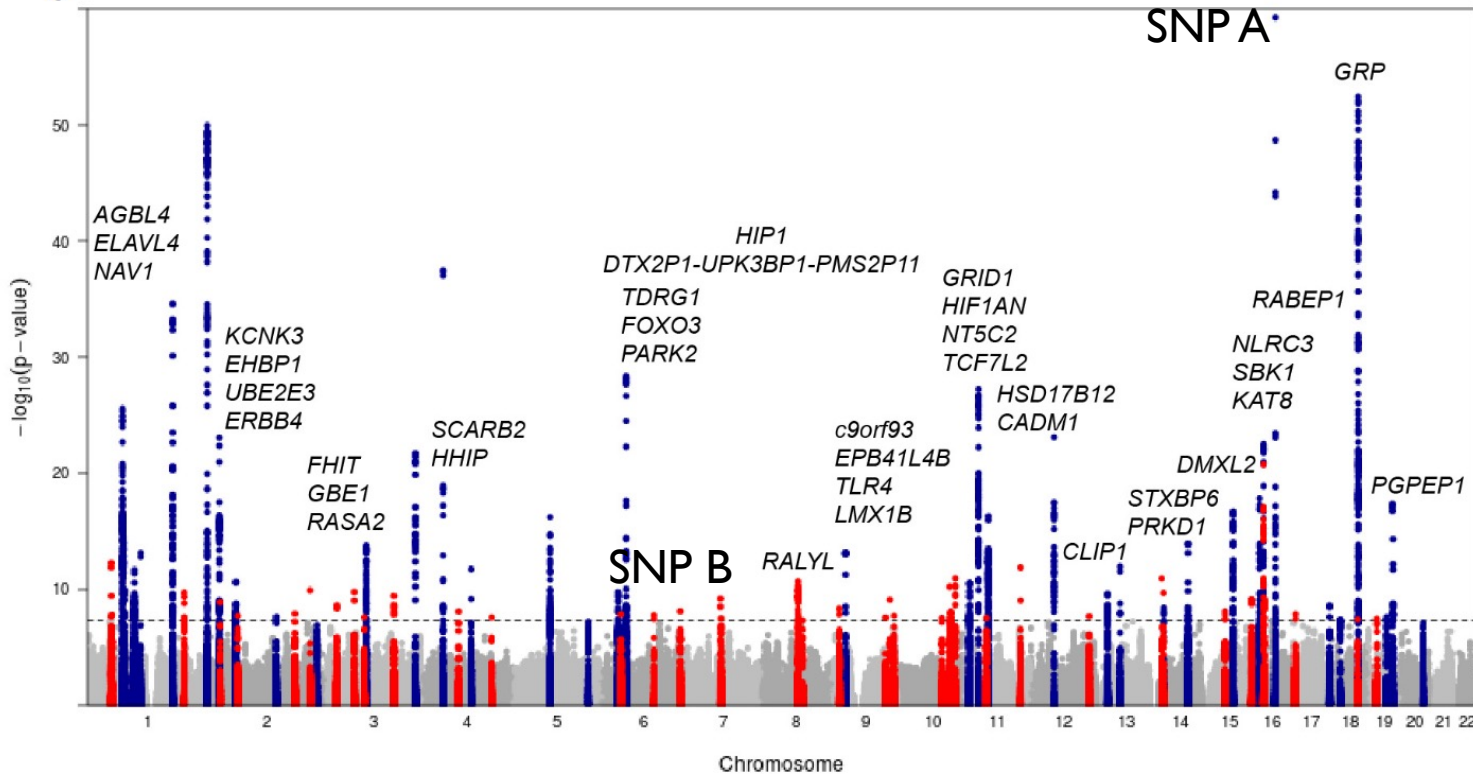


# GWAS 3

Matti Pirinen  
University of Helsinki  
March 20, 2023

# WHICH VARIANTS BECOME “SIGNIFICANT”?

**b** Locke et al. 2015 Nature

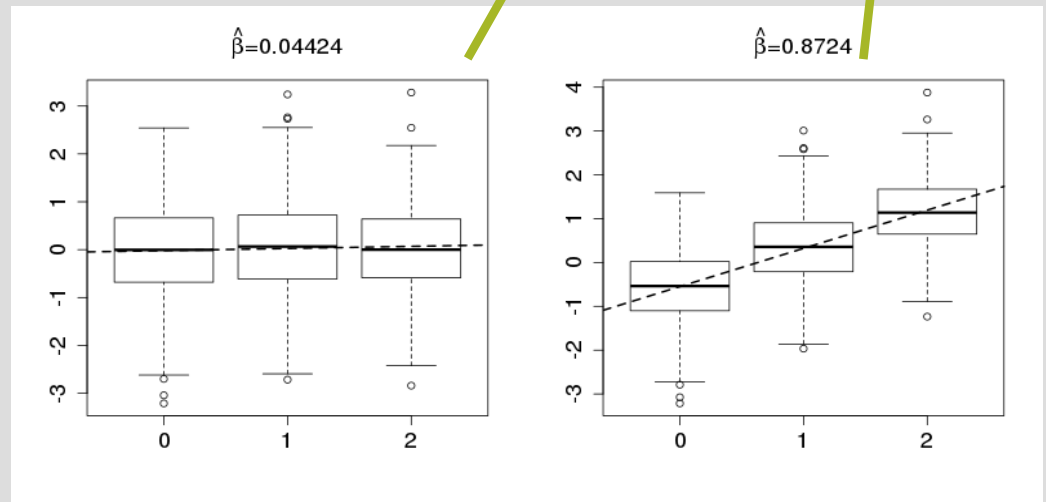
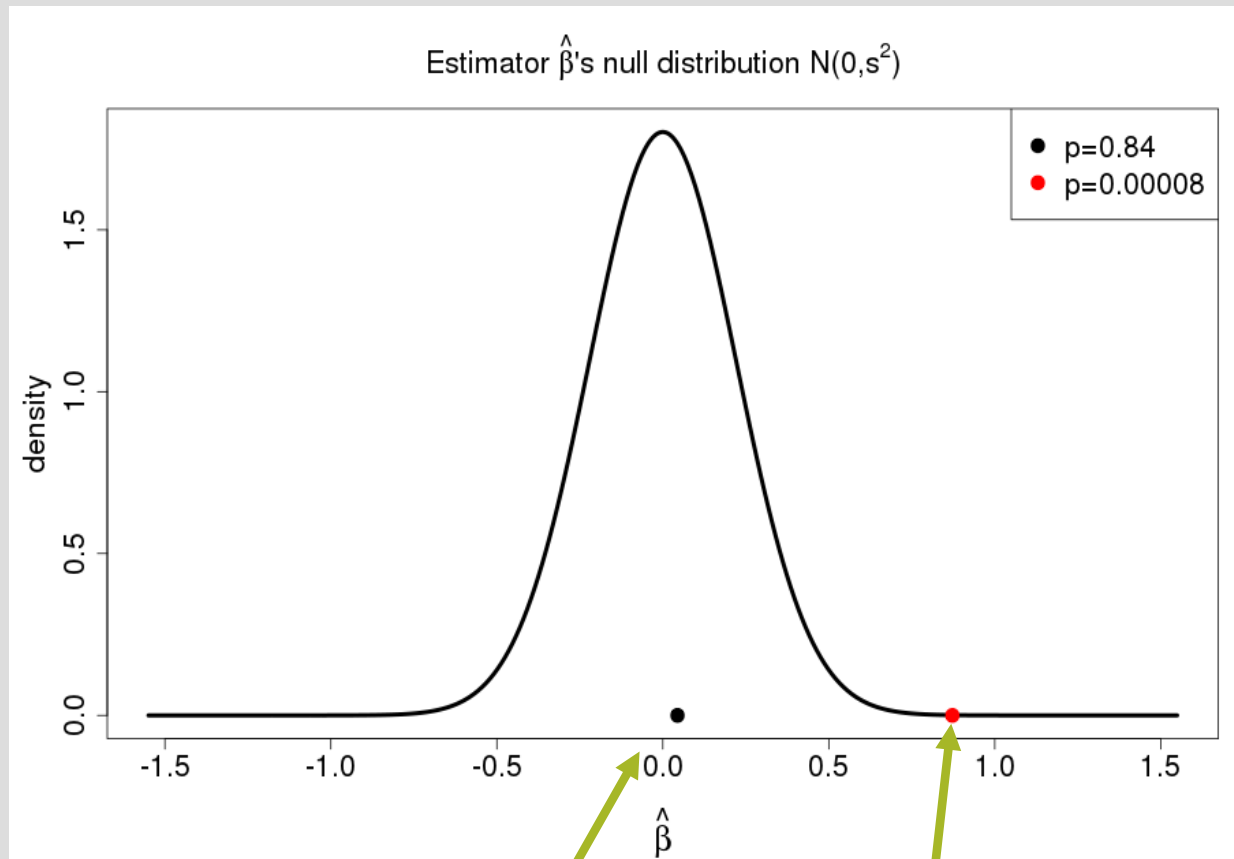


- With stringent threshold of  $P < 5e-8$  we have only few false positives
- What about true positives?
