Power, Sample Size and Design Considerations

SISG – Module 15

Dr Loic Yengo
l.yengo@imb.uq.edu.au
Institute for Molecular Bioscience
The University of Queensland

Outline

- Statistical power
- Ascertainment bias
- Designing new studies

Outline

- **Statistical power**
- Ascertainment bias
- Designing new studies

What is power?

When we set up a statistical test

- The null hypothesis (H_0) is EITHER
  - true
  - false
- With the data available we EITHER
  - reject the null hypothesis
  - fail to reject the null hypothesis

	Null hypothesis is true	**Null hypothesis is false**
Reject the null hypothesis	Type I error False positive	Correct Outcome True positive
Fail to reject the null hypothesis	Correct Outcome True negative	Type II error False Negative

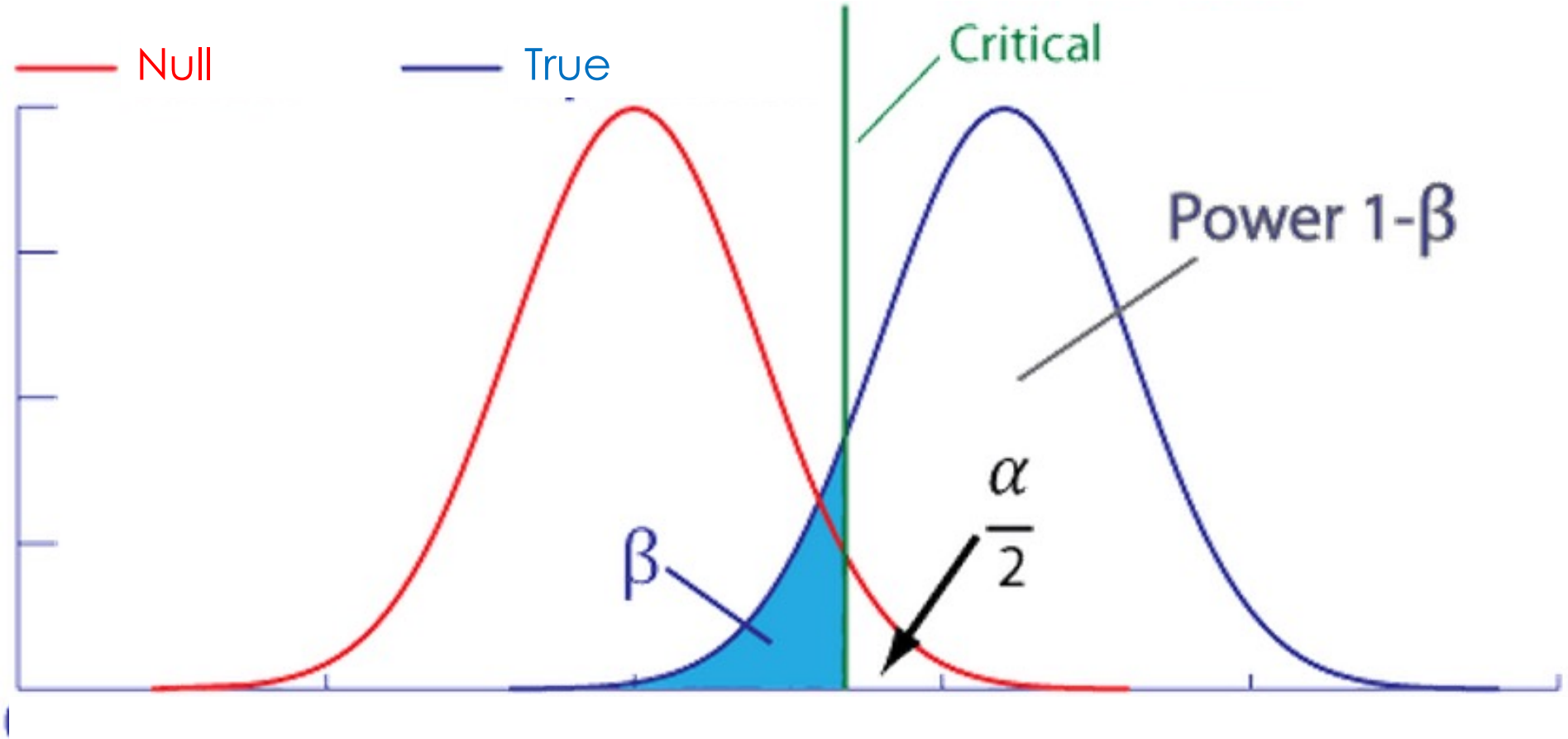
Power = probability of rejecting the null hypothesis when the null hypothesis is false

