Using Transcript Levels and Cell Culture Phenotypes to Extend Association Studies

Applications to Clinical Trials

Nancy J. Cox, Ph.D.
The University of Chicago
http://genemed.bsd.uchicago.edu
GWAS Works!

Every week
All top journals
Similar reports

GWAS success!

GWAS: Michigan
Sequencing Works!

Every week
All top journals
Similar reports

Sequencing success!

Sequencing Michigan
We’ve Been Picking the Cherries
The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. Brendan Maher shines a light on six places where the missing loot could be stashed away.

Even though genome-wide association studies (GWAS) turned up dozens of variants, they did “very little of the prediction that you would do just by asking people how tall their parents are”, says Jed Hirschhorn at the Broad Institute in Cambridge, Massachusetts who led one of the studies.

Height isn’t the only trait in which genes have gone missing, nor is it the most important. Studies looking at similarities between identical and fraternal twins estimate heritability at more than 90% for autism and more than 80% for schizophrenia. And genetics makes a major contribution to disorders such as obesity, diabetes and heart disease. GWAS, one of the most celebrated techniques of the past five years, promised to deliver many of the genes involved (see “Where’s the reward?” page 20). And to some extent they have identified more than 400 genetic variants that contribute to a variety of traits and common diseases. But even when dozens of genes have been linked to a trait, both the individual and cumulative effects are disappointingly small and nowhere near enough to explain earlier estimates of heritability. “It is the big topic in the genetics of common disease right now,” says Frances Collins, former head of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. The unexpected results left researchers at a point “where we all had to scratch our heads and say, ‘Huh!’”, he says.

Although fummoned by this missing heritability, geneticists remain optimistic that they can find more of it. “These are very early days, and there are things that are doable in the next year or two that may well explain another sizeable chunk of heritability,” says Hirschhorn. So where might it be hiding?
The case of the missing heritability

When scientists set out to map the human genome, they expected to find the genetic components of common traits and diseases. But when they were nowhere to be seen, Brendan Maher and colleagues began to look for other explanations.
The Real Problem?

Missing Biology
How Are We Doing with Identifying Primary Biology?

- Complement system in macular degeneration
- Autophagy in Crohn’s disease
- ?
- ?
How Can We Improve Our Ability to Identify Genes and Use Genetics to Understand Biology?
Transcriptome Studies

- No 3’ bias with this gene expression array
- Provide up to 4 probes per probeset (sub-exon)
- Core, extended, full probesets for flexible analyses

Affymetrix GeneChip® Human Exon 1.0ST Array in HapMap
Results of eQTL Discovery Studies Served in SCAN

- SNP and Copy number ANnotation database
- [http://www.scandb.org](http://www.scandb.org)
- Query by SNP (rs number), gene, or physical region
- eQTL information, multi-locus LD information, physical and functional annotations
Are Trait-Associated SNPs More Likely to Be eQTLs?

SCAN database
http://www.scandb.org

NHGRI GWAS Catalog
http://www.genome.gov/gwastudies/
SNPs with the highest eQTL function scores are enriched for WTCCC Crohn’s susceptibility loci.
WTCCC Crohn’s GWAS Has eQTL Enrichment, But …

• NO Enrichment of non-synonymous SNPs
• NO enrichment of SNPs in introns
• NO enrichment of SNPs in 3’ or 5’ UTR
SNPs with the strongest evidence for association in WTCCC Crohn’s disease are more likely to have eQTL function scores > 3
Results Are Robust Across Wide Range of Phenotypes

<table>
<thead>
<tr>
<th>Source</th>
<th># SNPs</th>
<th>$10^{-4}$</th>
<th>$10^{-6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using eQTLs from CEU+YRI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platform SNPs (MAF &gt; .05)</td>
<td>1,213,906</td>
<td>595285 (.490)</td>
<td>29347 (.024)</td>
</tr>
<tr>
<td>NHGRI</td>
<td>1598</td>
<td>972 (0.608)</td>
<td>83 (0.052)</td>
</tr>
<tr>
<td>Autoimmune Disorders</td>
<td>259</td>
<td>165 (0.637)</td>
<td>21 (0.081)</td>
</tr>
<tr>
<td>Cancers</td>
<td>93</td>
<td>56 (0.602)</td>
<td>4 (0.043)</td>
</tr>
<tr>
<td>Neurological/Psychiatric Disorders</td>
<td>63</td>
<td>41 (0.651)</td>
<td>2 (0.032)</td>
</tr>
</tbody>
</table>
Why Is This Important?

- Annotating SNPs with eQTL information (even from LCLs) will improve our ability to identify SNPs truly associated with disease.
- Utilizing this information should enhance our understanding of the underlying biology of the disease/trait.
- Magnitude of enrichment suggests that (for at least some disorders) we are nowhere near complete identification of associated common variants.
How Can We Make Use of the Information?