- Do we always find SNP A with  $P < 5e-8$  when we do a GWAS on BMI?
- What about SNP B?
- Which properties affect whether a true causal SNP becomes GWS?
- What is the probability that SNP A / B becomes GWS?
  - This is called “statistical power” to detect the SNP as associated

GWS = “genome-wide significant”, typically  $P < 5e-8$

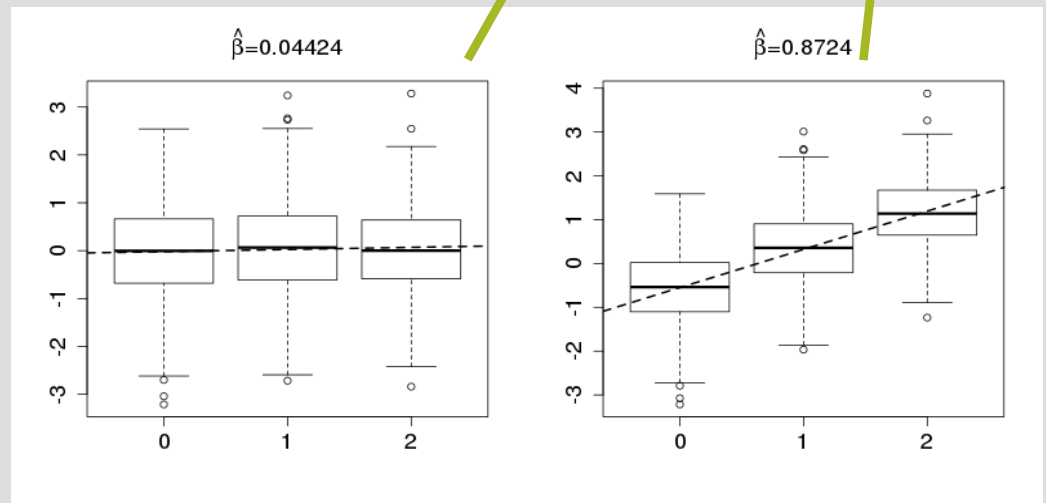
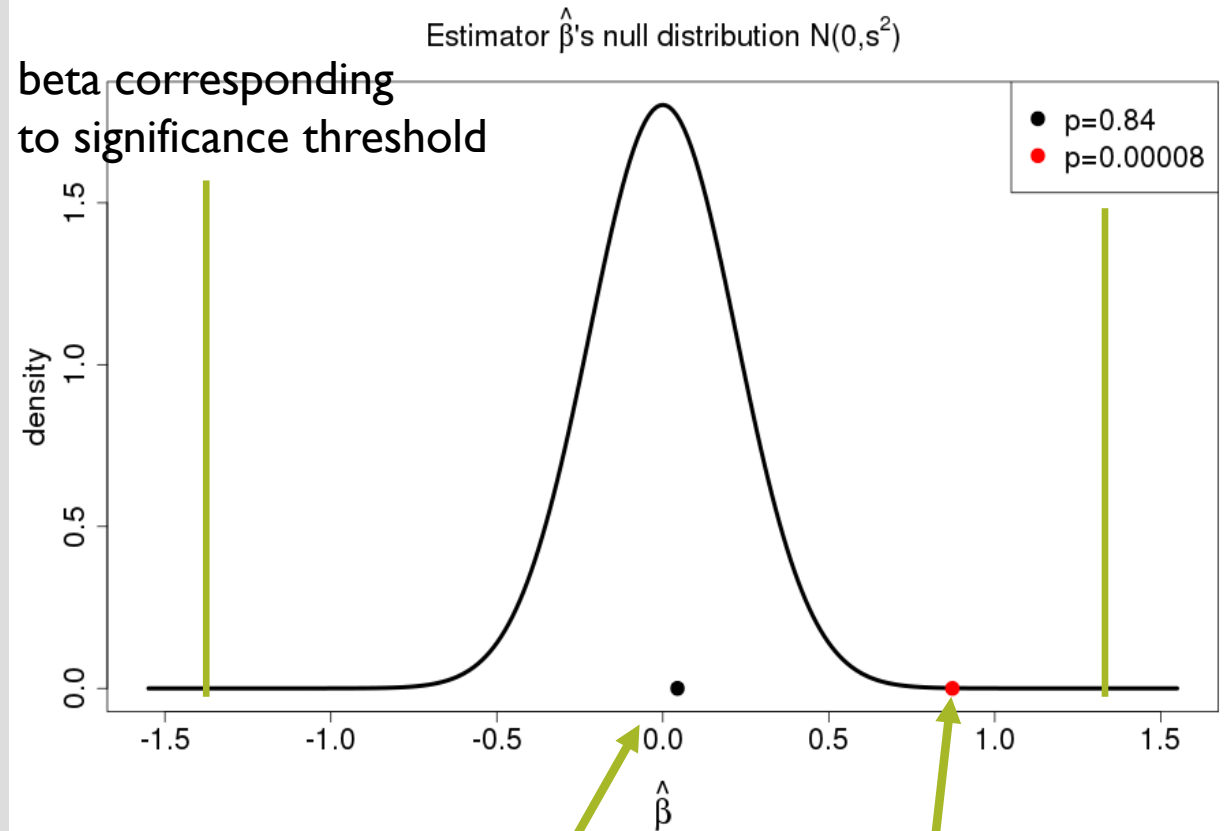
# REMINDER: P-VALUE

