Variation in the human genome & disease

Ingo Ruczinski

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

April 7, 2010
Linkage versus association

From Goncalo Abecasis

Variation in the human genome & disease
Single nucleotide polymorphisms

1  

SNP  

2

urgi.versailles.inra.fr

Ingo Ruczinski  Variation in the human genome & disease
Haplotypes

Haplotypes

Haplotype blocks

Kim and Dionne (2007)
Case-control design

Most recent common ancestor

Ancestral mutation

Time

Case
Control

Indirect association

Direct association

Disease phenotype

Typed marker locus

Unobserved causal locus

Haplotype

Table 1
Estimated coverage of commercially available fixed marker genotyping platforms

<table>
<thead>
<tr>
<th>Platform</th>
<th>HapMap population sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YRI</td>
</tr>
<tr>
<td>Affymetrix GeneChip 500K</td>
<td>46</td>
</tr>
<tr>
<td>Affymetrix SNP Array 6.0</td>
<td>66</td>
</tr>
<tr>
<td>Illumina HumanHap300</td>
<td>33</td>
</tr>
<tr>
<td>Illumina HumanHap550</td>
<td>55</td>
</tr>
<tr>
<td>Illumina HumanHap650Y</td>
<td>66</td>
</tr>
<tr>
<td>Perlegen 600K</td>
<td>47</td>
</tr>
</tbody>
</table>

Data represent percent of SNPs tagged at $r^2 \geq 0.8$. Values assume all SNPs on the platform are informative and pass quality control. YRI, Yoruba in Ibadan, Nigeria; CEU, subsample of Utah residents of Northern European ancestry selected from Centre d'Étude du Polymorphisme Humain samples; CHB, Han Chinese in Beijing, China; JPT, Japanese in Tokyo. From the International HapMap Consortium, 2007 (3).

Population stratification

Case

Population 1

Population 2

Control

Genes mirror geography within Europe

John Novembre¹,², Toby Johnson⁴,⁵,⁶, Katarzyna Bryc⁷, Zoltán Kutalik⁴,⁵,⁶, Adam R. Boyko⁷, Adam Auton⁷, Amit Indap⁷, Karen S. King⁷, Sven Bergmann⁴,⁵, Matthew R. Nelson⁷, Matthew Stephens⁷,⁹ & Carlos D. Bustamante⁷

Understanding the genetic structure of human populations is of fundamental interest to medical, forensic and anthropological sciences. Advances in high-throughput genotyping technology have markedly improved our understanding of global patterns of human genetic variation and suggest the potential to use large samples to uncover variation among closely spaced populations¹-⁴. Here we characterize genetic variation in a sample of 3,000 European individuals genotyped at over half a million variable DNA sites in the human genome. Despite low average levels of genetic differentiation among Europeans, we find a close correspondence between genetic and geographic distances; indeed, a geographical map of Europe arises naturally as an efficient two-dimensional summary of genetic variation in Europeans. The results emphasize that when mapping the genetic basis of a disease phenotype, spurious associations can arise if genetic structure is not properly accounted for. In addition, the results are relevant to the prospects of genetic ancestry testing; an individual’s DNA can be used to infer their geographic origin with surprising accuracy—often to within a few hundred kilometres.

The resulting figure bears a notable resemblance to a geographic map of Europe (Fig. 1a). Individuals from the same geographic region cluster together and major populations are distinguishable. Geographically adjacent populations typically abut each other, and recognizable geographical features of Europe such as the Iberian peninsula, the Italian peninsula, southeastern Europe, Cyprus and Turkey are apparent. The data reveal structure even among French-, German- and Italian-speaking groups within Switzerland (Fig. 1b), and between Ireland and the United Kingdom (Fig. 1a, IE and GB). Within some countries individuals are strongly differentiated along the principal component (PC) axes, suggesting that in some cases the resolution of the genetic data may exceed that of the available geographic information.

