Genome-Wide Association analysis significance, power and coverage

Yurii Aulchenko
Erasmus MC Rotterdam
26.08.2009

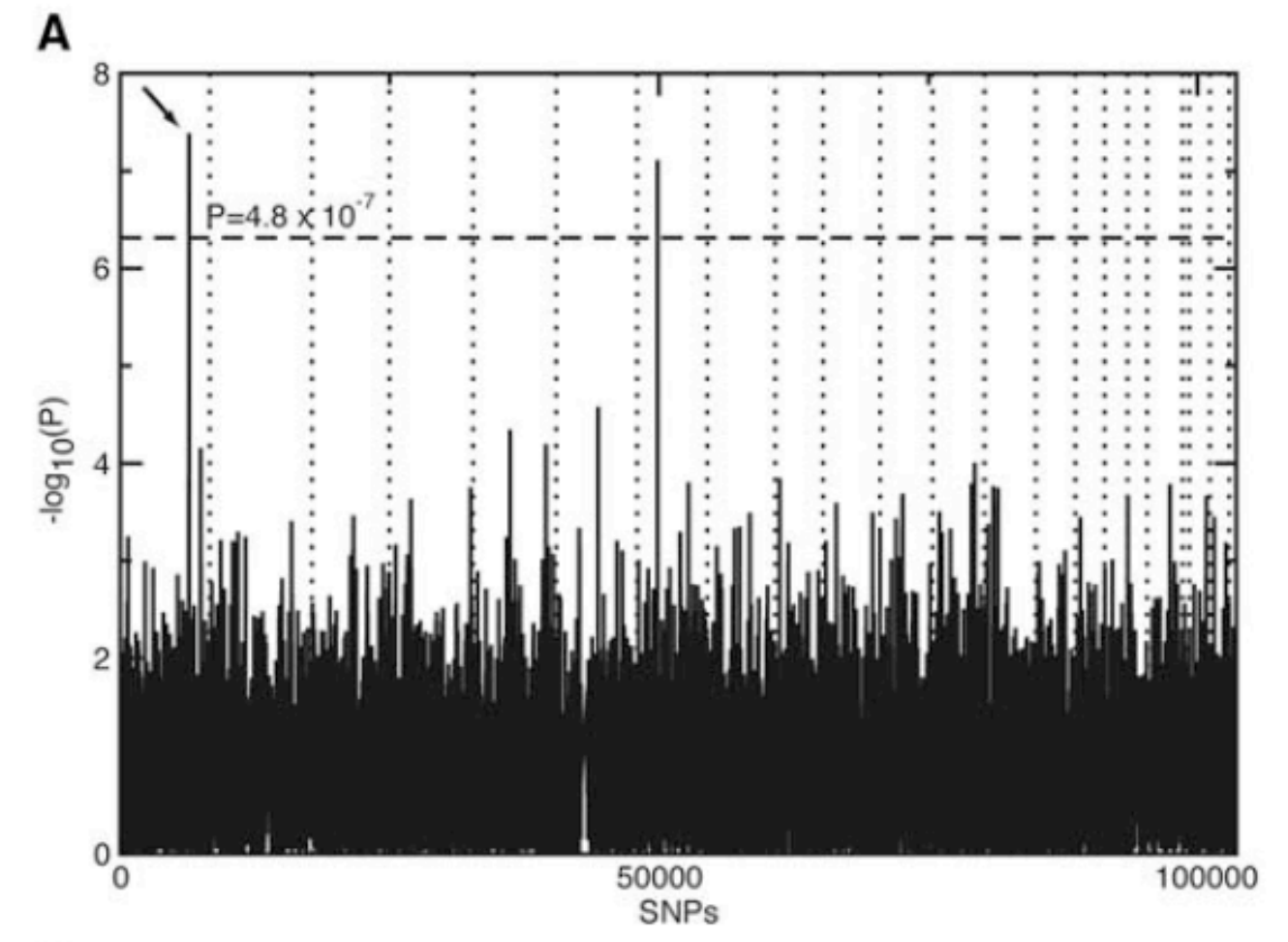
Outline

- AMD study of Klein et al.
- Single marker association models and tests
- Significance of GWA study
- Power of GWA study
- Coverage of GWA study
- Why AMD study was successful?