- Is the observed slope plausible if true slope = 0 ?
- *P*-value: Probability that we get at least as extreme estimate as we have observed, if true slope = 0
- $P = 0.84$ : No evidence for deviation from null
- $P = 8e-5$ : Unlikely under the null  $\rightarrow$  maybe not null

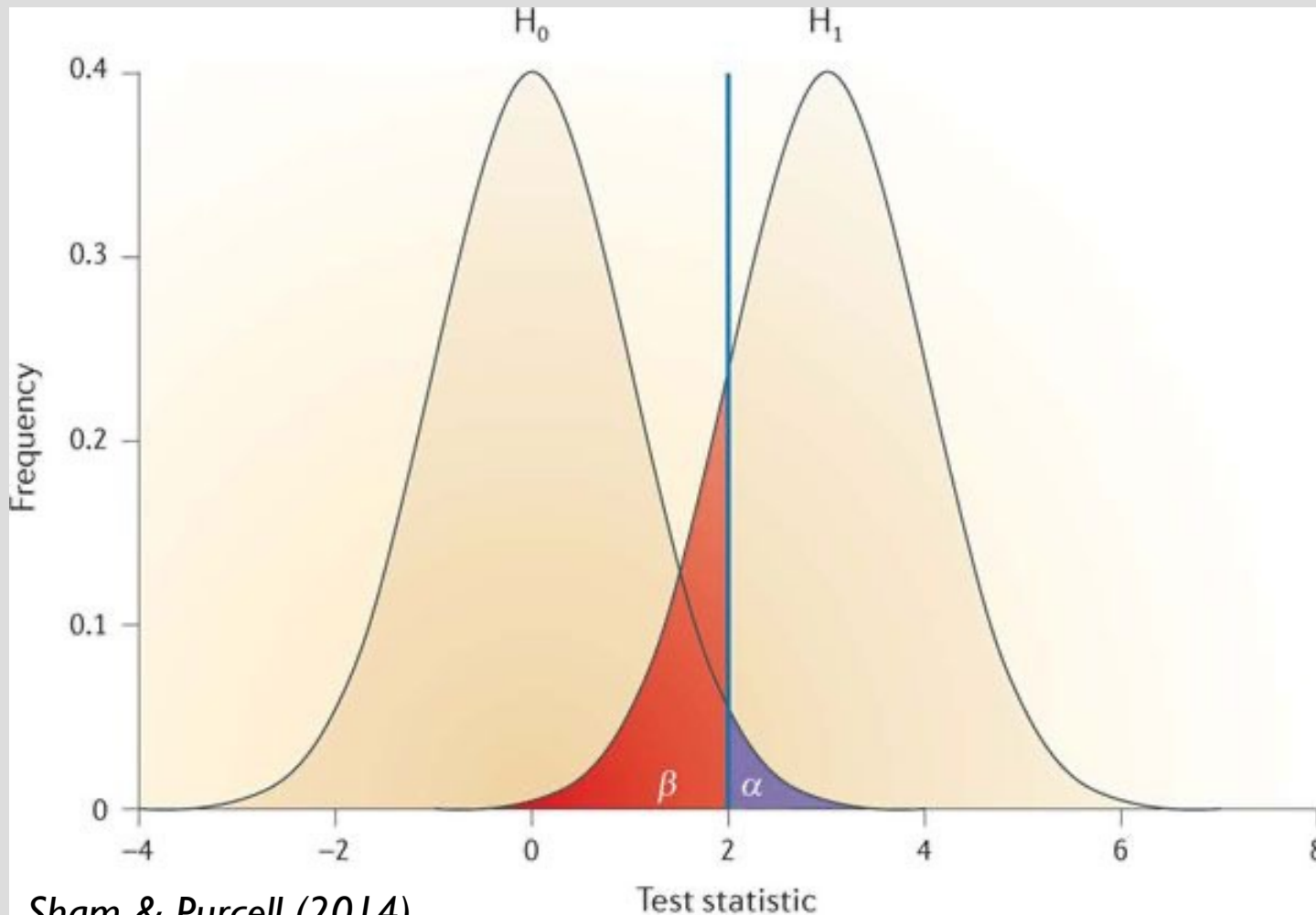


# SIGNIFICANCE THRESHOLD & POWER

- Significance threshold  $\alpha$  = Probability that a null variant has  $P$ -value  $\leq \alpha$
- What is the probability that a non-null variant has  $P$ -value  $\leq \alpha$ ?
- Depends on the properties of the variant and study
- Is called statistical power of the significance test



# TYPE I AND TYPE II ERRORS AND POWER



Sham & Purcell (2014)

Nature Reviews Genetics **15**: 335–346.