=1 – probability of failing to reject the null hypothesis when the null hypothesis is false

= 1 – probability(Type II error)

Power depends on **statistical test**, **effect size** to be detected, **sample size**, **acceptable level of Type I** error
Non-centrality parameter depends on statistical test, effect size to be detected, sample size

Power



β = probability of rejecting the null hypothesis when the alternative hypothesis is true
α = probability of rejecting the null hypothesis when the null hypothesis is true
Variance about mean values depends on sample size

Multiple testing

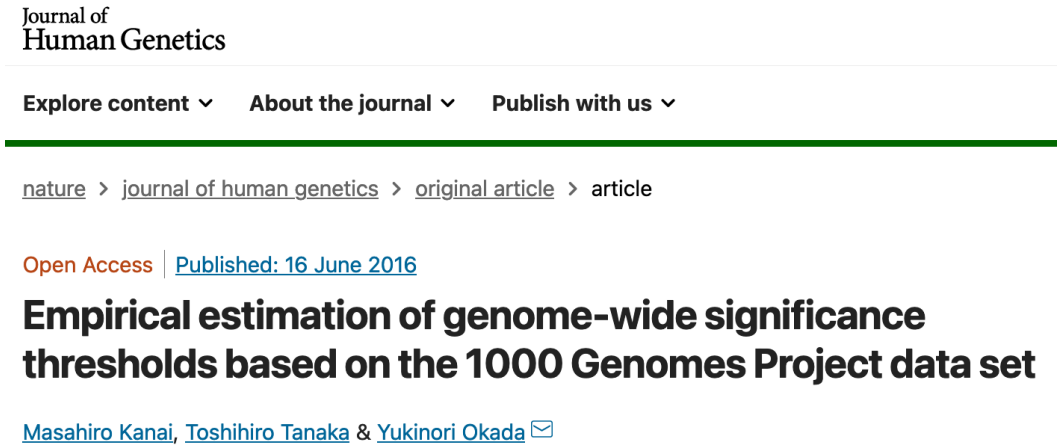
- For many tests, the distribution of p-values follows a uniform distribution under the null: $\mathbf{Prob}[P < \alpha] = \alpha$

- When running $\mathbf{M}$ (independent) tests with significance threshold α for each of them then the expected number of "detected" signals under the null is $\mathbf{M}\alpha$.

- To reduce that number it is often recommended use a more significant threshold (e.g. 0.05/M; a.k.a Bonferroni Correction)

Genome-wide significance threshold

- Corresponds to a Bonferroni threshold 0.05/1,000,000 (Me=1,000,000 – HapMap Project)
- Ancestries specific
- Threshold for WGS/Imputation-based GWAS is much smaller



Journal of Human Genetics

Explore content | About the journal | Publish with us

nature > journal of human genetics > original article > article

Open Access | Published: 16 June 2016

Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set

Masahiro Kanai, Toshihiro Tanaka & Yukinori Okada

Key ingredients of power

- The distribution of the test statistic (Q) under the null hypothesis (H_0) **AND** under the alternative hypotheses (H_1)
- The sample size (N)
- The significance threshold (e.g., α=5e-8)

How much power is enough?

- Great question!
- We can never reach 100% under realistic assumptions
- Rule of thumb is >80%

Why do I need a power calculation?

- To make sure that you have enough data to answer your question
- To set up baseline expectations that can inform what assumptions are violated in your study
- To convince funders!