When we quantitatively compare the geographic position of countries with their PC-based genetic positions, we observe few prominent differences between the two (Supplementary Fig. 1), and those that exist can be explained either by small sample sizes (for example, Slovakia (SK)) or by the coarseness of our geographic data (a problem for large countries, for example, Russia (RU)); see

Principal components


Ingo Ruczinski Variation in the human genome & disease
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium

There is increasing evidence that genome-wide association (GWA) studies represent a powerful approach to the identification of genes involved in common human diseases. We describe a joint GWA study (using the Affymetrix GeneChip 500K Mapping Array Set) undertaken in the British population, which has examined ~2,000 individuals for each of 7 major diseases and a shared set of ~3,000 controls. Case-control comparisons identified 24 independent association signals at \( P < 5 \times 10^{-7} \): 1 in bipolar disorder, 1 in coronary artery disease, 9 in Crohn's disease, 3 in rheumatoid arthritis, 7 in type 1 diabetes and 3 in type 2 diabetes. On the basis of prior findings and replication studies thus far completed, almost all of these signals reflect genuine susceptibility effects. We observed association at many previously identified loci, and found compelling evidence that some loci confer risk for more than one of the diseases studied. Across all diseases, we identified a large number of further signals (including 58 loci with single-point P values between \( 10^{-5} \) and \( 5 \times 10^{-7} \)) likely to yield additional susceptibility loci. The importance of appropriately large samples was confirmed by the modest effect sizes observed at most loci identified. This study thus represents a thorough validation of the GWA approach. It has also demonstrated that careful use of a shared control group represents a safe and effective approach to GWA analyses of multiple disease phenotypes; has generated a genome-wide genotype database for future studies of common diseases in the British population; and shown that, provided individuals with non-European ancestry are excluded, the extent of population stratification in the British population is generally modest. Our findings offer new avenues for exploring the pathophysiology of these important disorders. We anticipate that our data, results and software, which will be widely available to other investigators, will provide a powerful resource for human genetics research.

Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout

Owen M. Woodward\textsuperscript{a,1}, Anna Kottgen\textsuperscript{b,1}, Josef Coresh\textsuperscript{b}, Eric Boerwinkle\textsuperscript{c}, William B. Guggino\textsuperscript{a}, and Michael Kottgen\textsuperscript{d,2}

\textsuperscript{a}Department of Physiology, Johns Hopkins Medical Institutions, Baltimore, MD 21205; \textsuperscript{b}Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205; \textsuperscript{c}Human Genetics Center and Division of Epidemiology, University of Texas, Houston, TX 77030; and \textsuperscript{d}Department of Biological Chemistry, Johns Hopkins Medical Institutions, Baltimore, MD 21205

Edited by Maurice B. Burg, National Institutes of Health, Bethesda, MD, and approved May 4, 2009 (received for review February 4, 2009)

Genome-wide association studies (GWAS) have successfully identified common single nucleotide polymorphisms (SNPs) associated with a wide variety of complex diseases, but do not address gene function or establish causality of disease-associated SNPs. We recently used GWAS to identify SNPs in a genomic region on chromosome 4 that associate with serum urate levels and gout, a consequence of elevated urate levels. Here we show using functional assays that human ATP-binding cassette, subfamily G, 2 (ABCG2), encoded by the ABCG2 gene contained in this region, is a hitherto unknown urate efflux transporter. We further show that native ABCG2 is located in the brush border membrane of kidney proximal tubule cells, where it mediates renal urate secretion. Introduction of the mutation Q141K encoded by the common SNP rs2231142 by site-directed mutagenesis resulted in 53% reduced urate transport rates compared to wild-type ABCG2 ($P < 0.001$). Data from a population-based study of 14,783 individuals support rs2231142 as the causal variant in the region and show highly significant associations with urate levels [whites $P = 10^{-30}$, minor allele frequency (MAF) 0.11; blacks $P = 10^{-4}$, MAF 0.03] and gout (adjusted odds ratio 1.68 per risk allele, both races). Our data indicate that at least 10% of all gout cases in whites are attributable to this causal variant. With approximately 3 million US individuals suffering from often insufficiently treated gout, ABCG2 represents an attractive drug target. Our study completes the chain of evidence from association to causation and supports the common disease-common variant hypothesis in the etiology of gout.

