Genome-Wide Association analysis tests, power and coverage

Yurii Aulchenko
Erasmus MC Rotterdam
29.08.2007

Outline

- AMD study of Klein et al.
- Single marker tests for association
- What is significant (and what is replicated)?
- Power and 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

$P_{GW} = P_{nom}$ x 103,611 (threshold 4.8 x 10^{-7})

Y402H mutation in CFH gene

GWA reveals rs380390 (P_{GW}=0.004) 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 tests for association

What is significant (and what is replicated)?

Power of coverage of GWA study

Why AMD study was successful?

Single marker tests for comparison of two groups

General 2 x 3 table

	AA	AB	BB	Σ
Cases	r_o	r_1	r_2	R
Controls	s_o	s_1	s_2	S
Σ	n_0	n_1	n_2	N

Tests

Score tests with 1 d.f. (Armitage's trend test)

Score (χ^2) test with 2 d.f.

Tests for genotypic table

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

Score test 1 d.f. (aka Armitage's trend test)

$$X^2 = \frac{N\left[N(r_1+2r_2)-R(n_1+2n_2)\right]^2}{R(N-R)\left[N(n_1+4n_2)-(n_1+2n_2)^2\right]}$$

Equivalent to N x (genotype-phenotype correlation)^2

Decompositions to 2x2 table

Allele B is dominant

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

Allele B is recessive

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

Allelic table (number of alleles in cases and controls)

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

Tests

Score (χ^2) test with 1 d.f.

Fisher Exact test

Composite tests

When the model is right, the test exploiting this model is most powerful; when model is incorrect, the test is not powerful

The model is not known!

Why not testing each model and pick up “the best”?

- $P_{min} = \min \{P_{add}, P_{dom}, P_{rec}\}$

Why not taking the product of P values?

- $P_{FP} = P_{add} \cdot P_{dom} \cdot P_{rec}$

The null distribution is not known and should be derived empirically

Derivation of empirical null distribution for P_{FP}

a) Compute the original P_{FP}

b) Permute the case-control status at random

   Now any true association between the genotype and phenotype is destroyed

c) Re-compute and save the test statistics P_{null}

d) Repeat steps (b) and (c) say 10,000 times

e) The 5% quantile of the distribution of P_{null} gives threshold for 5% significance

f) Compare the original test statistics to this threshold. If the original test statistics P_{FP} is less then threshold, you can claim 5% significance

Comparison of tests

Test in allelic table is NOT recommended

- Is correct only when HWE holds
- When HWE holds it is equivalent to the Armitage's trend test in 2x3 table

Dominant/Recessive 2x2 test NOT recommended

- most powerful *when the underlying model is guessed rightly*
- The model is not known *a priori*

Recommended:

- Score test 1 d.f.
- General genotypic test
- FP test for additive/dominant/recessive…

Logistic regression

Wide range of models possible
Inclusion of covariates, interactions

Y is the vector of observations (1=case, 0=control)
g_{AB} is an indicator vector for AB genotype
g_{BB} is an indicator vector for BB genotype

$$E[Y] = \exp\{\alpha + \beta_{AB}g_{AB} + \beta_{BB}g_{BB}\}/(1 - \exp\{\alpha + \beta_{AB}g_{AB} + \beta_{BB}g_{BB}\})$$

Additive model: $\beta_{AB} = \frac{1}{2}\beta_{BB}$
Dominant model: $\beta_{AB} = \beta_{BB}$
Recessive model: $\beta_{AB} = 0$

Single marker tests for comparison of quantitative traits

Analysis in not selected cohort

- Comparison of means between genotype groups
  - ANOVA, t-test, Z-test,…
  - Kruskal-Wallis test
- Linear Models (Regression)
  - Wide range of models possible
  - Inclusion of covariates, interactions

Sampling from extremes

- back to case-control
- specific tests

Especially for QTs, estimating empirical significance is a good idea (asymptotic tests rely on Normal distribution)

Outline

AMD study of Klein et al.

Single marker tests for association

What is significant (and what is replicated)?

Power of coverage of GWA study

Why AMD study was successful?

Multiple testing

Null hypothesis is true

- $\alpha = 0.05$ (1 test in 20 is “significant”)
- χ^2 test with 1 d.f. has a threshold of 3.84

We do two independent tests

- What is the chance we will obtain $\chi^2 \geq 3.84$ in any or both?
- I. e. what is type 1 error (α) at the threshold of 3.84?

  $\alpha = 1 - (1 - 0.05)^2 = 0.0975$

To keep $\alpha = 0.05$

Solve $1 - (1 - x)^2 = 0.05$

Nominal significance x should be 0.02532

Threshold χ^2 is 5.002

Šidak and Bonferroni corrections

If N tests are done, to keep type 1 error of α nominal threshold significance should be *x*

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

Šidak correction is given by solution of this equation

Bonferroni correction

When $x \rightarrow 0$

$(1 - x)^N \approx 1 - N \cdot x$

then $x = \alpha/N$

Already for N=5 and α=0.05

Šidak x = 0.0102

Bonferroni x = 0.05/5 = 0.01

Estimating GW significance

Bonferroni correction

- GW α = 0.05 corresponds to nominal P = 0.05/(# SNPs)

FDR procedures

- Less conservative compared to Bonferroni
- Benjamini & Hochberg 1995: R library “GenABEL”
- Storey 2006: R library “qvalue”

SNPs are not independent (LD), therefore Bonferroni and FDR are conservative

Use permutation tests!