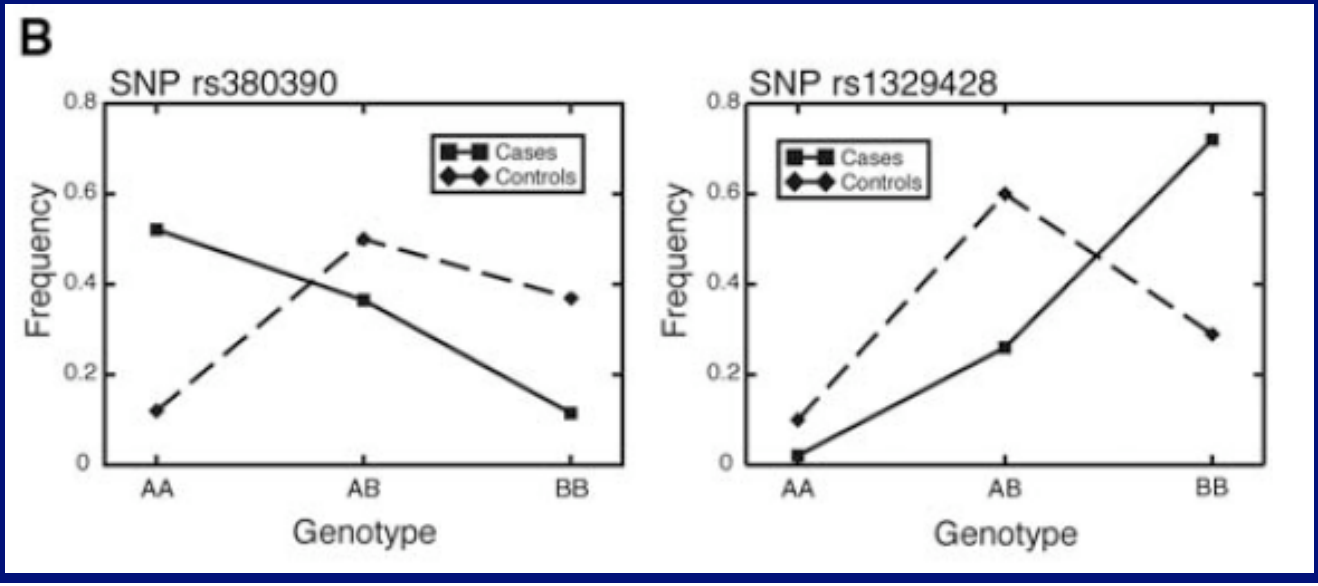
AMD story (Klein et al., 2005)

- Age-related Macular Degeneration (AMD)
- Sample of 96 cases and 50 controls
- 116K → 106K after QC → 104K on autosomes
- Allelic test from 2x2 tables for each SNP
- Bonferroni $P_{GW} = 0.05 / 104K = 4.8 \times 10^{-7}$

Results AMD scan



Characteristics of best hits



Attribute	rs380390 (C/G)	rs1329428 (C/T)
Risk allele	C	C
Allelic association χ^2 nominal P value	4.1×10^{-8}	1.4×10^{-6}
Odds ratio (dominant) (95% CI)	4.6 (2.0–11)	4.7 (1.0–22)
PAR (95% CI)	70% (42–84%)	80% (0–96%)
Frequency in HapMap CEU	0.70	0.82
Odds ratio (recessive) (95% CI)	7.4 (2.9–19)	6.2 (2.9–13)
PAR (95% CI)	46% (31–57%)	61% (43–73%)
Frequency in HapMap CEU	0.23	0.41

Y402H mutation in CFH gene

- GWA scan reveals significant (Bonferroni P=0.004) association with rs380390 in the Complement Factor H (CFH) gene
- Re-sequencing reveals common Y402H mutation
- Independent sample of 1238 cases and 934 controls
- Y402H P-value = 10^{-59}
- Confirmed by many other studies

Outline

AMD study of Klein et al.

Single marker association models and tests

Significance of GWA study

Power of GWA study

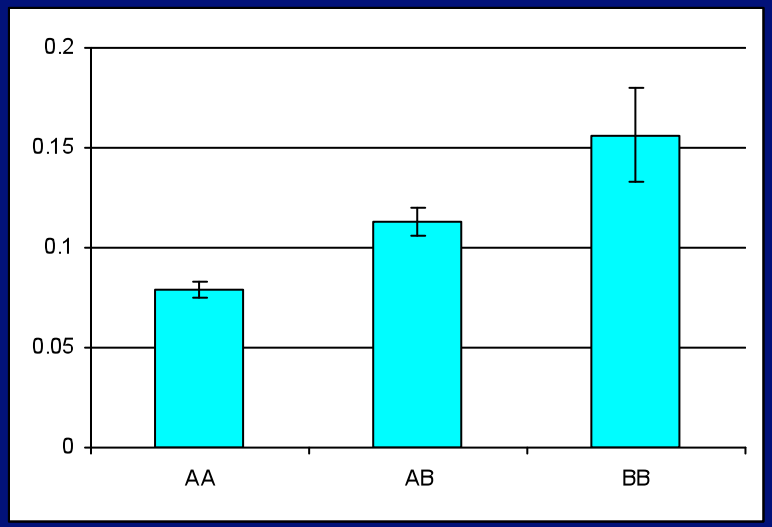
Coverage of GWA study

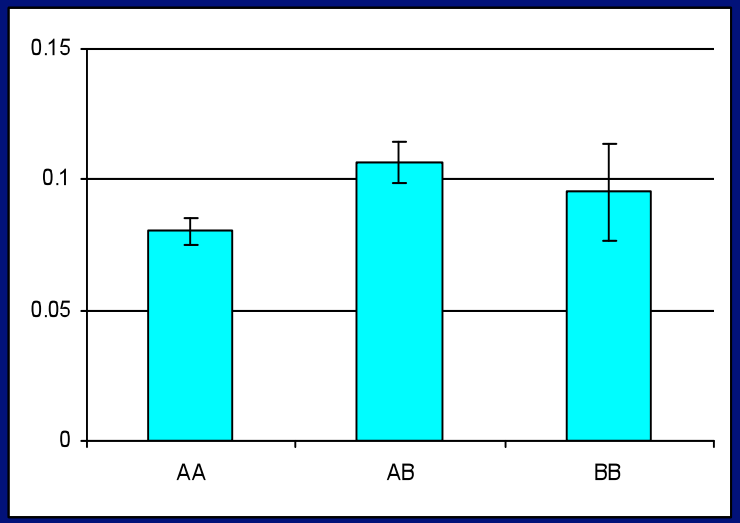
Why AMD study was successful?

