




Summer Institute
in Statistical Genetics 2017

Genetics and Genomics

4. Association Studies



ggibson.gt@gmail.com
http://www.cig.gatech.edu

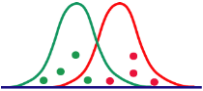



Outline

- General overview of association studies
- Sample Results
- Three steps to GWAS:
 - Primary scan
 - Replication
 - Fine mapping

Principle of Association Studies

Individual	Site	Score
1	A T C C G A	9
2	A C T C G A	8
3	A C C A - G	3
4	T T C A G A	5
5	A T C A G A	2
6	A C C C - G	7
7	T C T A - G	4
8	A T C C G A	8

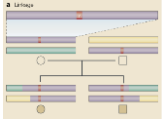


Are the phenotype scores associated with each class of SNP drawn from the same or different distributions ?

Linkage versus Association

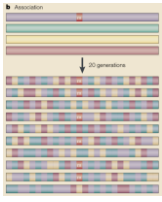
Linkage examines recent recombination events in a pedigree:

- over just several generations
- large chromosomal regions detected
- no information on allele frequency




Association examines historical recombination events in a population:

- basically a 10,000 generation pedigree
- resolution to single genes
- estimates effect size and frequency



Cardon & Bell, 2003. Nat. Rev. Genet. 2: 91

Why LD happens



When a mutation occurs, by definition it is only on one chromosome and hence "associated" with the genotypes elsewhere on that chromosome.

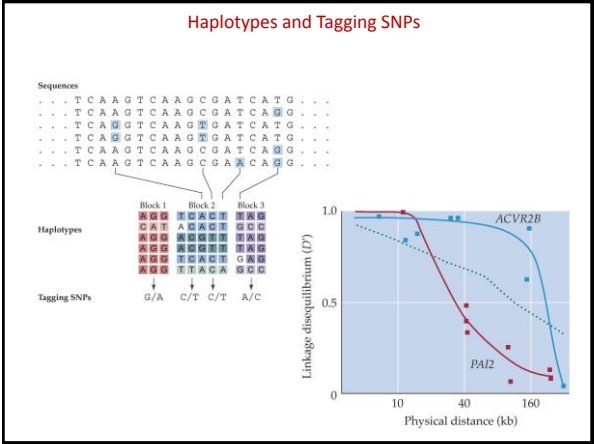
Over time, the mutation increases in frequency and becomes a polymorphism. It remains in LD with the genotypes on the chromosome it appeared on.

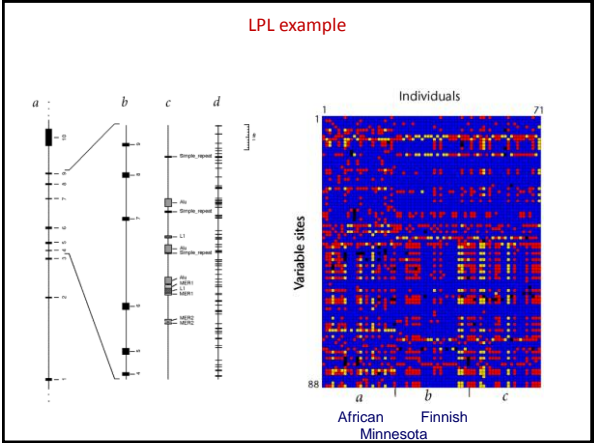
Eventually recombination breaks up the LD, in proportion to genetic distance.

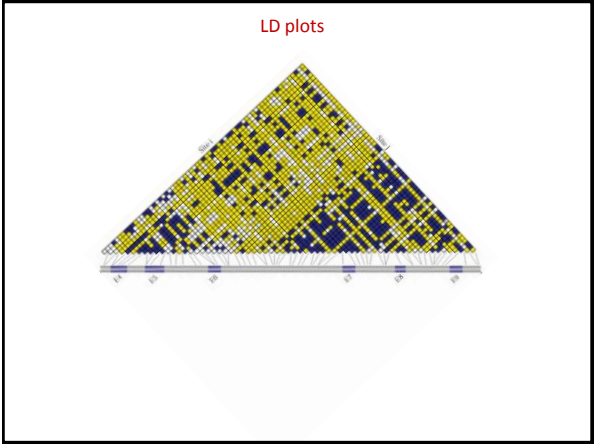
Measurement of LD

LD is the non-random association of genotypes.