Single locus disease model

Single locus disease model:

G = genotype; D=disease; K = overall disease risk in population;

p = risk allele frequency;

	P(G)	P(D|G)
aa	$(1\text{-}p)^2$	f_0
Aa	$2p(1\text{-}p)$	f_0R
AA	p^2	f_0R^2

$f_0=K/(1+p(R\text{-}1))^2$

Single locus disease model

Single locus disease model:

G = genotype; D=disease; K = overall disease risk in population;

p = risk allele frequency;

f_0 = baseline risk for homozygote non-risk allele – UNKNOWN

R = relative risk for heterozygote; assume risk is multiplicative (on this scale)

	P(G)	P(D|G)
aa	$(1\text{-}p)^2$	f_0
Aa	$2p(1\text{-}p)$	f_0R
AA	p^2	f_0R^2

$f_0 = K/(1+p(R\text{-}1))^2$

Single locus disease model

Single locus disease model:

G = genotype; D=disease; K = overall disease risk in population;

p = risk allele frequency;

f_0 = baseline risk for homozygote non-risk allele – UNKNOWN

R = relative risk for heterozygote; assume risk is multiplicative (on this scale)

	P(G)	P(D|G)	P(DandG) =P(D|G)P(G)
aa	$(1-p)^2$	f_0	$(1-p)^2 f_0$
Aa	$2p(1-p)$	f_0R	$2p(1-p) f_0R$
AA	p^2	f_0R^2	$p^2 f_0R^2$
			Sum= K

$$P(\text{Disease})=K=f_0(1-p)^2+f_0R2p(1-p)+f_0R^2p^2=f_0(1+p(R-1))^2$$

$$f_0=K/(1+p(R-1))^2$$

Single locus disease model

Single locus disease model:

G = genotype; D=disease; K = overall disease risk in population;

p = risk allele frequency;

f_0 = baseline risk for homozygote reference allele – UNKNOWN

R = relative risk for heterozygote; assume risk is multiplicative (on this scale)

	P(G)	P(D|G)	P(DandG) =P(D|G)p(G)	P(G|D) =P(DandG)/P(D)
aa	$(1\text{-}p)^2$	f_0	$(1\text{-}p)^2 f_0$	$(1\text{-}p)^2 f_0/K$
Aa	$2p(1\text{-}p)$	f_0R	$2p(1\text{-}p) f_0R$	$2p(1\text{-}p) f_0R/K$
AA	p^2	f_0R^2	$p^2 f_0R^2$	$p^2 f_0R^2/K$
			Sum= K	

$$P(\text{Disease})=K=f_0(1\text{-}p)^2 + f_0R2p(1\text{-}p) + f_0R^2p^2 = f_0(1+p(R\text{-}1))^2$$

$$f_0=K/(1+p(R\text{-}1))^2$$

Allele Frequency in Cases

Single locus disease model:

G = genotype; D=disease; K = overall disease risk in population

	P(G)	P(D|G)	P(DandG) =P(D|G)p(G)	P(G|D) =P(DandG)/P(D)
aa	$(1-p)^2$	f_0	$(1-p)^2 f_0$	$(1-p)^2 f_0/K$
Aa	$2p(1-p)$	f_0R	$2p(1-p) f_0R$	$2p(1-p) f_0R/K$
AA	p^2	f_0R^2	$p^2 f_0R^2$	$p^2 f_0R^2/K$
			Sum= K	

$P(\text{Disease})=K=f_0(1-p)^2 + f_0R2p(1-p) + f_0R^2p^2 = f_0(1+p(R-1))^2$

$f_0=K/(1+p(R-1))^2$

$p_1 = \frac{1}{2} P(Aa|D)+P(AA|D)$ Allele frequency in cases

$= f_0pR\big((1-p)+pR\big)/K = \frac{pR}{(1+p(R-1))}$

Find allele frequency in controls in the same way

$p_0 = \frac{p}{1-K}\left(1-\frac{KR}{(1+p(R-1))}\right)$

Allele Frequency in Controls

Single locus disease model:

G = genotype; D=disease; K = overall disease risk in population

	P(G)	P(D'|G)	P(D'andG) =P(D'|G)p(G)	P(G|D') =P(D'andG)/P(D')
aa	$(1-p)^2$	$(1-f_0)$	$(1-p)^2\ (1-f_0)$	$(1-p)^2\ (1-f_0)/(1-K)$
Aa	$2p(1-p)$	$(1-f_0R)$	$2p(1-p)\ (1-f_0R)$	$2p(1-p)\ (1-f_0R)/(1-K)$
AA	p^2	$(1-f_0R^2)$	$p^2\ (1-f_0R^2)$	$p^2\ (1-f_0R^2)/(1-K)$
			Sum= 1-K	