Single marker data for a binary outcome

- Data presented as 2 x 3 contingency table

	AA	AB	BB
Cases	R_{AA}	R_{AB}	R_{BB}
Controls	S_{AA}	S_{AB}	S_{BB}





General and additive models

- "Genotypic" or "General" 2 d.f. test
  - Prevalence is not equal across genotypes
  - χ^2 test with 2 d.f. = $\Sigma_i (O_i-E_i)^2/E_i$
- "Trend" or "Additive effect" 1 d.f. test
  - Prevalence is not equal across genotypes
  - It changes constantly with the number of "B" alleles
  - Armitage's trend test: χ^2 test with 1 d.f. = $\Sigma_i (O_i-E_i)^2/E_i$
  - Compared to 2 d.f. test, E_i are computed differently

Dominance and recessive models

- "B is dominant" 1 d.f. test
  - Prevalence is equal for AB and BB, and it is different to that of AA

	AA	B-
Cases	R_{AA}	$(R_{AB}+R_{BB})$
Controls	S_{AA}	$(S_{AB}+S_{BB})$

- "B is recessive" 1 d.f. test
  - Prevalence is equal for AA and AB, and it is different to that of BB

	A-	BB
Cases	$(R_{AA} + R_{AB})$	R_{BB}
Controls	$(S_{AA} + S_{AB})$	S_{BB}

  - χ^2 test with 1 d.f. = $\Sigma_i (O_i-E_i)^2/E_i$
  - Fisher Exact test

“Allelic” model

- Frequency of the “B” allele is different between cases and controls
- 2 x 2 table: counts of A and B in cases and controls

	A	B
Cases	$2 \cdot R_{AA} + R_{AB}$	$R_{AB} + 2 \cdot R_{BB}$
Controls	$2 \cdot S_{AA} + S_{AB}$	$S_{AB} + 2 \cdot S_{BB}$

  - χ^2 test with 1 d.f. = $\Sigma_i (O_i - E_i)^2 / E_i$
  - Fisher Exact test

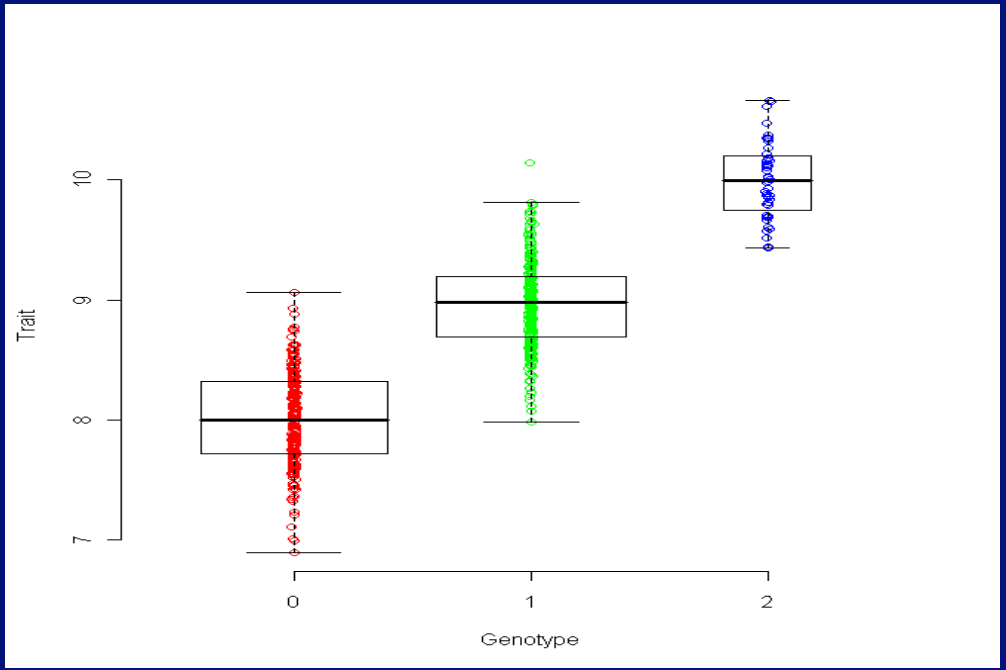
Single marker data for a continuous outcome

Raw data

ID	Trait	Genotype
1	8.3	AA
2	8.5	AB
3	9.1	AB
…		
N	10.0	BB

Table of N, means and SD

	AA	AB	BB
N	222	230	48
Mean	8.01	8.96	10.01
SD	0.39	0.37	0.33



Tests : any test for comparison of means between groups

- General 2 d.f. test: ANOVA
- Dominance/recessive test: T-test, Z-test, …

Linear regression

- Wide range of models; it is possible to include the covariates, interactions, etc.