Nature Reviews | Genetics

The probability distributions of test statistic under  $H_0$  and  $H_1$ , the critical threshold for significance (blue line), the probability of type I error ( $\alpha$ ; purple) and the probability of type 2 error ( $\beta$ ; red).

Type I error: "false positive", wrongly reject  $H_0$  when  $H_0$  holds. Making significance level very low **avoids** Type I errors.

We can lower  $\alpha$  by dragging blue line to right.

Type II error: "false negative", wrongly accept  $H_0$  when  $H_0$  is not true.

Making significance level very low **creates** Type II errors.

Power =  $1 - \beta = P(\text{reject } H_0 \mid H_1 \text{ true})$ .

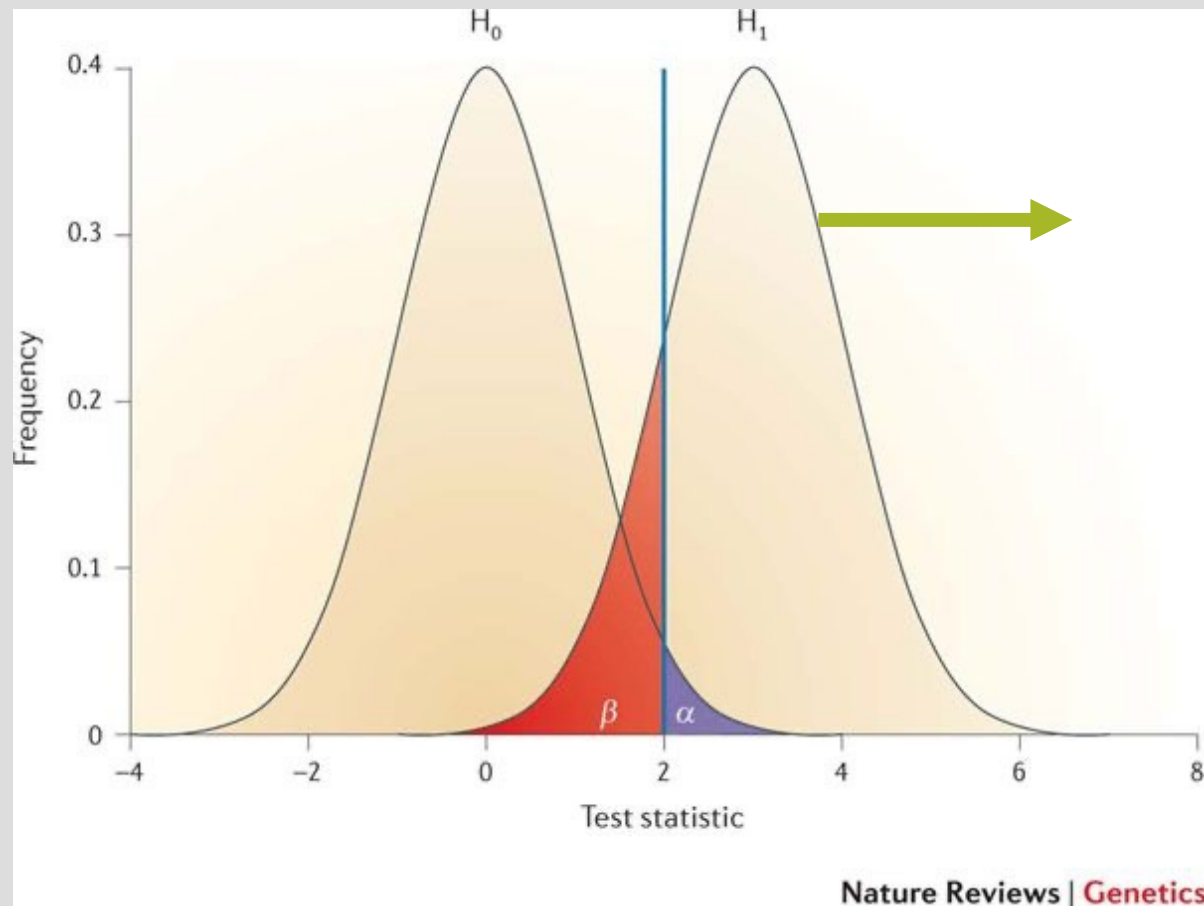
# HYPOTHESIS TESTING TERMINOLOGY

- $H_0$  (NULL HYPOTHESIS): Variant has no effect on phenotype
- $H_1$  (ALTERNATIVE HYPOTHESIS): Variant has a non-zero effect on the phenotype
  - In power calculations we must be specific about  $H_1$  (What are MAF,  $N$ , effect size?)
- Significance level  $\alpha$ : “Reject  $H_0$ ” and “accept  $H_1$ ” if  $P$ -value  $< \alpha$ , where  $P$ -value is calculated under  $H_0$
- If  $\alpha$  is defined before the experiment, then the proportion of false rejections of  $H_0$  out of all true  $H_0$  would be  $\alpha$  in repeated experiments
  - By making  $\alpha$  small (say  $5e-8$ ) we can protect from false positive findings (Type I errors) but increase false negative findings (Type II errors)
  - By keeping  $\alpha$  larger (say 0.05) we have more statistical power to reject  $H_0$  (avoid Type II error) but we are more likely to make a false positive finding (Type I error)

