Understanding the contribution of copy number polymorphisms to multigenic traits

Rob Scharpf

October 14, 2013
Serum uric acid concentrations

- serum uric acid concentrations are highly heritable
- high levels of serum uric acid (hyperuricemia) is associated with multiple diseases (gout, cardiovascular disease, and renal complications)
- SNPs in $SLC2A9$ and $ABCG2$ increase the risk for hyperuricemia and gout
Function of *SLC2A9* and *ABCG2* in the kidney

Woodward *et al.*, PNAS (2009)
Atherosclerosis Risk in Communities Study

- 8,598 European ancestry (EA), 3,392 African Americans (AA) passing QC
- IQR age: 49-59 (median 54)
- 47% of EA and 38% of AA participants are male
Copy number estimation

1. Preprocess Affy 6.0 CEL files (R package `crlmm`)

2. Fit 6-state HMM to each sample (R package `VanillaICE`) → genomic intervals with estimated copy number
CNVs (European ancestry)
Copy number estimation

3. Translate list of CNVs to a rectangular matrix (rows are genomic intervals, columns are samples)

4. Define copy number polymorphisms (CNPs) as intervals for which at least 1% of ARIC participants have a CNV
   - 14,678 such intervals in ARIC corresponding to approximately 434 regions
Many regions are complex
Association of copy number and uric acid

5. Regress uric acid on copy number estimates
   • model copy number as continuous
Copy number GWAS
Copy number GWAS

A

B

observed $-\log_{10} (p\text{-value})$

expected $-\log_{10} (p\text{-value})$
Signal is from two non-overlapping CNPs

- CNP regions are small and data is noisy, but CNPs are near *SLC2A9*
- *SLC2A9* is transcribed in the reverse direction
CNP-9Mb breakpoints

chr 4

Sample index

Mb

0
1000
2000
3000
4000
5000
6000

CNP-9Mb average log R ratios

Marginal distribution across samples – peaks are well separated
CNP-10Mb vs CNP-9Mb log R ratios
Marginal distributions are a mixture of normals
Marginal distributions are a mixture of normals

- we implemented a Gibbs’ sampler to approximate the posterior distribution and classify copy number
- different implementations have been developed by others (Korn et al., Nature Genetics (2008))
SNP in CNP-10Mb locus

![SNP in CNP-10Mb locus](image-url)
Association from HMM estimates
Association from mixture model
Gender-specific slopes

- Copy number estimates are negatively correlated
- Relationship between log (uric acid) and copy number is approximately linear
Replication in Framingham Heart Study

- Older array technology in Framingham (250k chips)
- No markers in the CNP-9Mb locus
- One SNP in the CNP-10Mb locus
- Only genotype calls available from our Framingham collaborators

HapMap data for 250k chip
Replication in Framingham Heart Study

<table>
<thead>
<tr>
<th>Population</th>
<th>CNP−9Mb</th>
<th>CNP−10Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC white female</td>
<td>4,566</td>
<td></td>
</tr>
<tr>
<td>FHS white female</td>
<td>2,106</td>
<td></td>
</tr>
<tr>
<td>ARIC black female</td>
<td>2,106</td>
<td></td>
</tr>
<tr>
<td>combined female</td>
<td>6,672</td>
<td></td>
</tr>
<tr>
<td>ARIC white male</td>
<td>4,032</td>
<td></td>
</tr>
<tr>
<td>FHS white male</td>
<td>1,286</td>
<td></td>
</tr>
<tr>
<td>ARIC black male</td>
<td>1,286</td>
<td></td>
</tr>
<tr>
<td>combined male</td>
<td>5,318</td>
<td></td>
</tr>
</tbody>
</table>

$\beta \pm 2 \times se$

* Missing genotype calls used as a surrogate for deletion genotypes
SNPs in *SLC2A9* increase risk of hyperuricemia
CNP association is independent of SNPs in SLC2A9
Remarks regarding independence

• Reverse is also true: adjusted for CNP estimates, SNP rs7675964 remains genome-wide significant.