  Y is the vector of observations
  g_{AB} is an indicator vector for AB genotype
  g_{BB} is an indicator vector for BB genotype
  C is the vector containing covariate

  Expected value of Y is modeled as a linear function

  Linear model: $E[Y] = \mu + \beta_{AB}g_{AB} + \beta_{BB}g_{BB} + \beta_c C$

  Logistic model: $E[\text{logit}(Y)] = \mu + \beta_{AB}g_{AB} + \beta_{BB}g_{BB} + \beta_c C$

  - Additive model: $\beta_{AB} = ½\ \beta_{BB}$
    [logistic without covariates = multiplicative on OR scale]
  - Dominant B model: $\beta_{AB} = \beta_{BB}$
  - Recessive B model: $\beta_{AB} = 0$

Comparison of tests

- **"Allelic" model NOT recommended**
  - Not correct for quantitative outcomes
  - For binary outcomes studied using contingency tables
    - Is correct [only when Hardy-Weinberg Equilibrium holds]
    - When HWE holds it is equivalent to the trend test in 2x3 table
- **Dominant/Recessive tests NOT recommended**
  - most powerful *when the underlying model is guessed rightly*
  - The model is not known *a priori*
- **Recommended:**
  - Score (=Armitage's trend) test 1 d.f.
  - General genotypic test
  - Composite e.g. Fisher Product based tests [Darina of Friday]

Not the most powerful, but maintain power across a range of models

Outline

AMD study of Klein et al.

Single marker association models and tests

Significance of GWA study

Power of GWA study

Coverage of GWA study

Why AMD study was successful?

Multiple Testing

Test claimed “significant” when nominal $p \leq 0.05$

Number of Test conducted	Probability of ≥1 false +
1	0.05
2	0.0975
10	0.401263061
100	0.994079471
1000	1
1000000	1

Multiple testing

- Null hypothesis is true

  $\alpha = 0.05$ (1 test in 20 is “significant”)

- We do two independent tests

  What is type 1 error (α) at nominal p=0.05?

  I.e. what is the chance we will obtain $p \leq 0.05$ in any or both tests?

  $\alpha = P(p_1 \leq 0.05 \;\&\; p_2 \leq 0.05) + P(p_1 \leq 0.05 \;\&\; p_2 > 0.05) +$
  $P(p_1 > 0.05 \;\&\; p_2 \leq 0.05) =$
  $1 - P(p_1 > 0.05 \;\&\; p_2 > 0.05) =$
  $1 - P(p_1 > 0.05)\, P(p_2 > 0.05) =$
  $1 - 0.95^2 = 1 - (1 - 0.05)^2 = 0.0975$

- Thus to keep $\alpha = 0.05$, nominal p-value should be …

  p: solution of $1 - (1 - p)^2 = 0.05$

To have $\alpha = 0.05$ after two tests, nominal p should be 0.02532

Šidak and Bonferroni corrections

If N tests are done, to keep type 1 error of α nominal threshold significance should be *p* given by solution of

$\alpha = 1 - (1 - p)^N$

This is called “Šidak” correction

- Bonferroni correction
  - When $p \rightarrow 0$
  - $(1 - p)^N \approx 1 - N \cdot p$
  - then $p = \alpha/N$

Already for N=5 and α=0.05

- Šidak *p* = 0.0102
- Bonferroni *p* = 0.05/5 = 0.01

Šidak/Bonferroni for GW significance?

- Šidak/Bonferroni correction
  - GW α = 0.05 corresponds to nominal P = 0.05/(# SNPs)
- Problems:
  - Šidak/Bonferroni/FDR assume that tests are independent
  - **SNPs are not** (because of LD)
  - Therefore Šidak/Bonferroni/etc are conservative

  - 550K SNPs were typed, and imputations were done to 2.5M SNPs using HapMap panel. How many tests are done? 0.5M or 2.5M? … or neither?

Empirical GW significance?

- Empirical estimation of GW (experiment-wise) significance gives exact answer, taking the LD structure and phenotype distribution into account
- Works very well for a single one-stage study
- Problems:
  - May be technically demanding (no problem for few dozens of traits, but is a problem for 100s)
  - More complex design: e.g. two-stage, or multiple independent studies
  - Knowledge accumulation (meta-analysis)

Multiple testing burden: fixed threshold

- Pe'er et al, Genetic Epi, 2008, 32: 381-385
- If we measured all common SNPs in the genome, what number of "independent" SNPs could mimic the null distribution of the test statistics?
- ~1M tests ➔ GW 5% ~ nominal P = 0.05/1M = $5\ 10^{-8}$
- To keep in mind:
  - Above is true for CEU (2M for Yoruba)
  - Estimated using $1/600^{th}$ of the genome (ENCODE)

So is my *p*-value significant or not?!