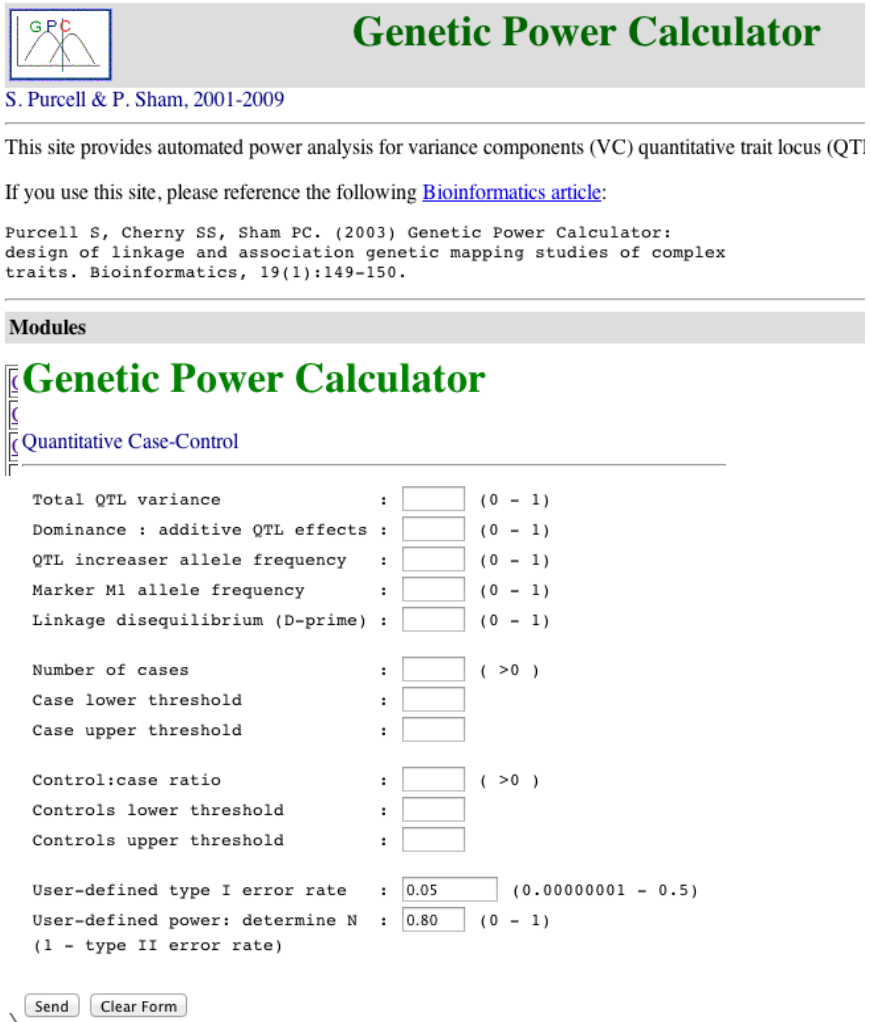
$f_0=K/(1+p(R-1))^2$

$p_0 = ½\ P(Aa|D')+P(AA|D')$ Allele frequency in controls

$$= \frac{p}{1-K}\left(1-\frac{KR}{(1+p(R-1))}\right)$$

Using the single locus disease model to calculate power in an association study

Genetic Power Calculator



Genetic Power Calculator

S. Purcell & P. Sham, 2001-2009

This site provides automated power analysis for variance components (VC) quantitative trait locus (QT

If you use this site, please reference the following Bioinformatics article:

Purcell S, Cherny SS, Sham PC. (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics, 19(1):149-150.