in oocytes and urate efflux rates from oocytes were measured. Urate accumulation was significantly decreased by 75.5% in oocytes expressing ABCG2 compared with water-injected control oocytes (Fig. 1A). In addition, ABCG2-expressing oocytes showed lower urate concentrations than oocytes expressing the known urate efflux transporter MRP4 (7). The reduced urate accumulation in ABCG2-expressing oocytes was absent in the presence of fumitremorgin C (FTC), a specific ABCG2 inhibitor, or after introduction of a mutation in ABCG2 (187T) that is known to disrupt transport of chemotherapeutic agents (5, 6) (Fig. 1B). ABCG2-expressing oocytes showed significantly lower urate accumulation over a wide range of extracellular concentrations compared with control cells or cells expressing the loss-of-function mutation ABCG2 187T (Fig. 1C). The reduced urate accumulation was due to ABCG2-mediated export of urate out of the cells as shown in experiments monitoring the decrease of the intracellular urate concentration over time in oocytes preloaded with radiolabeled urate (Fig. 1D). Furthermore, ABCG2-mediated urate efflux was dependent on the intracellular urate concentration and significantly higher in ABCG2-expressing oocytes than in control oocytes (Fig. 1E). Together, these data show that ABCG2 is a urate efflux transporter.

in Renal Epithelial Cells. In mammals, the proximal tubule is the major site of renal urate handling (1). To study the urate transport capacity of endogenous ABCG2 in polarized renal epithelial cells, we measured urate accumulation in native LLC-PK1 cells. Inhibition of ABCG2 by FTC resulted in a significant and statistically significant increase in urate accumulation compared to control cells (Fig. 1F). The importance of ABCG2 in urate homeostasis is further supported by the identification of a common variant in the ABCG2 gene associated with serum urate levels and the observation that ABCG2-mediated urate efflux is dependent on the intracellular urate concentration.
Science in medicine

A HapMap harvest of insights into the genetics of common disease

Teri A. Manolio, Lisa D. Brooks, and Francis S. Collins

National Human Genome Research Institute, Bethesda, Maryland, USA.

The International HapMap Project was designed to create a genome-wide database of patterns of human genetic variation, with the expectation that these patterns would be useful for genetic association studies of common diseases. This expectation has been amply fulfilled with just the initial output of genome-wide association studies, identifying nearly 100 loci for nearly 40 common diseases and traits. These associations provided new insights into pathophysiology, suggesting previously unsuspected etiologic pathways for common diseases that will be of use in identifying new therapeutic targets and developing targeted interventions based on genetically defined risk. In addition, HapMap-based discoveries have shed new light on the impact of evolutionary pressures on the human genome, suggesting multiple loci important for adapting to disease-causing pathogens and new environments. In this review we examine the origin, development, and current status of the HapMap; its prospects for continued evolution; and its current and potential future impact on biomedical science.

Catalogued findings for the following phenotypes:

1. Eye diseases
2. Diabetes
3. Cancer
4. Gastrointestinal disorders
5. Cardiovascular conditions and lipid metabolism
6. Neuropsychiatric conditions
7. Autoimmune and infectious diseases
8. Various traits