Empirical GW significance

a) Compute the original test statistics GW
b) Permute the case-control status at random
    Now any true association between the genotype and phenotype is destroyed
c) Re-compute the test statistics GW, and save the maximal statistics T_{max}
d) Repeat steps (b) and (c) say 1,000 times
e) The 95% quantile of the distribution of T_{max} gives the empirical genome-wide threshold at α=0.05
f) Compare the original test statistics to this threshold

What problems are (not) solved by empirical procedures

Permutation of phenotypes generate null distribution of the test statistics (GW)

Is the correct and the only practically available general procedure when

- Genotypes are correlated
- Phenotypes are correlated
- Distribution of the test statistics is not known (e.g. P_{min})

Is NOT correct when there are additional, other than genotypes themselves, sources of correlation between genotypes and phenotypes, e.g. stratification

Estimating GW significance threshold

Desired α = 0.05, 500K SNP array

Threshold

- nominal P = $0.05/(5\ 10^5) = 0.0000001 = 10^{-7}$
- χ^2 test with 1 d.f. 28.37
- χ^2 test with 2 d.f. 32.34

This is reasonably close to the empirical threshold

What is a significant and what is a replicated finding?

Significant: has experiment-wise type 1 error of 5%

Replicated (puristic): given **significant** finding in the first study, a second independent study shows experiment-wise **significant** finding

Two-stage design

- Stage 1 (discovery): one sample is used for GWA screening
- Stage 2 (replication): other sample is used typed for a number of SNPs selected based on the first stage

Significant and replicated in 2-stage

Replicated:

S1: a SNP is 5% GW significant (**significant**)

S2: the SNP is 5% experiment-wise significant in S2 data (**replicated**)

Significant:

S1: no SNPs is 5% GW significant

S2: a SNP is 5% experiment-wise significant in S2 data or in joint analysis (**significant**)

Nothing:

S1: no SNPs is 5% GW significant

S2: no SNP is 5% experiment-wise significant in S2 data or in joint analysis

Outline

AMD study of Klein et al.

Single marker tests for association

What is significant (and what is replicated)?

Power of coverage of GWA study

Why AMD study was successful?

Estimating power of GWA study

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)

- Expected mean (-1) of your test statistics
- NCP = (no. samples)$\cdot V_{SNP}$

Power to achieve threshold T is $Pr(\chi^2_{NCP} \geq T)$

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

Power of GWA study

Sample size	V_{SNP}	NCP	Power to achieve $\chi^2 \geq 28.37$
500	1%	5	0.1%
	2%	10	2%
	3%	15	7%
1000	1%	10	2%
	2%	20	20%
	3%	30	56%
2000	1%	20	20%
	2%	40	84%
	3%	60	99%
5000	1%	50	96%
	2%	100	100%
	3%	150	100%

A note on adjustment for the covariates

Consider a QTL which explains 1% of height variation

Expected power in a study of 2000 people is 20%

Sex and age together explain ~50% of height variation

Therefore in the adjusted data the QTL explains 2%

The power to detect it GW is thus 84%

Power to detect SNP in LD with genotyped SNPs

We assumed that the SNP explaining some proportion of variance is in the genotyping set

If it is not, but is in LD (as measured by r^2) with the causal variant, this scales NCP by r^2.

Power of GWA study with LD

Sample size	V_{SNP}	NCP	Power to achieve $\chi^2 \geq 28.37$	NCP·r^2	Power to achieve $\chi^2 \geq 28.37$
500	1%	5	0.1%	4	0.04%
	2%	10	2%	8	0.6%
	3%	15	7%	12	3%
1000	1%	10	2%	8	0.6%
	2%	20	20%	16	9%
	3%	30	56%	24	33%
2000	1%	20	20%	16	9%
	2%	40	84%	32	63%
	3%	60	99%	48	95%
5000	1%	50	96%	40	84%
	2%	100	100%	80	99.99%
	3%	150	100%	120	100%

Painting everything black?

Cohort of 2000 people

- Power to detect a SNP explaining 1% is 9%
- This is power to detect a **particular single** QTL

How many SNPs like this we may expect?

- Assume 20 independent SNPs
- Chance to detect **none** of these is $(1-0.09)^{10} = 0.15$
- Thus power to detect **at least one** is 85%

Genomic coverage by standard panels

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

Table 1 Genomic coverage of commercial GWAS products for common SNPs at $r^2 \geq 0.8$

		CEU		JPT+CHB		YRI	
	Type	Coverage (%)	Mean r^2	Coverage (%)	Mean r^2	Coverage (%)	Mean r^2
Illumina HumanHap300	Tag	75	0.961	63	0.964	28	0.961
Affymetrix 500K	Random	65	0.975	66	0.974	41	0.971
Affymetrix 111K	Random	31	0.960	31	0.957	15	0.957
Affymetrix 500k + 175K tag	Combination	86	0.975	79	0.978	49	0.973
Illumina Human-1	Gene	26[a]	0.957	28[a]	0.955	12[a]	0.956

Barret & Cardon, NatGenet, 2006

Coverage pitfalls

With 1,000K SNP panels we may expect good coverage of common variants for any human population

Some diseases (e.g. T2D) are expected to be explained in large part by common variants

For other disease multiple rare variants may play large role

Coverage is very poor (<10%) for such rare variants

Outline

AMD study of Klein et al.

Single marker tests for association

What is significant (and what is replicated)?

Power of 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 is 2%

…you need some luck…