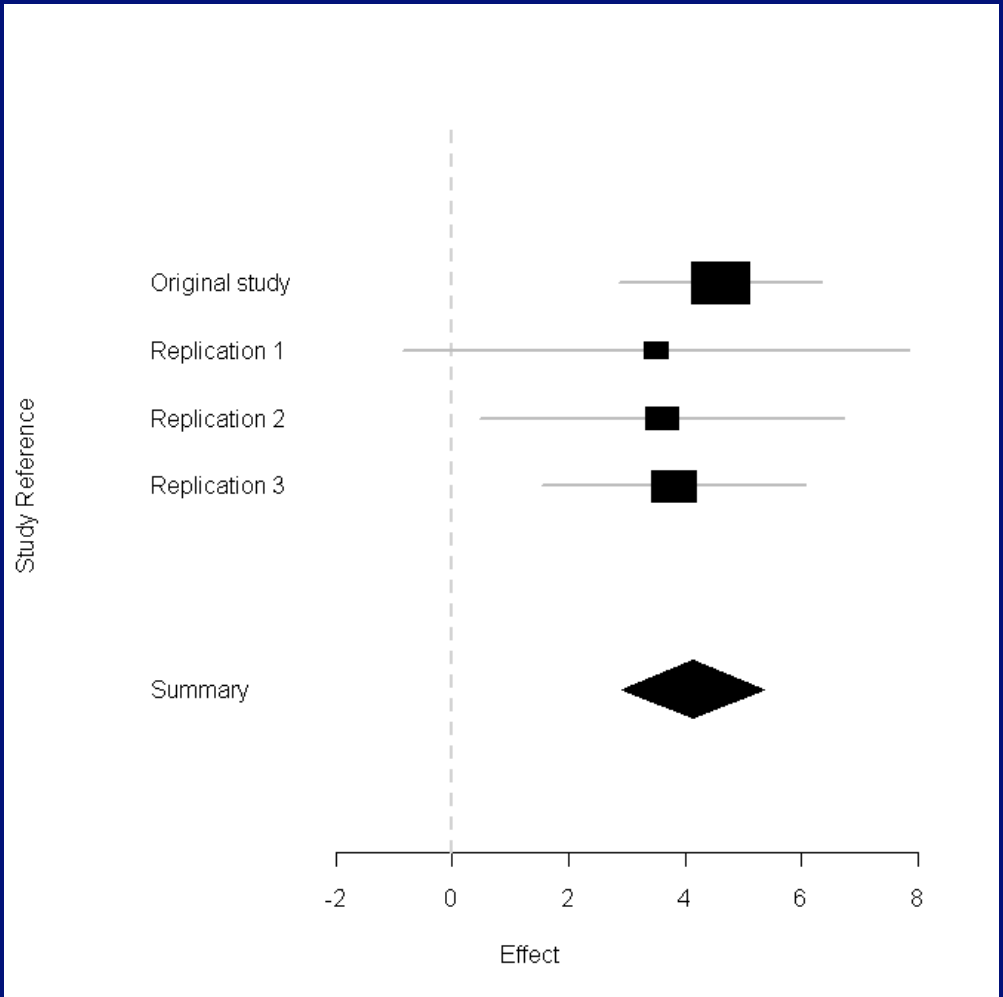
- You (your referees) may be convinced (or not) by a *p*-value which pass
  - Permutation procedure
  - Šidak/Bonferroni/FDR correction
  - $P < 5 \times 10^{-8}$
  - …
- Ultimate answer: **replication**
- This is a way to
  - Achieve “overwhelming significance”
  - Exclude possibility that the finding is “study-specific”

Example

- A genetic study estimates effect of the SNP rs724016*C allele on height as +4.6 mm (s.e. = 0.88)
  - Nominal *p*-value = 2×10^{-7}
  - Permutation-based *p*-value = **0.045**
  - Bonferroni *p*-value = **0.06**
  - Fixed threshold: **$2 \times 10^{-7} > 5 \times 10^{-8}$**
- Is that a true finding or not?
- Replicate!

Replication in three populations

Study	Effect	S.E.	*P*-value
Original	4.6	0.88	2×10^{-7}
Rep 1	3.5	2.21	0.11
Rep 2	3.6	1.59	0.02
Rep 3	2.8	1.15	0.001
Total	4.14	0.62	**2×10^{-11}**



Outline

AMD study of Klein et al.

Single marker association models and tests

Significance of GWA study

Power of GWA study

Coverage of GWA study

Why AMD study was successful?