Results

http://www.genome.gov/GWAstudies/

Ingo Ruczinski

Variation in the human genome & disease
# Results

Phenotypes under investigation in collaborative GWA studies

<table>
<thead>
<tr>
<th>GAIN&lt;sup&gt;a&lt;/sup&gt;</th>
<th>GEI&lt;sup&gt;b&lt;/sup&gt;</th>
<th>STAMPEED&lt;sup&gt;c&lt;/sup&gt;</th>
<th>CGEMS&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention deficit hyperactivity disorder</td>
<td>Type 2 diabetes</td>
<td>Early-onset myocardial infarction</td>
<td>Prostate cancer</td>
</tr>
<tr>
<td>Major depressive disorder</td>
<td>Maternal metabolism and birth weight</td>
<td>Asthma</td>
<td>Breast cancer</td>
</tr>
<tr>
<td>Bipolar I disorder</td>
<td>Preterm birth</td>
<td>Platelet phenotypes</td>
<td>Pancreatic cancer</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>Oral clefts</td>
<td>CHD and other heart, lung and blood disorders</td>
<td>Lung cancer</td>
</tr>
<tr>
<td>Type I diabetic nephropathy</td>
<td>Dental caries</td>
<td>Childhood respiratory outcomes</td>
<td>Bladder cancer</td>
</tr>
<tr>
<td>Psoriasis</td>
<td>Coronary disease</td>
<td>Hematopoietic cell transplant outcome</td>
<td>Renal cancer</td>
</tr>
<tr>
<td></td>
<td>Lung cancer</td>
<td>Arteriosclerosis in hypertensives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Addiction</td>
<td>Asthma and lung function</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cardiovascular risk factors</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Atherosclerosis pathway genes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CV events</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Early coronary artery disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phenotypic variability in sickle cell anemia</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Centenarians</td>
<td></td>
</tr>
</tbody>
</table>


Missing heritability


Finding the missing heritability of complex diseases

Genome-wide association studies have identified hundreds of genetic variants associated with complex human diseases and traits, and have provided valuable insights into their genetic architecture. Most variants identified so far confer relatively small increments in risk, and explain only a small proportion of familial clustering, leading many to question how the remaining, ‘missing’ heritability can be explained. Here we examine potential sources of missing heritability and propose research strategies, including and extending beyond current genome-wide association approaches, to illuminate the genetics of complex diseases and enhance its potential to enable effective disease prevention or treatment.
### Missing heritability

#### Estimates of heritability and number of loci for several complex traits

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of loci</th>
<th>Proportion of heritability explained</th>
<th>Heritability measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age-related macular degeneration$^{72}$</td>
<td>5</td>
<td>50%</td>
<td>Sibling recurrence risk</td>
</tr>
<tr>
<td>Crohn's disease$^{21}$</td>
<td>32</td>
<td>20%</td>
<td>Genetic risk (liability)</td>
</tr>
<tr>
<td>Systemic lupus erythematosus$^{73}$</td>
<td>6</td>
<td>15%</td>
<td>Sibling recurrence risk</td>
</tr>
<tr>
<td>Type 2 diabetes$^{74}$</td>
<td>18</td>
<td>6%</td>
<td>Sibling recurrence risk</td>
</tr>
<tr>
<td>HDL cholesterol$^{75}$</td>
<td>7</td>
<td>5.2%</td>
<td>Residual phenotypic variance</td>
</tr>
<tr>
<td>Height$^{15}$</td>
<td>40</td>
<td>5%</td>
<td>Phenotypic variance</td>
</tr>
<tr>
<td>Early onset myocardial infarction$^{76}$</td>
<td>9</td>
<td>2.8%</td>
<td>Phenotypic variance</td>
</tr>
<tr>
<td>Fasting glucose$^{77}$</td>
<td>4</td>
<td>1.5%</td>
<td>Phenotypic variance</td>
</tr>
</tbody>
</table>

*Residual is after adjustment for age, gender, diabetes.

---

Missing heritability

1000 Genomes

A Deep Catalog of Human Genetic Variation

PROJECT OVERVIEW

The 1000 Genomes Project is an international research consortium formed to create the most detailed and medically useful picture to date of human genetic variation. The project involves sequencing the genomes of approximately 1200 people from around the world and receives major support from the Wellcome Trust Sanger Institute in Hinxton, England, the Beijing Genomics Institute Shenzhen in China and the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH).

Drawing on the expertise of multidisciplinary research teams, the 1000 Genomes Project will develop a new map of the human genome that will provide a view of biomedically relevant DNA variations at a resolution unmatched by current resources. As with other major human genome reference projects, data from the 1000 Genomes Project will be made swiftly available to the worldwide scientific community through freely accessible public databases.

On 4 September 2008, the co-chairs of the analysis group and overall project co-chairs drafted a letter to the NIH Council about 1000 Genomes Project. This letter, available here, reviews the goals, describes the current status, and provide an update on the critical tasks the Analysis Group must accomplish in order to deliver a valuable community resource and achieve the Project's goals.