- Bayesian (and related frequentist approaches) that allow you to put higher priors on functional SNPs
- Suggests new ways of looking for gene-gene interactions
- Provides novel approaches to building pathways for disease
- New types of GWAS
Enhancements to SCAN

- Results of eQTL studies on brain
- Results of eQTL studies on liver
- Results of eQTL studies on adipose tissue
- Results of eQTL studies on skeletal muscle
- New software for function-based GWAS
Results Databases

- SCAN (SNP and Copy number Annotation) [http://www.scandb.org](http://www.scandb.org)
- PACdb (Pharmacogenetics And Cell lines) [http://www.pacdb.org](http://www.pacdb.org)
- PGScore (Pharmacogenomics Score) [http://www.pgscore.org](http://www.pgscore.org)
# Phenotype / Genotype Association

## Select Study:
- **80303: A study on pancreatic cancer**

## Select phenotype:
- **Hypertension grades 2–4**

## Select drug:
- **gemcitabine + bevacizumab**

## Enter SNPs:

### Or

**Upload SNP list:**
- **Choose File** no file selected

**Enter p-value threshold:**
- **.00001**

---

<table>
<thead>
<tr>
<th>Chrom</th>
<th>SNP</th>
<th>Position</th>
<th>P_value</th>
<th>OR</th>
<th>Minor Allele</th>
<th>MAF</th>
<th>HWE</th>
<th>CEU Minor Allele</th>
<th>CEU MAF</th>
<th>YRI Minor Allele</th>
<th>YRI MAF</th>
<th>ASN Minor Allele</th>
<th>ASN MAF</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs231432</td>
<td>56425941</td>
<td>3.9e-05</td>
<td>0.3056</td>
<td>G</td>
<td>0.217172</td>
<td>0.009597</td>
<td>G</td>
<td>0.279661</td>
<td>G</td>
<td>0.156009</td>
<td>G</td>
<td>0.136364</td>
<td>neuro34</td>
</tr>
<tr>
<td>1</td>
<td>rs4843142</td>
<td>97315400</td>
<td>7.1e-05</td>
<td>0.2483</td>
<td>A</td>
<td>0.180471</td>
<td>0.35812</td>
<td>A</td>
<td>0.190588</td>
<td>A</td>
<td>0.048359</td>
<td>1</td>
<td>0.35933</td>
<td>neuro34</td>
</tr>
<tr>
<td>8</td>
<td>rs735912</td>
<td>56352572</td>
<td>7.1e-05</td>
<td>0.2453</td>
<td>G</td>
<td>0.210436</td>
<td>0.4884</td>
<td>C</td>
<td>0.4286</td>
<td>C</td>
<td>0.400359</td>
<td>0.0284</td>
<td>A</td>
<td>0.117978</td>
</tr>
<tr>
<td>18</td>
<td>rs1505601</td>
<td>66642466</td>
<td>3.4e-05</td>
<td>0.3016</td>
<td>A</td>
<td>0.189893</td>
<td>0.5599</td>
<td>A</td>
<td>0.216467</td>
<td>0.4147</td>
<td>0.401786</td>
<td>0.7807</td>
<td>0.078517</td>
<td>0.4252</td>
</tr>
<tr>
<td>7</td>
<td>rs4843142</td>
<td>97315400</td>
<td>7.1e-05</td>
<td>0.2483</td>
<td>A</td>
<td>0.180471</td>
<td>0.35812</td>
<td>A</td>
<td>0.190588</td>
<td>A</td>
<td>0.048359</td>
<td>1</td>
<td>0.35933</td>
<td>neuro34</td>
</tr>
<tr>
<td>11</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>4</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>5</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>6</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>8</td>
<td>rs735912</td>
<td>56352572</td>
<td>7.1e-05</td>
<td>0.2453</td>
<td>G</td>
<td>0.210436</td>
<td>0.4884</td>
<td>C</td>
<td>0.4286</td>
<td>C</td>
<td>0.400359</td>
<td>0.0284</td>
<td>A</td>
<td>0.117978</td>
</tr>
<tr>
<td>18</td>
<td>rs1505601</td>
<td>66642466</td>
<td>3.4e-05</td>
<td>0.3016</td>
<td>A</td>
<td>0.189893</td>
<td>0.5599</td>
<td>A</td>
<td>0.216467</td>
<td>0.4147</td>
<td>0.401786</td>
<td>0.7807</td>
<td>0.078517</td>
<td>0.4252</td>
</tr>
<tr>
<td>7</td>
<td>rs4843142</td>
<td>97315400</td>
<td>7.1e-05</td>
<td>0.2483</td>
<td>A</td>
<td>0.180471</td>
<td>0.35812</td>
<td>A</td>
<td>0.190588</td>
<td>A</td>
<td>0.048359</td>
<td>1</td>
<td>0.35933</td>
<td>neuro34</td>
</tr>
<tr>
<td>11</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>4</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>5</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
<tr>
<td>6</td>
<td>rs1092841</td>
<td>13139650</td>
<td>9.2e-05</td>
<td>0.2231</td>
<td>C</td>
<td>0.308081</td>
<td>0.07601</td>
<td>A</td>
<td>0.200288</td>
<td>0.7275</td>
<td>0.342105</td>
<td>0.5594</td>
<td>0.486304</td>
<td>0.8216</td>
</tr>
</tbody>
</table>
Example Application