Estimating power

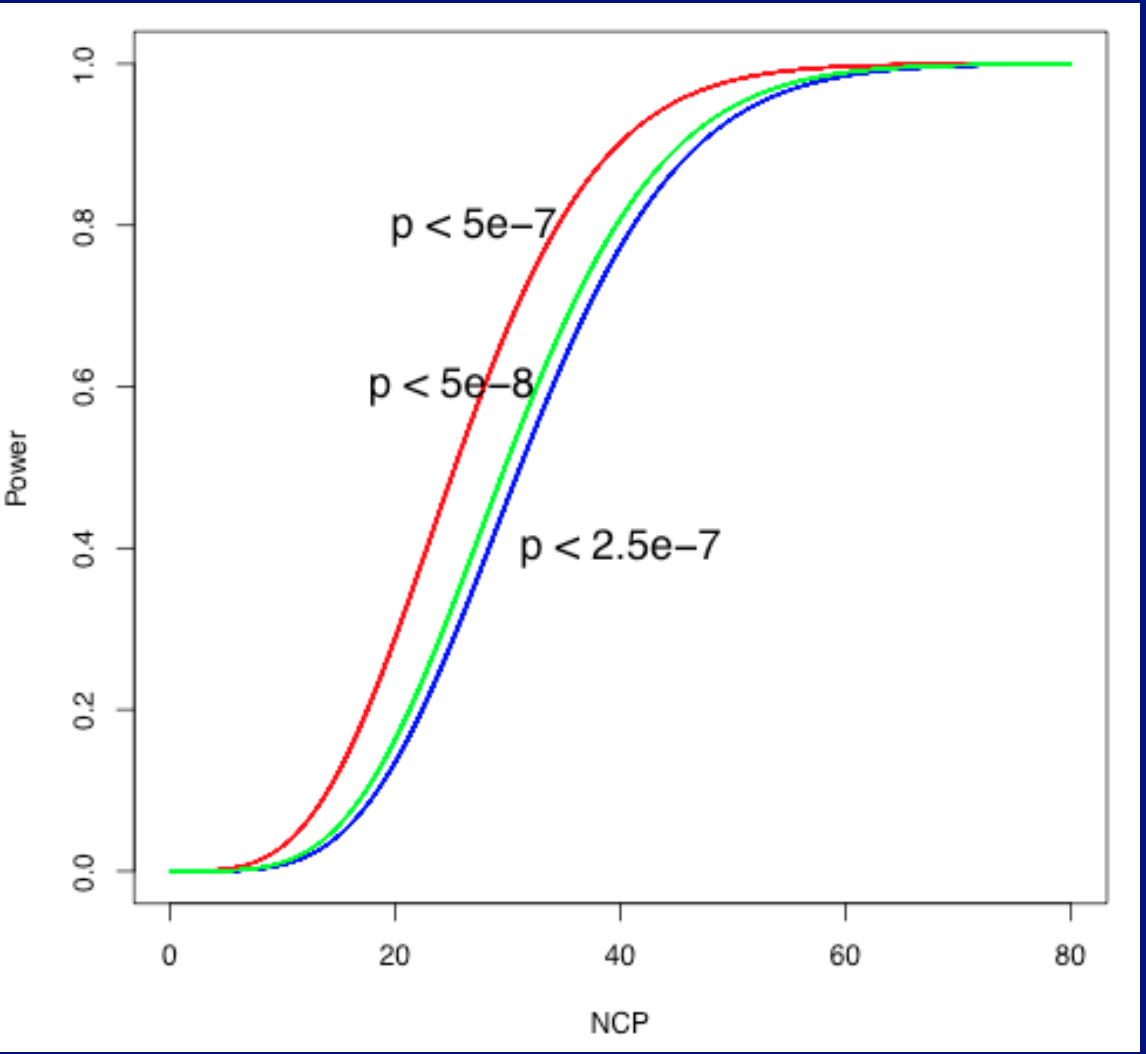
Exact (not known!) model of the gene action is to be assumed

Is study large enough to detect anything assuming some reasonable model?

- Proportion of trait variance (V_{SNP}) explained by the SNP
- The non-centrality parameter (*NCP*)
  - Measures how much the χ^2 test statistic is expected to deviate from it's expectation under the null
  - *NCP* = (no. samples)$\cdot V_{SNP}$
- Power to achieve threshold *T* is $Pr(\chi^2_{NCP} \geq T)$

  Can be computed in R using pchisq(*T*,df=1,ncp=*NCP*,low=FALSE)

Power as function of NCP



Power of GWA study

"Biggest common loci":

- HDL: *CETP* ~ 2.5%
- Total chol.: *APOE* ~ 0.5%
- Height: *HMGA2* ~ 0.3%

Sample size	V_{SNP}	NCP	Power to achieve $p < 5 \times 10^{-8}$
1,000	3%	30	51%
	1%	10	1%
	0.5%	5	<1%
	0.1%	1	<1%
5,000	3%	150	100%
	1%	50	95%
	0.5%	25	33%
	0.1%	5	<1%
10,000	3%	300	100%
	1%	100	100%
	0.5%	50	95%
	0.1%	10	1%

A note on adjustment for the covariates

- Consider *HMGA2* which explains 0.15% of height variation
- Expected power in a study of 14000 people is 20%
- Sex and age together explain ~50% of height variation
- Therefore in the adjusted data the QTL explains 0.3%
- The power to detect it GW is thus 84%

Power to find an untyped variant

- We assumed that a SNP explaining some proportion of variance is in the genotyping set
- Not all genome polymorphisms are genotyped!
- We can detect the region because the “causative” variant may be in LD (measured with r^2) with a typed SNP(s)
  - Proportion of variance explained by not typed variant = V_{var}
  - NCP (if we would have typed the variant) = N x V_{var}
  - Expected proportion of the variance of the trait explained by the SNP, which is in LD (r^2) with the variant is $V_{SNP} = V_{var} \times r^2$
  - NCP from the SNP in LD = N x V_{var} x r^2

Outline

AMD study of Klein et al.