# WALD TEST

- Assuming that the GWAS model is correct (i.e., there are no biases), the regression coefficient estimator  $\hat{\beta} \sim N(\beta, SE^2)$
- Wald statistic  $z = \hat{\beta} / SE \sim N\left(\frac{\beta}{SE}, 1\right)$ 
  - $z \sim N(0, 1)$  under the null ( $\beta = 0$ ), and this is how we compute P-values
  - Under the alternative hypothesis, the mean of the distribution of  $z$  depends on true  $\beta$  and  $SE$
- Chi-square statistic  $z^2 \sim \chi_1^2\left(\text{NCP} = \beta^2 / SE^2\right)$ , where NCP is the “non-centrality parameter”
  - General definition: When  $Y \sim N(\mu, \sigma^2)$  then  $\frac{Y^2}{\sigma^2} \sim \chi_1^2\left(\text{NCP} = \mu^2 / \sigma^2\right)$
  - $z^2 \sim \chi_1^2$  under the null, i.e., the central (NCP = 0) chi-square distribution with 1 df

$$Z = \hat{\beta} / SE \sim N\left(\frac{\beta}{SE}, 1\right)$$



- The alternative's test statistic distribution will move farther from the null distribution when  $|\beta|/SE$  grows
- For a fixed significance threshold, the power will thus increase as  $|\beta|$  increases or as SE decreases
- Makes sense:
  - “Larger effects are easier to find”
  - “More precise estimates help separating real effects from noise”



## FORMULAS FOR SE

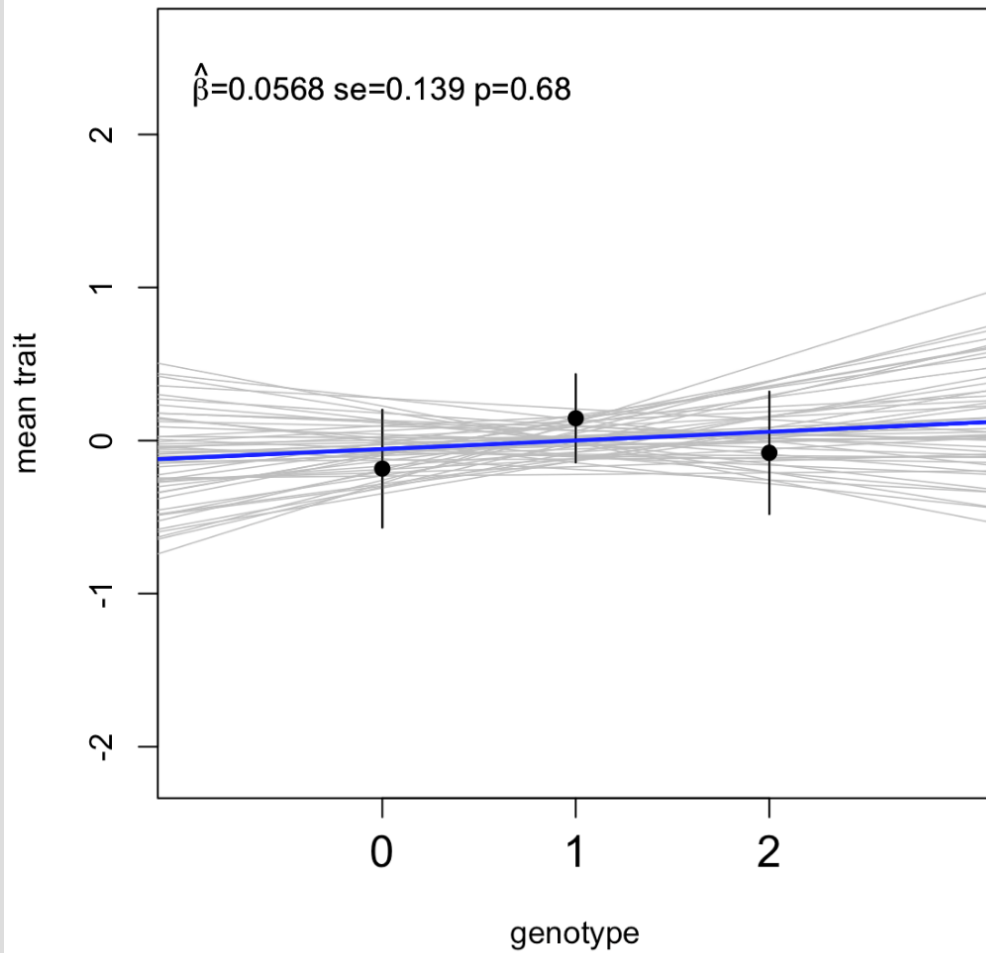
- Liner model GWAS has  $SE \approx \frac{\sigma}{\sqrt{2 n f (1-f)}}$
- Logistic model GWAS has  $SE \approx \frac{1}{\sqrt{2 n \phi (1-\phi) f (1-f)}}$
- $\sigma$  is the error variance
- $n$  is the total sample size
- $f$  is the minor allele frequency
- $\phi$  is the proportion of cases among all samples

## FORMULAS FOR $NCP = \beta^2 / SE^2$

- Linear model GWAS has  $NCP \approx 2 n f (1 - f) \beta^2 / \sigma^2$
- Logistic model GWAS has  $NCP \approx 2 n \phi (1 - \phi) f (1 - f) \beta^2$
- $\sigma$  is the error variance
- $n$  is the total sample size
- $f$  is the minor allele frequency
- $\beta$  is the effect size
- $\phi$  is the proportion of cases among all samples