Additional project details are are available in the September 2007 meeting report.

http://www.1000genomes.org
Trisomy

Variation in the human genome & disease

Ingo Ruczinski
Karyotypes

General Cytogenetics Information  http://members.aol.com/chrominfo/

Ingo Ruczinski  Variation in the human genome & disease
Courtesy of the Pevsner Laboratory
De novo deletion

Ingo Ruczinski

Variation in the human genome & disease
## Copy number variants


<table>
<thead>
<tr>
<th>Disease</th>
<th>Locus</th>
<th>Type of CNV</th>
<th>Size (kb)</th>
<th>Population frequency</th>
<th>Case frequency</th>
<th>Effect size (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare CNVs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism/IMR(^59)</td>
<td>16p11.2</td>
<td>De novo deletion</td>
<td>600</td>
<td>1 \times 10^{-4}</td>
<td>1%</td>
<td>100</td>
</tr>
<tr>
<td>Autism(^59)</td>
<td>16p11.2</td>
<td>Rare duplication</td>
<td>600</td>
<td>3 \times 10^{-4}</td>
<td>0.50%</td>
<td>16</td>
</tr>
<tr>
<td>Schizophrenia(^60,78)</td>
<td>1q21.1</td>
<td>Rare deletion</td>
<td>1,400</td>
<td>2 \times 10^{-4}</td>
<td>0.30%</td>
<td>15</td>
</tr>
<tr>
<td>Schizophrenia(^60,78)</td>
<td>1q21.1</td>
<td>Rare deletion</td>
<td>1,400</td>
<td>2 \times 10^{-4}</td>
<td>0.47%</td>
<td>Not observed in 4,737 controls</td>
</tr>
<tr>
<td>Epilepsy(^80)</td>
<td>15q13.3</td>
<td>Rare deletion</td>
<td>1,600</td>
<td>2 \times 10^{-4}</td>
<td>0.20%</td>
<td>12</td>
</tr>
<tr>
<td>IMR(^79,81)</td>
<td>15q13.3</td>
<td>Rare deletion</td>
<td>1,600</td>
<td>2 \times 10^{-4}</td>
<td>1.0%</td>
<td>Not observed in 3,899 controls</td>
</tr>
<tr>
<td>Schizophrenia(^82)</td>
<td>22q11.2</td>
<td>Rare deletion</td>
<td>3,000</td>
<td>2.5 \times 10^{-4}</td>
<td>0.30%</td>
<td>Not observed in 960 controls</td>
</tr>
<tr>
<td>Common CNPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crohn’s disease(^83)</td>
<td>IRGM</td>
<td>Deletion polymorphism</td>
<td>20</td>
<td>7%</td>
<td>11%</td>
<td>1.5</td>
</tr>
<tr>
<td>Body mass index(^61)</td>
<td>NEGR1</td>
<td>Deletion polymorphism</td>
<td>45</td>
<td>65%</td>
<td>Quantitative trait</td>
<td>&lt;1 kg</td>
</tr>
<tr>
<td>Psoriasis(^84)</td>
<td>LCE3C</td>
<td>Deletion polymorphism</td>
<td>30</td>
<td>55%</td>
<td>65%</td>
<td>1.3</td>
</tr>
</tbody>
</table>

IMR, Idiopathic mental retardation.
The DNA Age - Parents Whose Children Share Genetic Mutations Seek Each Other for Support - New York Times

December 28, 2007

THE DNA AGE

After DNA Diagnosis: ‘Hello, 16p11.2. Are You Just Like Me?’

By AMY HARMON

The girls had never met, but they looked like sisters.

There was no missing the similarities: the flat bridge of their noses, the thin lips, the fold near the corner of their eyes. And to the families of 14-year-old Samantha Napier and 4-year-old Taygen Lane there was something else, too. In the likeness was lurking an explanation for the learning difficulties, the digestion problems, the head-banging that had troubled each of them, for so long.

Several of the adults wiped tears from their eyes. “It’s like meeting family,” said Jessica Houk, Samantha’s older sister, who accompanied her and their mother to a Kentucky amusement park last July to greet Taygen.