Modules

Genetic Power Calculator

Quantitative Case-Control

Total QTL variance : (0 - 1)
Dominance : additive QTL effects : (0 - 1)
QTL increaser allele frequency : (0 - 1)
Marker M1 allele frequency : (0 - 1)
Linkage disequilibrium (D-prime) : (0 - 1)

Number of cases : (>0)
Case lower threshold :
Case upper threshold :

Control:case ratio : (>0)
Controls lower threshold :
Controls upper threshold :

User-defined type I error rate : 0.05 (0.00000001 - 0.5)
User-defined power: determine N : 0.80 (0 - 1)
(1 - type II error rate)

Send Clear Form

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A) : (0 - 1)
Prevalence : (0.0001 - 0.9999)
Genotype relative risk Aa : (>1)
Genotype relative risk AA : (>1)

D-prime : (0 - 1)
Marker allele frequency (B) : (0 - 1)

Number of cases : (0 - 10000000)
Control : case ratio : (>0)
(1 = equal number of cases and controls)

Unselected controls? (* see below)

User-defined type I error rate : 0.05 (0.00000001 - 0.5)
User-defined power: determine N : 0.80 (0 - 1)
(1 - type II error rate)

Process Reset

Created by Shaun Purcell 24.Oct.2008

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	.2	(0 - 1)
Prevalence	:	.01	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	.2	(0 - 1)
Number of cases	:	5000	(0 - 10000000)
Control : case ratio	:	1	(>0)
			(1 = equal number of cases and controls)
		☐	Unselected controls? (* see below)
User-defined type I error rate	:	0.00000005	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

Case-control statistics: allelic 1 df test (B versus b)

Sample NCP = 28.59

Alpha	Power	N cases for 80% power
0.1	0.9999	1081
0.05	0.9996	1372
0.01	0.9972	2042
0.001	0.9802	2985
5e-08	0.4586	6924

Power of a case-control study

Notations

p = frequency of the risk allele in the population
$\hat{p}_1$ = estimated frequency of the risk allele in cases
$\hat{p}_0$ = estimated frequency of the risk allele in controls
π = proportion in a sample of N that are cases
$\bar{p}$ = $\pi p_1 + (1-\pi)p_0$, expected allele frequency in the entire sample

Test statistic

$$Z^2 = \frac{(\hat{p}_1 - \hat{p}_0)^2}{\widehat{\text{var}}\,[\hat{p}_1 - \hat{p}_0]} = \frac{(\hat{p}_1 - \hat{p}_0)^2}{\frac{\hat{p}_1(1-\hat{p}_1)}{N\pi} + \frac{\hat{p}_0(1-\hat{p}_0)}{N(1-\pi)}}$$

Z^2 follows a χ_1^2 (chi-square distribution with 1 degree of freedom) under the null and a non-central χ_1^2 distribution under the alternative.

Non-centrality parameter

$$\text{NCP} = \frac{(E[\hat{p}_1] - E[\hat{p}_0])^2}{\text{var}[\hat{p}_1 - \hat{p}_0]} = \frac{(p_1 - p_0)^2}{\frac{\bar{p}(1-\bar{p})}{2N}\left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)} = \mathbf{\frac{2N\pi(1-\pi)(p_1 - p_0)^2}{\bar{p}(1-\bar{p})}}$$

Yang et al (2009) Comparing Apples and Oranges: Equating the Power of Case-Control and Quantitative Trait Association Studies. Genetic Epidemiology

How does it look in R…

```
# Input Parameters
N=10000        # Total sample size
p=0.2          # Risk allele frequency in the population
pi=0.5         # proportion of cases
alpha=5e-8     # significance threshold
R=1.2          # Relative Risk
K=0.01         # Population prevalence

p1         <- p * R  / (p * R + 1 - p)
p0         <- ( p/(1-K) ) * (1 - K*R/(1 + p*(R-1) ))
pb         <- pi*p1 + (1-pi)*p0
NCP        <- 2*N*pi*(1-pi)*((p1-p0)^2) / ( pb*(1-pb) )
Theshold   <- qchisq(p=alpha,df=1,lower.tail = FALSE)
Power      <- pchisq(q=Theshold,df=1,ncp=NCP,lower.tail = FALSE)
Power
```

Yang et al (2009) Comparing Apples and Oranges: Equating the Power of Case-Control and Quantitative Trait Association Studies. Genetic Epidemiology

NCP for quantitative traits

- $\mathrm{NCP} = \frac{Nq^2}{1-q^2}$
- $q^2 = 2p(1-p)\beta^2$ is the (expected) variance explained by tested SNPs.
- p=Allele Frequency
- β= per allele effect (e.g., (expected) GWAS BETA)