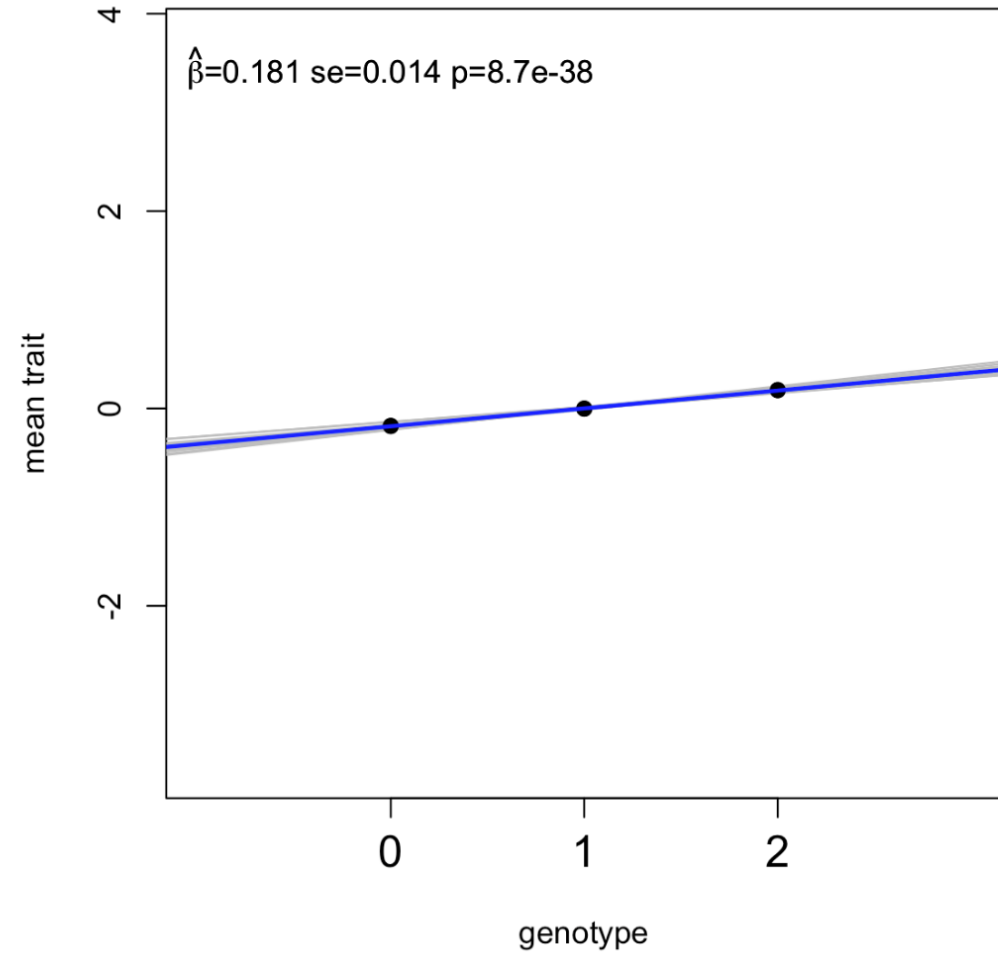
# WHY N INCREASES POWER?