		Expected					Observed		
		AA	AG	GG			AA	AG	GG
		24	48	24			24	48	24
TT	24	6	12	6	TT	24	24	0	0
TC	48	12	24	12	TC	48	0	48	0
CC	24	6	12	6	CC	24	0	0	24








Key Parameters for LD Mapping


- Polymorphism:
Flies 1/30 bp, 10x > Human 1/kb
- Haplotype structure/LD:
Fly LD decays over 200 bp, Human LD decays over 100 kb
- Population structure:
Panmixia and clinality v. Structure and admixture
- Allele frequencies:
Much more power for common alleles (infinitesimal model)

Case-Control and TDT designs

Affected $f(M) = 0.1$



Unaffected $f(M) = 0.2$



	Observed		Expected	
	Allele M	Allele m	Allele M	Allele m
Affected	34	278	61	265
Unaffected	69	256	62	269

$\chi^2 = 14.0$ $P < 0.001$

$\frac{M}{m} \times \frac{M}{m}$

↓

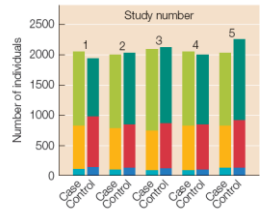
$f(M) = 0.63$

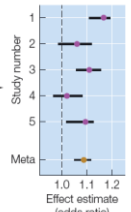
Transmitted Allele

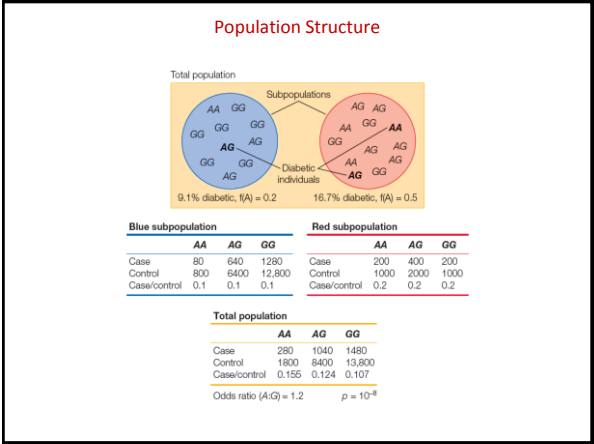
	M	m
Observed	78	46
Expected	62	62

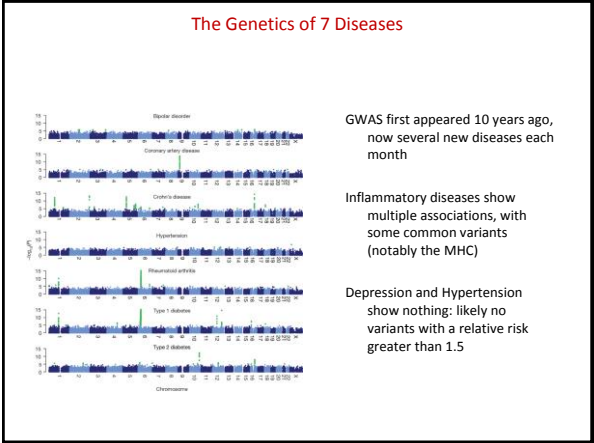
$\chi^2 = 8.2$ $P < 0.001$

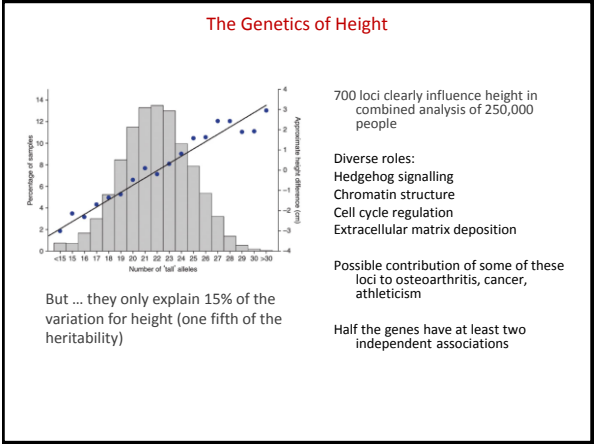
Repeatability and Forest Plots



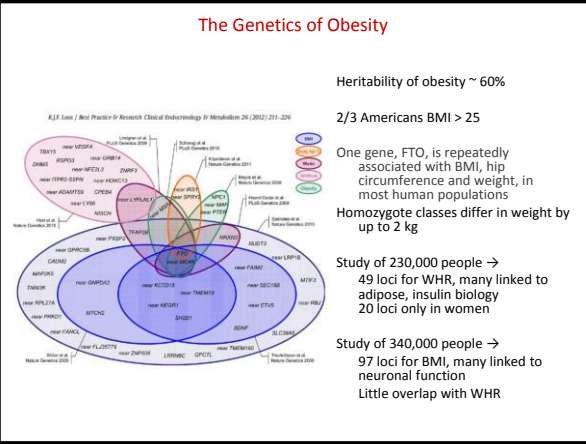








The Genetics of Obesity



Heritability of obesity ~ 60%

2/3 Americans BMI > 25

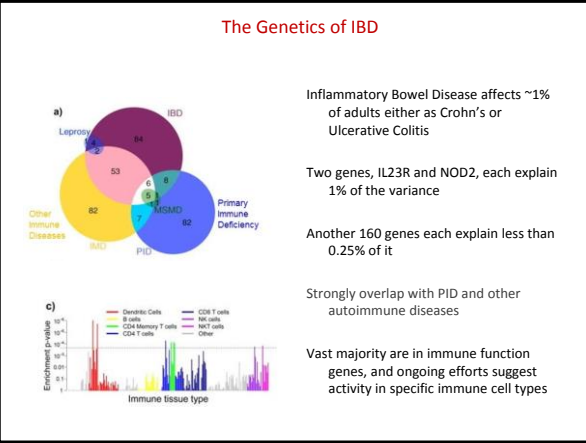
One gene, FTO, is repeatedly associated with BMI, hip circumference and weight, in most human populations

Homozygote classes differ in weight by up to 2 kg