Single marker association models and tests

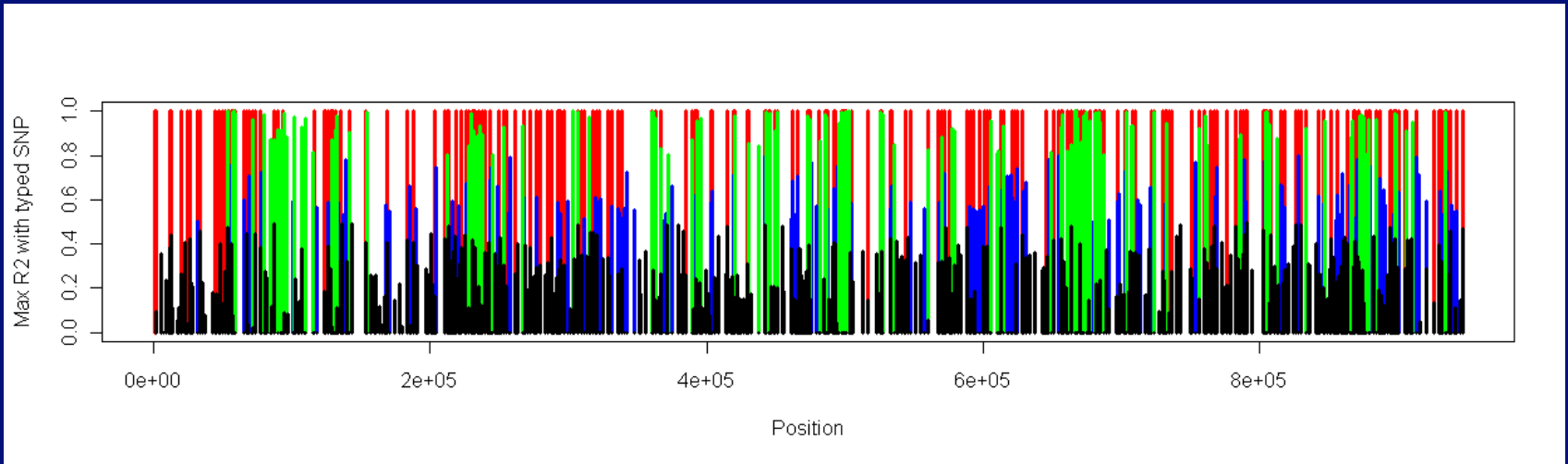
Significance of GWA study

Power of GWA study

Coverage of GWA study

Why AMD study was successful?

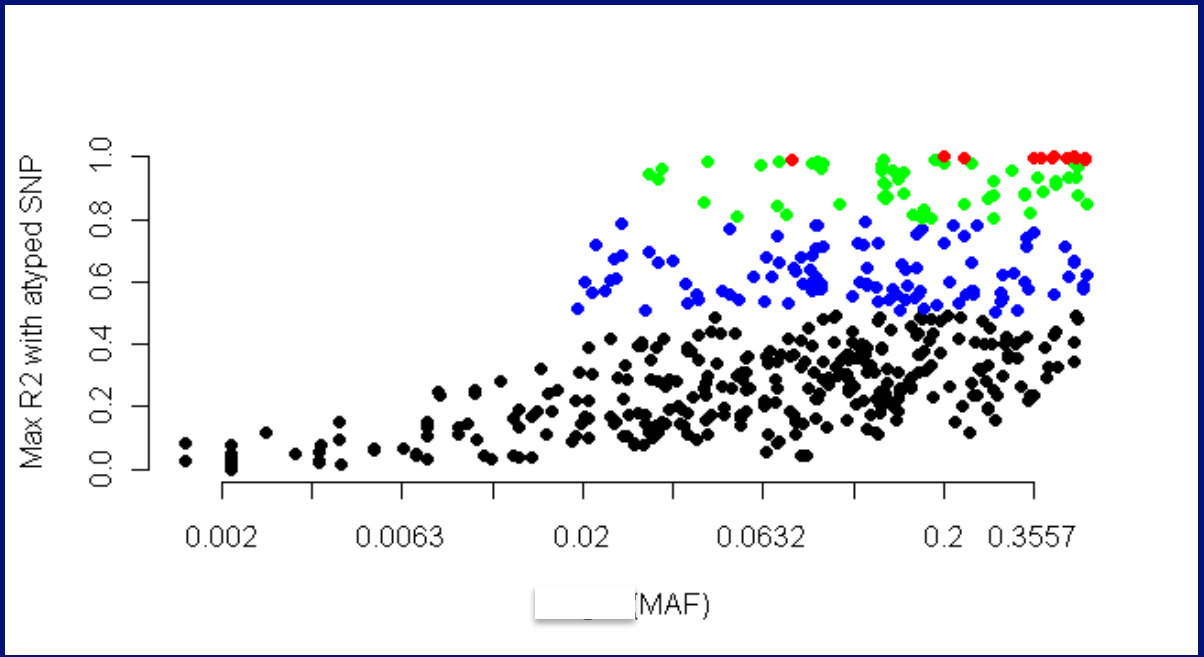
How many SNPs we capture?



- Red: typed SNPs

- Green: SNPs with R2≥0.8 with a typed SNP (well-captured)
- Blue: SNPs with 0.8>R2≥0.5 with a typed SNP (captured)
- Black: SNPs with R2<0.5

Max R2 with a typed SNP depends on MAF



- Selected SNPs are likely to be common (if it is very rare, it is not likely to be known!)
- High R2 between two SNPs is possible only if their frequencies are similar

Genomic coverage by standard panels

- What proportion of common SNPs (MAF≥0.05) are in the genotyped set or are in high LD (r^2>0.8) with at least one genotyped SNP?

		HapMap population		
SNP panel	Type	CEU	JPT+CHB	YRI
Affymetrix 111K	Random	31	31	15
Affymetrix 500K	Random	65	66	41
Affymetrix 1M	Combined	80		
Illumina 300K	Tag	75	63	28
Illumina 550K	Tag	87		
Illumina 1M	Tag	91		

Barret & Cardon, NatGenet, 2006
Anderson et al., AJHG, 2008

Coverage pitfalls

- With 1,000K-2,000K SNP panels we may expect good coverage of common variants for any human population
- Some diseases/traits may be expected to be explained in large part by common variants
- For other disease multiple rare variants may play large role
- Coverage is poor for rare variants

Outline

AMD study of Klein et al.