**n=100 afreq=0.5 b=0.2**



We are unsure whether slope is positive

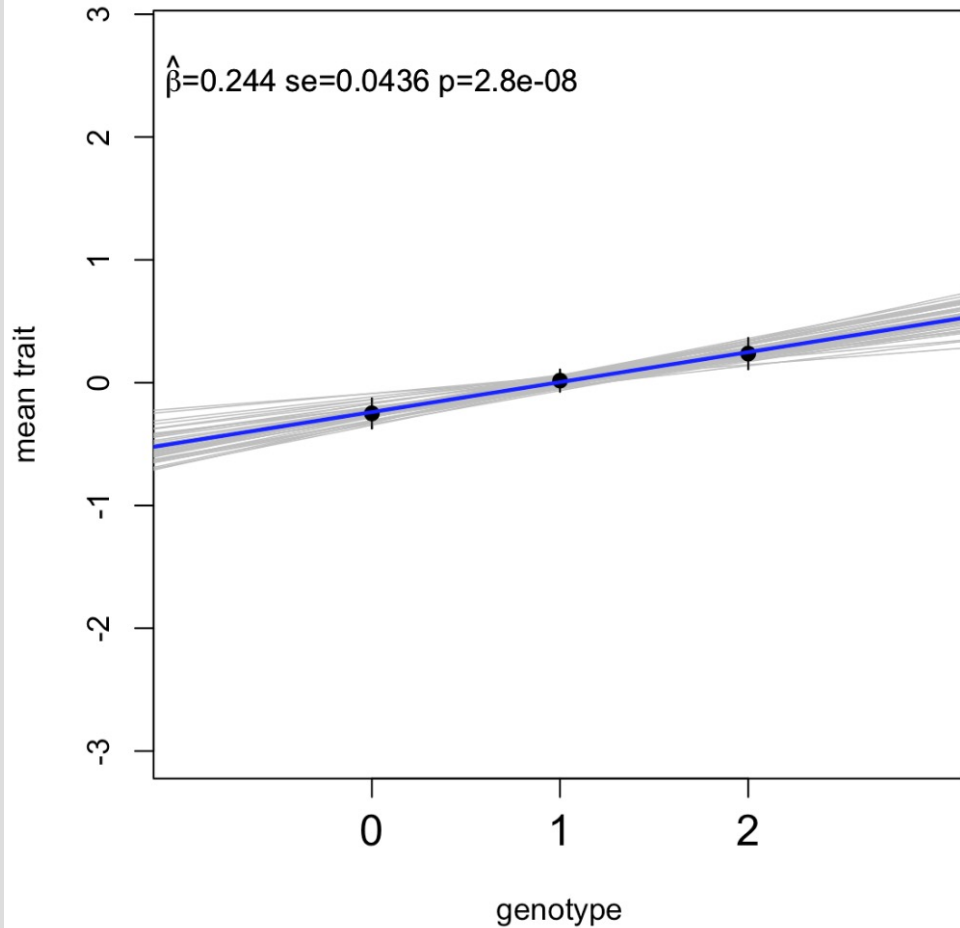
**n=10000 afreq=0.5 b=0.2**



We are confident that the slope is positive

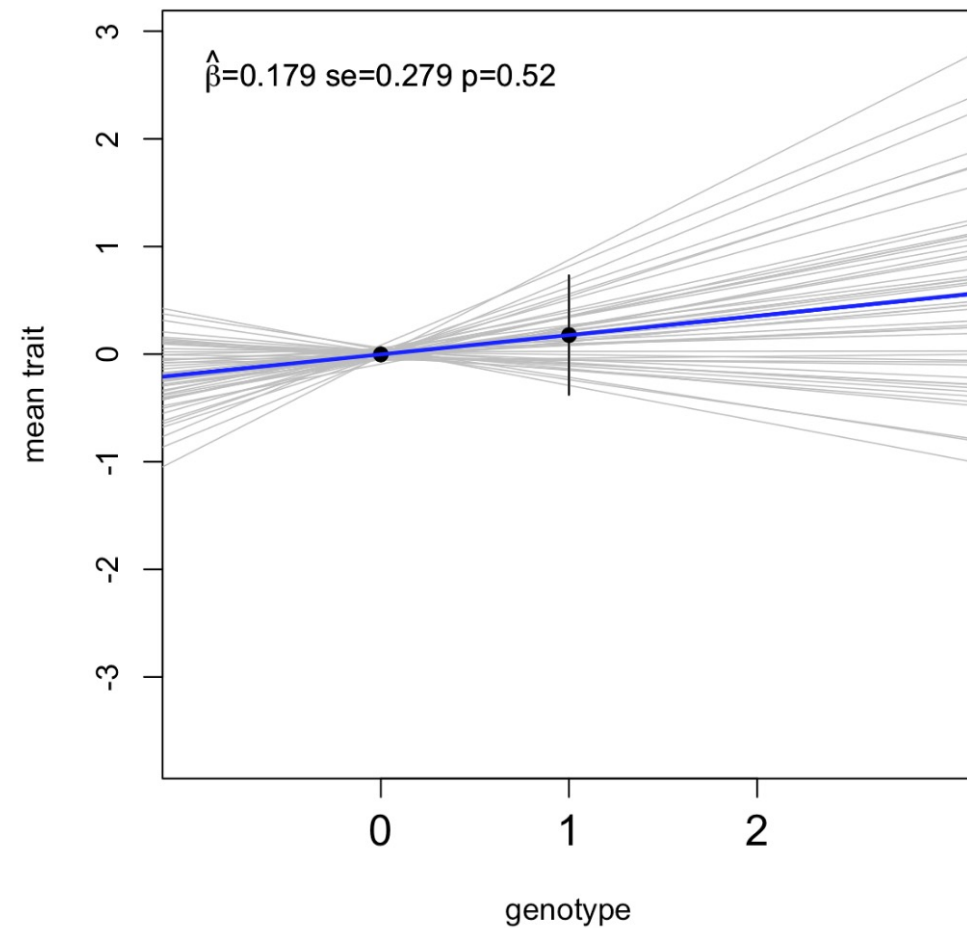
# WHY MAF INCREASES POWER?

