Genome-wide Association Studies

David W. Craig, Ph.D.
Investigator
Neurogenomics Division
dcraig@tgen.org
Identification of the genetic basis of disease

Diagnostic (Personalized Medicine)

New targets for treatment

Better treatment
Family-based Linkage Studies

- Single Nucleotide Polymorphisms (SNPs)
- Family-based Linkage Studies

SNP-Genotyping Arrays

Affymetrix
(10K, 100K, 500K, 5.0, 1M)

Illumina
(100K, 320K, 230S, 550K, 650K, 430S)
Requirements
Genome-wide Association Studies (common variant hypothesis)

Genome-wide scan of 300K-1M single nucleotide polymorphisms (SNPs) captures the majority of the common variation.

Common Variant Hypothesis

Technology driven

SNP Genotyping Microarrays
- Affymetrix
- Illumina

Content increasing
100’s 1,000s 10,000s of samples

Rare variant hypothesis
Alzheimer’s Disease & Apoe-ε4

Fisher’s exact p-values across APOE locus

- APOE E4 (rs429358) uncorrected p-value = 1.76 x 10^-22
- rs4420638 uncorrected p-value = 3.27 x 10^-16

Bonferroni p-value correction (p = 10^-7)

Chr 19: rs429358

Common Variant Diseases
- Age-related Macular Degeneration
- Diabetes Type II
- Prostate Cancer (gene*gene)
- Breast Cancer
- Crohn’s Disease
- Multiple Sclerosis
- Cardiovascular disease
- Colon cancer
- etc.

Other large DNA microarray studies
- Autism
- Bipolar Disorder
- Common Variant Diseases
- Age-related Macular Degeneration
- Diabetes Type II
- Prostate Cancer (gene*gene)
- Breast Cancer
- Crohn’s Disease
- Multiple Sclerosis
- Cardiovascular disease
- Colon cancer
- etc.
Targeted Sequencing for Many Individuals

Type II Diabetes

Nature Reviews Genetics 8, 657-662 (September 2007)
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 identifying genes involved in common human diseases. We describe a joint GWA study (using the 500K Mapping Array Set) undertaken in the British population, which has examined 2,000 individual cases and a shared set of 3,000 controls. Case-control comparisons identified 24 independent P < 10^{-7} loci in bipolar disorder, 13 coronary artery disease, 9 in Crohn's disease, 3 in rheumatoid arthritis and 3 in type 2 diabetes. On the basis of prior findings and replication studies, such loci reflect genuine susceptibility effects. We observed association at many previously identified loci. This study is a further 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, and that a genome-wide gene database can be used for future studies of common diseases in the British population; and shows that, while Europeans are at risk of the disorders, the extent of population stratification in the British population is generally moderate. Our findings suggest avenues for exploring the pathophysiology of these important disorders. We anticipate that these data, results and software, which will be widely available to other investigators, will provide a powerful resource for human genetics research.

A nonsynonymous SNP in PRKCH (protein kinase C η) increases the risk of cerebral infarction

A genome-wide association study identifies novel risk loci for type 2 diabetes

Common Kibra Alleles Are Associated with Human Memory Performance

Anandar Papasotroupolou,

Dietrich A. Stephan,

Matthew J. Hausdorff,

Frederic J. Heesen,

David W. Craig,

John V. Pearson,

Kim-Dung Nguyen,

Fabienne Brunner,

Jason Corneil,

David Osbornes,

Malcolm Hull,

Amanda Aerni,

Daniel Cobza,

Jürgen Härnig,

Christian A. R. Monteleone,

Andreas Buchmann,

Eric M. Kelman,

Richard J. Corbell,

Katharina Hennke,

Dominique J.-F. de Quervain.

Human memory is a polygenic trait. We performed a genome-wide screen to identify memory-related gene variants. A genomic locus encoding the brain protein KIBRA was significantly associated with memory performance in three independent, genetically unrelated cohorts from Switzerland and the United States. Gene expression studies showed that KIBRA was expressed in memory-related brain structures. Functional magnetic resonance imaging imaging detected KIBRA allele high statistical confidence (10). Two SNPs fulfilled these selection criteria and were prioritized for subsequent individual genotyping to exclude pooling-related false positives: rs17077145 and rs439898. Both SNPs map within genes expressed in the human brain: rs17077145 is a common T → C substitution within the ninth exon of KIBRA (Klfox accession number NM_013238), encoding a neuronal protein, and rs439898 is a common T → C substitution within the first intron of GSF2 (encoding the synaptic protein calprenylin 2) (NM_022131).

Both the KIBRA and GSF2 SNPs were also significantly associated with different human memory performance when we analyzed the full sample (p < 0.05).
Identification of a gene associated with memory

KIBRA
Best SNP rank in window: #50

Chromosome 5 locus

Chromosome 5 Physical Position (Build 35.1)
Genotype training cohort with a few SNPs individually to validate pooling results
The Elephant in the Room

Common variants:
Late onset with big effect size in a common disorder
  - One: ApoE-ε4
  - Why?
Most small effect size.
  - Increase risk 20-30% for common diseases.
  - E?
Some, still, a larger effect size:
  - Generally, though, rare disorders.
    - Multiple Sclerosis, Progressive Supranuclear Palsy
  - E?
Responder / non-responder
  - Not fully explored.
Why not more E discoveries?
  - Annotation and Power?
Where is the major push for the next-generation GWA?

Bipolar as an example, 10K cases & 100K controls

2007
- Baum et al
- WTCCC

2008
- Sklar
- GAIN

Forthcoming
- Pritzker
- BiGS

Future?
- 10K cases & 100K controls
Personalized Medicine

Promise of personalized medication inherently linked to whole-genome sequencing
‘$1,000 Genome’ within reach 5 years, in 2008 a genome is $100K
Massive data production Tb per day in Images

Sequencing center in 2001,
20 Sequencers ~ 25 Mbases/yr

Termed ‘next-generation’
~500 Mbases/day/machine

Acknowledgements

TGen

Szabolcs Szelinger
Margot Josephson
Sotiris Mitropanopolous
Aswin Sekar

Dietrich Stephan, Ph.D.
Matthew J. Huentelman, Ph.D.
John V. Pearson

Funding

Stardust Foundation
R01 MH078151
U01 HL086528