Study of 230,000 people → 49 loci for WHR, many linked to adipose, insulin biology
20 loci only in women

Study of 340,000 people → 97 loci for BMI, many linked to neuronal function
Little overlap with WHR

The Genetics of IBD



Inflammatory Bowel Disease affects ~1% of adults either as Crohn's or Ulcerative Colitis

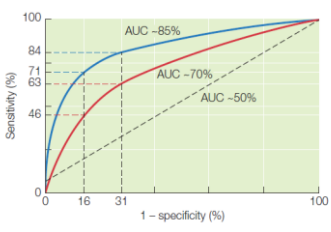
Two genes, IL23R and NOD2, each explain 1% of the variance

Another 160 genes each explain less than 0.25% of it

Strongly overlap with PID and other autoimmune diseases

Vast majority are in immune function genes, and ongoing efforts suggest activity in specific immune cell types

Prediction, Specificity, and Sensitivity



	Cases	Controls	
Called positive	True positive (TP), 25	False positive (FP), 215	PPV, (TP / (TP + FP)) = 10.4%
Called negative	False negative (FN), 10	True negative (TN), 1150	NPV, (TN / (FN + TN)) = 99.1%
Sensitivity	(TP / (TP + FN)) = 71%		Specificity
			(TN / (FP + TN)) = 84%

The Genetics of Type 2 Diabetes

54 established Fasting Glucose or Fasting Insulin loci tend to have pancreatic islet cell functions

23 replicate in African Americans in Trans-ethnic analyses

TCF7L2 is the strongest risk locus for T2D in Caucasians. The ancestral allele is the risk allele: it is found in 90% of Africans, 40% of Europeans, and just 5% of Asians

Considering the top 18 loci:
 1% of CAU have >24 alleles
 2% of CAU have <12 alleles
 They differ 4-fold in odds of T2D