• Some SNPs remain genome-wide significant in the rs7675964-adjusted model (i.e., rs6449213).
Phasing SNP allelic haplotypes with copy number

- We phased copy number estimates at CNP-9Mb and CNP-10Mb with the rs7675964 and rs6449213 genotypes

- Example haplotype

  Haplotype 1: $\neg\neg a\neg\neg b\neg\neg 0\neg\neg 1\neg\neg$
  Haplotype 2: $\neg\neg b\neg\neg b\neg\neg 1\neg\neg 1\neg\neg$

  (Haplotype for subject that is heterozygous at SNP rs7675964, homozygous for the minor allele at SNP rs6449213, hemizygous deletion at CNP-9Mb, and diploid at CNP-10Mb)

- Thanks to Dan Arking for the suggestion of phasing
Phasing *SLC2A9* SNPs and CNPs

- Only three allelic haplotypes exhibit variation in the corresponding copy number haplotypes:
  1. H1: $a-a-a$  
     H2: $a-a-a$
  2. H1: $b-a-a$  
     H2: $a-a-a$
  3. H1: $b-b-a$  
     H2: $a-a-a$

- Among subjects with the same allelic haplotype, do we see variation in uric acid levels associated with copy number haplotypes?
Association of CNP haplotypes

\[ \beta \pm 2 \times \text{se} \]

\begin{align*}
H1 &-a-a-a- & H1 &-b-a-a- & H1 &-b-b-b- \\
H2 &-a-a-a- & H2 &-a-a-a- & H2 &-a-a-a- \\
\end{align*}

<table>
<thead>
<tr>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1-1--1--</td>
<td>141</td>
</tr>
<tr>
<td>H2-0--1--</td>
<td>93</td>
</tr>
<tr>
<td>H1-1--0--</td>
<td>244</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
<tr>
<td>H1-0--1--</td>
<td>1029</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
<tr>
<td>H1-0--0--</td>
<td>883</td>
</tr>
<tr>
<td>H2-0--1--</td>
<td></td>
</tr>
<tr>
<td>H1-1--0--</td>
<td>1029</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
</tbody>
</table>

\begin{align*}
H1 &-a-a-a- & H2 &-a-a-a- & H1 &-b-b-b- \\
\end{align*}

<table>
<thead>
<tr>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1-1--1--</td>
<td>141</td>
</tr>
<tr>
<td>H2-0--1--</td>
<td>93</td>
</tr>
<tr>
<td>H1-1--0--</td>
<td>244</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
<tr>
<td>H1-0--1--</td>
<td>1029</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
<tr>
<td>H1-0--0--</td>
<td>883</td>
</tr>
<tr>
<td>H2-0--1--</td>
<td></td>
</tr>
<tr>
<td>H1-1--0--</td>
<td>1029</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
</tbody>
</table>

\begin{align*}
H1 &-a-a-a- & H2 &-a-a-a- & H1 &-b-b-b- \\
\end{align*}

<table>
<thead>
<tr>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1-1--1--</td>
<td>141</td>
</tr>
<tr>
<td>H2-0--1--</td>
<td>93</td>
</tr>
<tr>
<td>H1-1--0--</td>
<td>244</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
<tr>
<td>H1-0--1--</td>
<td>1029</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
<tr>
<td>H1-0--0--</td>
<td>883</td>
</tr>
<tr>
<td>H2-0--1--</td>
<td></td>
</tr>
<tr>
<td>H1-1--0--</td>
<td>1029</td>
</tr>
<tr>
<td>H2-1--0--</td>
<td></td>
</tr>
</tbody>
</table>
Discussion and Summary

- Fewer copies of CNP-9Mb
  ⇒ reduced \textit{SLC2A9} expression
  ⇒ less reabsorption of uric acid from urine
  ⇒ decreased levels of serum uric acid concentrations
Discussion and Summary

- Deletions at the CNP-10Mb span a locus identified by ENCODE with strong evidence of DNase hypersensitivity and histone makers in several normal tissues, but nothing yet reported in kidney.

- To date, no promoters have been reported at the CNP-9Mb locus.

- Gene expression data on the same set of individuals in target kidney and liver tissues is needed to evaluate whether deletions effect transcription of $SLC2A9$ as hypothesized, and to evaluate gender differences in $SLC2A9$ expression.

- Relationship of CNP-9Mb and CNP-10Mb with gout has not been evaluated.
Acknowledgements

• Linda Kao
• Lynn Mireles
• Adrienne Tin
• Stephen Cristiano
• Ingo Ruczinski
• Anna Köttgen
• Josef Coresh
• Aravinda Chakravarti
• Eric Boerwinkle
• Caroline Fox and Qiong Yang (Framingham Heart Study)
• Members of ARIC writing group