Outline

- Statistical power
- **Ascertainment bias**
- Designing new studies

Non-random sampling

- Studies are generally not random samples from the *Population*
- **Example 1**: Population = all individuals affected with a certain disease vs Sample = diagnosed individuals, hospitalized
- **Example 2**: General Population vs Sample = people who enroll in studies (healthy enough to do so, hypochondriac, very keen people)
- **Example 3**: Batch effects

Implications of non-random sampling

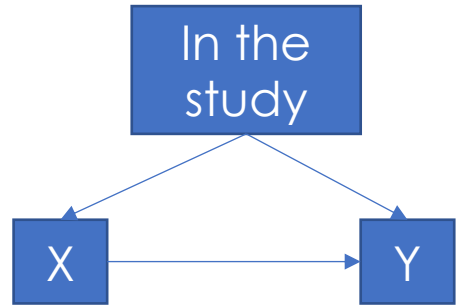
- Biases in estimated effects
- False positive or false negative associations
- Loss of power

How to detect these issues?

- Compare phenotypes of participants with that of the general population (e.g., census data)
- Compare allele frequencies in sample vs reference population (e.g., 1000 Genomes, UK10K, TOPMed, etc)
- Principal component analyses can inform ascertainment biases

How to correct these issues?

- Not always possible (e.g., perfect confounding between disease status and ancestries/genotyping batches)
- Using regression methods to adjust for confounders
- For population stratification => use within-family designs (non-standard GWAS lecture)
- Sensitivity analyses
- Replication!



Outline

- Statistical power
- Ascertainment bias
- **Designing new studies**

Questions to ask

- What population?
  - Need for genetic studies in understudied populations
  - Diseased population (prognostic / complications GWAS)

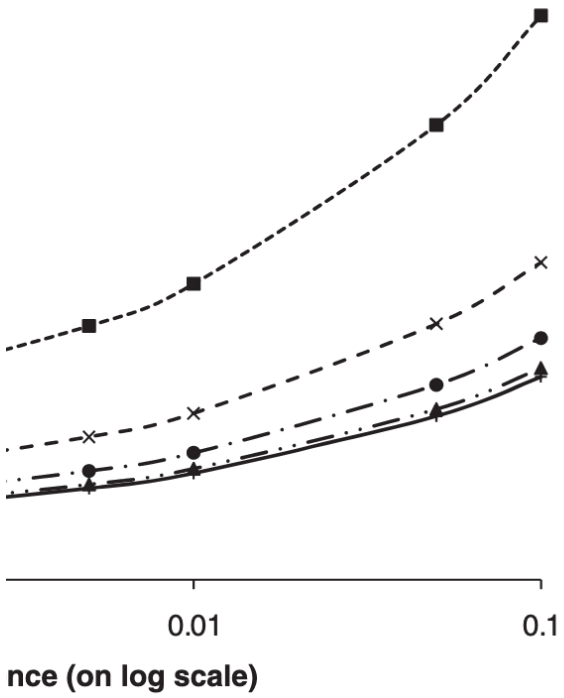
- What sample size (power calculation)?
  - Large! obviously depends on budget.

- What genotyping platform (WGS? WES?)
  - High-coverage WGS is the best investment
  - Low-coverage WGS + imputation is extremely appealing

Other considerations

Case-control design or continuou
Obesity)

ratio of NCP (Fig. 3). When $NCP_{01} = NCP_{QT}$ and $v = 0.5$, Equation (3) reduces to $N_{01}/N_{QT} = 4(1-K)^2/i^2$, which ranges from 0.26 to 1.05 for K from 0.00001 to 0.1. Therefore, for diseases with prevalence <0.1, CC association study with equal numbers of cases and controls needs smaller sample size than QT association study to achieve equivalent power. For example, a CC association study of schizophrenia with prevalence of 0.01 [Sullivan et al., 2003], needs only 55% of sample size required for a QT association study of height, assuming equal effect sizes on the liability scale for schizophrenia and the observed scale of height.



Other considerations

Longitudinal or cross-sectional?

1) Useful question if interested in time-varying effects of SNPs

2) If measurement error is large

3) Confounding may increase over time (drop-out)