The Genetics of Schizophrenia

128 independent SNP associations from GWAS of 37,000 cases

Strong enrichment in genes expressed in certain neuronal cell types or implicated in synaptic transmission

But at least 5% of cases attributable to CNV: copy number variation

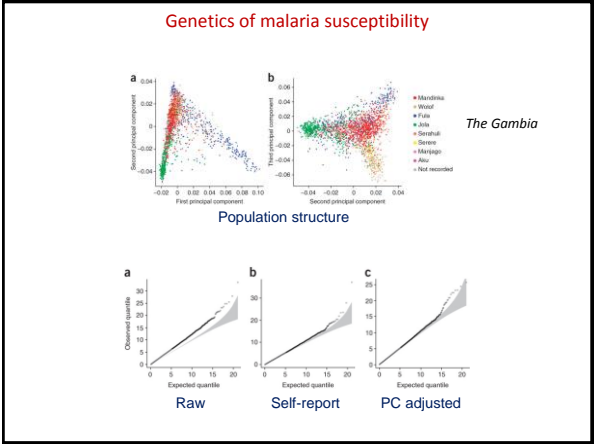
3 major chromosomal deletions of >100kb at frequency <1% are almost exclusively found in schizophrenics

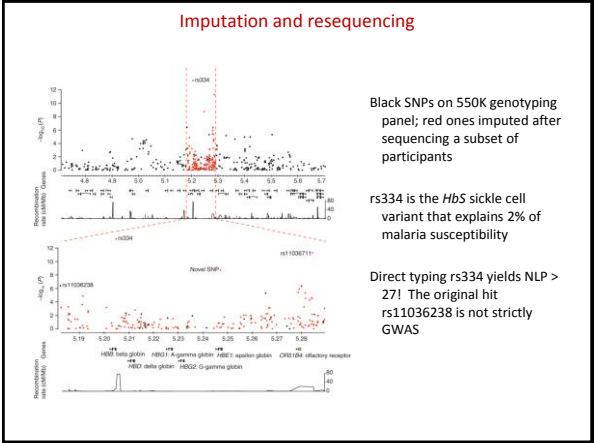
23andme studies

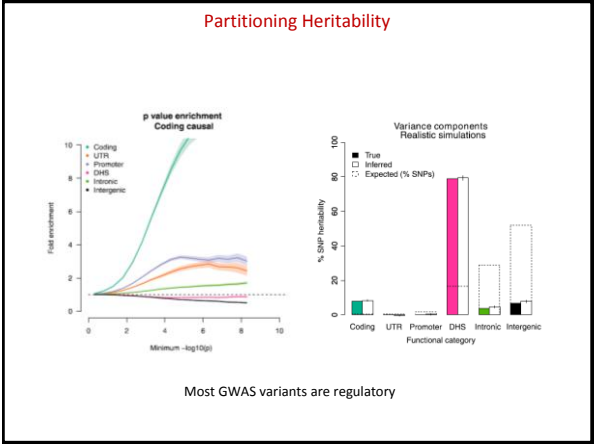
Other interesting traits:

- Endurance Runner vs Sprinter (30% of people change their answer if they know their ACTN3)
- Left vs Right Handedness (nothing striking)
- Have you ever needed braces or wisdom teeth surgery?
- Breast size (finds breast cancer risk loci)
- Hand-clasp dominance ...

Eriksson et al, (2010) PLoS Genetics 6: e1000993







Some references

- The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**: 661-678
- Ioannidis, J.P., G. Thomas and M.J. Daly (2009) Validating, augmenting and refining genome-wide association signals. *Nature Reviews Genetics* **10**: 318-329
- Shungin, D. et al. (2015) New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**: 187-196.
- Locke, A.E. et al (2015) Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**: 197-206.
- Wood, A.R., et al (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* **40**: 575-583
- Jostins, L. et al (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**: 119-124.
- DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics* **46**: 234-244.
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**: 421-427.
- Gusev, A. et al (2015) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* **95**: 535-552.
- Jallow, M. et al (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nature Genetics* **41**: 657-665
- Visscher, P.M, Brown, M.A., McCarthy, M.I., and Yang, J. (2012) Five years of GWAS discovery. *Am J Hum Genet.* 907-24.
