect many genes...
- ...and you can use cross-trait methods!

Association data

\[
\begin{bmatrix}
    z_{1,1} & z_{1,2} & \ldots & z_{1,p} \\
    z_{2,1} & & & \\
    \vdots & & & \\
    z_{s,1} & & & \ldots & z_{s,p}
\end{bmatrix}
\]

Cross-phenotype meta-analysis

\[
S_{\text{CPMA}} \sim \frac{\ln(\text{data} | \lambda=1)}{\ln(\text{data} | \lambda=1)}
\]

Cotsapas et al, PLoS Genetics
CAN WE LEARN REGULATORY VARIATION?

69-80% of cis associations are cell type-specific (0.001 permutation threshold)

- cis association sharing increases slightly when significance thresholds are relaxed
- Cell type specificity verified experimentally for subset of eQTLs

Dimas et al. Science 2009
Slide courtesy Antigone Dimas

Number of genes with cis-eQTL associations
8 extended HapMap populations
SRC: permutation threshold

<table>
<thead>
<tr>
<th></th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU</td>
<td>2869</td>
<td>0.06</td>
</tr>
<tr>
<td>CHB</td>
<td>2832</td>
<td>0.06</td>
</tr>
<tr>
<td>GIH</td>
<td>2959</td>
<td>0.06</td>
</tr>
<tr>
<td>JPT</td>
<td>2900</td>
<td>0.06</td>
</tr>
<tr>
<td>LWK</td>
<td>3818</td>
<td>0.05</td>
</tr>
<tr>
<td>MEX</td>
<td>2832</td>
<td>0.06</td>
</tr>
<tr>
<td>MKK</td>
<td>4222</td>
<td>0.04</td>
</tr>
<tr>
<td>YRI</td>
<td>3961</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>CEU</th>
<th></th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td># genes</td>
<td>2869</td>
<td>657</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDR</td>
<td>0.06</td>
<td>0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CHB</td>
<td>774</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GIH</td>
<td>698</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JPT</td>
<td>795</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LWK</td>
<td>773</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MEX</td>
<td>472</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MKK</td>
<td>947</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>YRI</td>
<td>799</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Non-redundant: 12494 3130
≥2 pops: 6889 0.55 1074 0.34
8 pops: 151 0.01 63 0.02

Stranger et al., in review

Shared association in 8 HapMap populations

Direction of allelic effect
same SNP-gene combination across populations

Population 1
AGREEMENT

Population 2
OPPOSITE

APOH: apolipoprotein H
Stranger et al., in review
**Direction of associations concordant across populations**

Shared SNP-probe associations

Spearman rho: 98.8%-100% direction concordance

0.001 permutation threshold

Stranger et al., in review

---

**POPULATION DIFFERENCES**

Population differences could have non-genetic basis

- Differences due to environment? (Idaghdour et al. 2008)
- Differences in cell line preparation? (Stranger et al. 2007)
- Differences due to batch effects? (Akey et al. 2007)

(Reviewed in Gilad et al. 2008)

---

**Gene expression experiment**

Does gene expression in 60 CEU + 60 YRI vary with ancestry?

Does gene expression in 89 AA vary with % Eur ancestry?

60 CEU + 60 YRI from HapMap, 89 AA from Coriell HD100AA

Gene expression measurements at 4,197 genes obtained using Affymetrix Focus array
Gene expression differences in African Americans validate CEU-YRI differences

\[ c = 0.43 \pm 0.02 \]  
\( (P\text{-value} < 10^{-25}) \)

RNA/seq questions

- Standard eQTLs
  - Montgomery et al, Pickrell et al Nature 2010
- Isoform eQTLs
  - Depth of sequence!
    - Long genes are preferentially sequenced
    - Abundant genes/isoforms ditto
  - Power!?  
    - Mapping biases due to SNPs
- Bulk trans tests

RNA/seq combined with other techs

- Regulons: TF gene sets via CHiP/seq
  - Look for trans effects
- Open chromatin states (Dnase I; methylation)
  - Find active genes
  - Changes in epigenetic marks correlated to RNA
    - Genetic effects
- RNA/DNA comparisons
  - Simultaneous SNP detection/genotyping
  - RNA editing ???

APPLICATION TO GWAS

eQTLs as intermediate traits

Schadt et al Nat Genet 2005
Exploring eQTLs in the relevant cell type is important for disease association studies.

Importance of cataloguing regulatory variation in multiple cell types.


ENCODE (and beyond).