Top hits, neuropenia (3-4), gemcitabine, from PGScore

<table>
<thead>
<tr>
<th>Chrom</th>
<th>SNP</th>
<th>Position</th>
<th>P-value</th>
<th>OR</th>
<th>Minor Allele</th>
<th>MAF</th>
<th>HWE</th>
<th>CEU Minor Allele</th>
<th>CEU MAF</th>
<th>CEU HWE</th>
<th>YRI Minor Allele</th>
<th>YRI MAF</th>
<th>YRI HWE</th>
<th>ASN Minor Allele</th>
<th>ASN MAF</th>
<th>ASN HWE</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>rs2216322</td>
<td>56228414</td>
<td>3.88e-06</td>
<td>0.3058</td>
<td>G</td>
<td>0.217172</td>
<td>0.009597</td>
<td>G</td>
<td>0.279661</td>
<td>0.7539</td>
<td>G</td>
<td>0.190909</td>
<td>0.1803</td>
<td>G</td>
<td>0.136364</td>
<td>1</td>
<td>neutro34</td>
</tr>
<tr>
<td>1</td>
<td>rs4847428</td>
<td>93715030</td>
<td>7.71e-06</td>
<td>2.238</td>
<td>A</td>
<td>0.380471</td>
<td>0.5382</td>
<td>A</td>
<td>0.330508</td>
<td>0.2389</td>
<td>G</td>
<td>0.0438596</td>
<td>1</td>
<td>A</td>
<td>0.353933</td>
<td>1</td>
<td>neutro34</td>
</tr>
<tr>
<td>8</td>
<td>rs6473987</td>
<td>56525712</td>
<td>1.18e-05</td>
<td>2.453</td>
<td>A</td>
<td>0.210438</td>
<td>0.4884</td>
<td>T</td>
<td>0.2</td>
<td>0.4286</td>
<td>C</td>
<td>0.403509</td>
<td>0.02833</td>
<td>T</td>
<td>0.117978</td>
<td>1</td>
<td>neutro34</td>
</tr>
</tbody>
</table>

Identify overlap with cytotoxicity phenotypes in PACdb

<table>
<thead>
<tr>
<th>Population</th>
<th>Drug</th>
<th>Phenotype</th>
<th>SNP</th>
<th>Allele</th>
<th>Chrom</th>
<th>Position</th>
<th>Annotation (Host Gene, SNP Function)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU</td>
<td>ARAC</td>
<td>IC50</td>
<td>rs10088270</td>
<td>A/G</td>
<td>8</td>
<td>56550383</td>
<td><strong>XKR4</strong>, intron[NM_052898.1]</td>
<td>0.0074</td>
</tr>
<tr>
<td>YRI</td>
<td>ARAC</td>
<td>AUC</td>
<td>rs10088270</td>
<td>A/G</td>
<td>8</td>
<td>56550383</td>
<td><strong>XKR4</strong>, intron[NM_052898.1]</td>
<td>0.0093</td>
</tr>
</tbody>
</table>
IL17F, ch 1, 10^{-8}, q 0.04, HR 3.27

coding nonsynonymous

rs763780 (Caucasian Subset)
P-value (log-rank test)= 2.66e-08

Call: survfit(formula = SURV[CAUC] ~ snpi[CAUC])

<table>
<thead>
<tr>
<th></th>
<th>n events</th>
<th>median</th>
<th>0.95LCL</th>
<th>0.95UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>snpi[CAUC]=C/A</td>
<td>22</td>
<td>22</td>
<td>3.24</td>
<td>2.43</td>
</tr>
<tr>
<td>snpi[CAUC]=C/C</td>
<td>272</td>
<td>269</td>
<td>6.64</td>
<td>5.82</td>
</tr>
</tbody>
</table>
The Team

Nancy Cox
Dan Nicolae
M. Eileen Dolan
Wei Zhang and Shiwei Duan
Eric Gamazon

Supported by ENDGAMe and PGRN (PAAR) resources