Single marker association models and tests

Significance of GWA study

Power of GWA study

Coverage of GWA study

Why AMD study was successful?

Why AMD scan was successful?

- Observed
  - $\chi^2 \sim 30$
  - "Wrong" allelic table test
  - $P_{nom} = 4 \cdot 10^{-8}$, P = 0.004 after Bonferroni correction

- 100 cases and 50 controls
- If we plug in the model for rs380390 (q=0.23, GRR=3), expected
  - NCP ~ 16
  - Power to detect a SNP in the 100K: 18%
  - Power to detect a SNP with r^2=0.8: 7%
  - Coverage of Affy 100K at r^2=0.8 is 31%

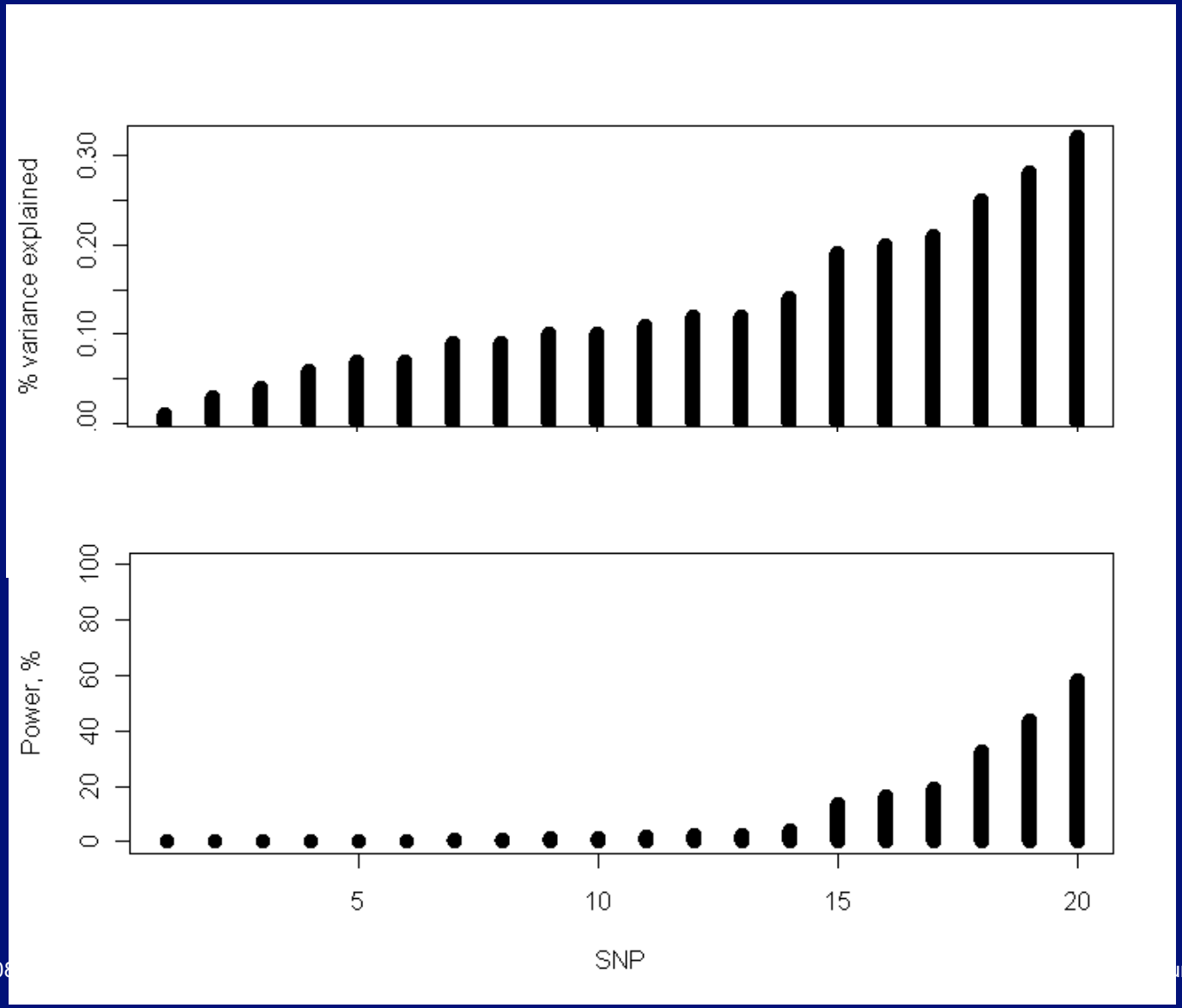
- **A priori chance to detect this common mutation in such study is ~3%**

- …you need some luck…

Painting everything black?!

- Study of height in a cohort of 10,000 people of European origin, using Illumina 300K
  - 25% of common variants are **not** captured
  - Vast majority of rare variants is **not** captured
  - Power to detect even the biggest effect using the directly typed SNP (rs724016 at HMGA2, explains 0.3%) is only 58%!

Power to detect height loci in a study of 10,000 people



What is the chance to miss ALL loci?

- P(miss all) = P(miss locus 1) x P(miss locus 2) x …
- P(miss locus N-1) x P(miss locus N)
- Chance to miss all 20 loci from the height paper of Weedon: only 8%
- Thus you will find at least one with power of 92%
- There are much more than 20 loci involved in height
- **Your chances** (to find ≥1 loci) **are very good!**