n=1000 afreq=0.5 b=0.2



We are confident that the slope is positive

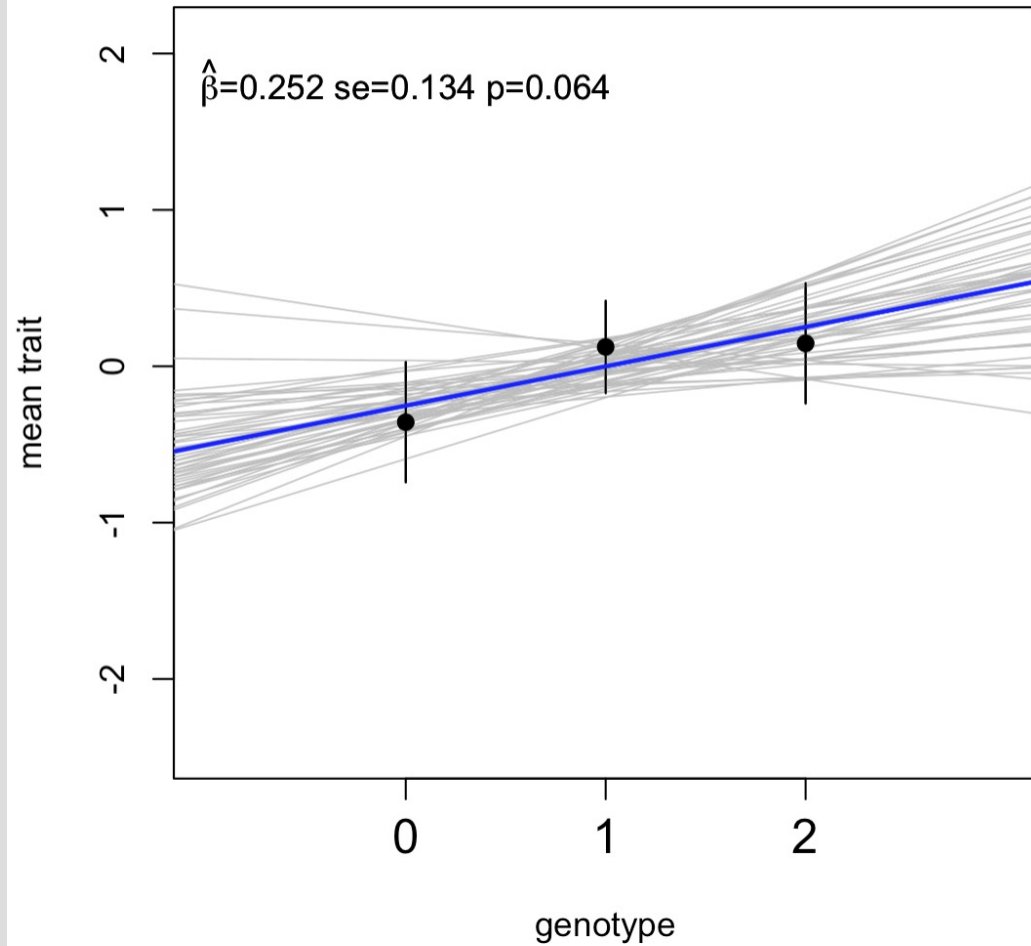
n=1000 afreq=0.01 b=0.2



We are unsure whether the slope is positive

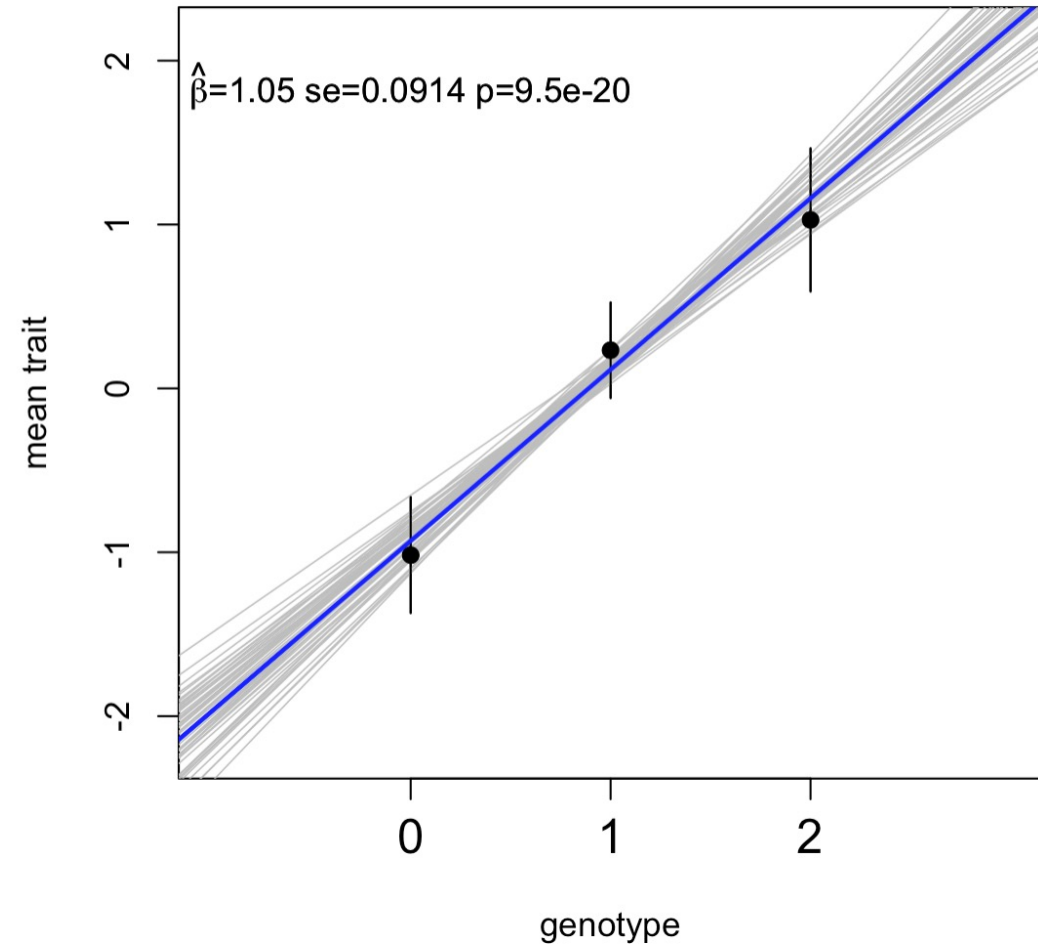
# WHY $|\beta|$ INCREASES POWER?

**n=100 afreq=0.5 b=0.2**



We are unsure whether the slope is positive

**n=100 afreq=0.5 b=1**

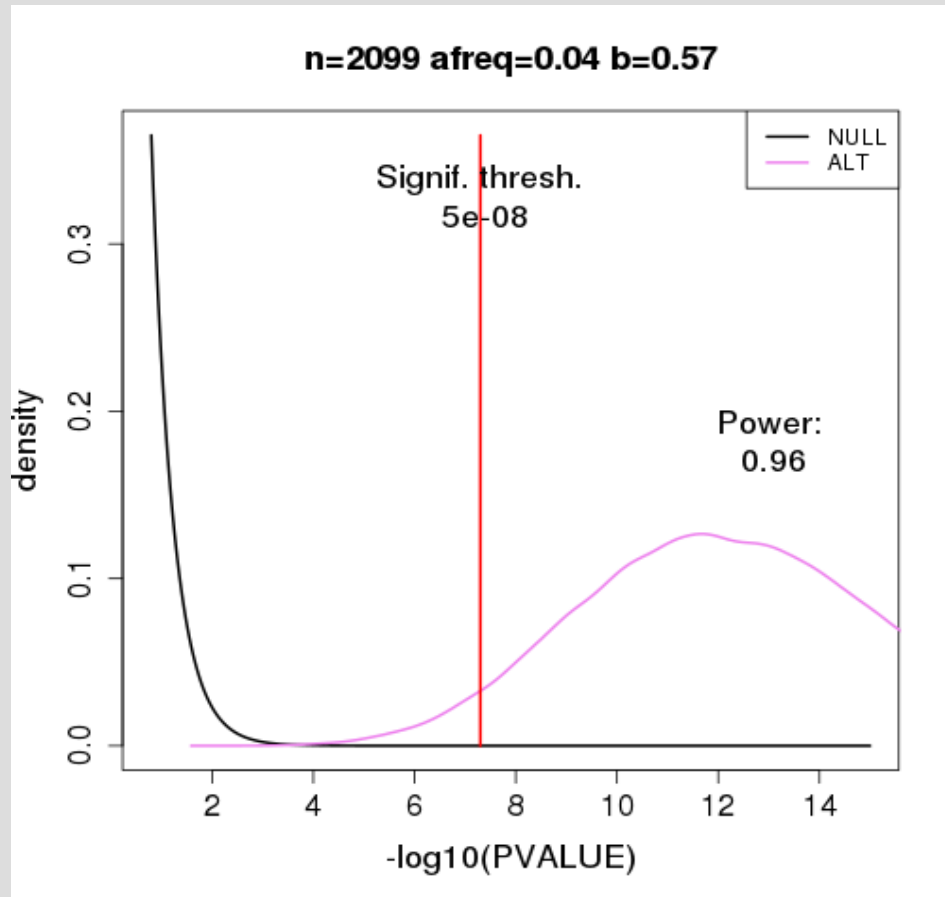


We are confident that the slope is positive

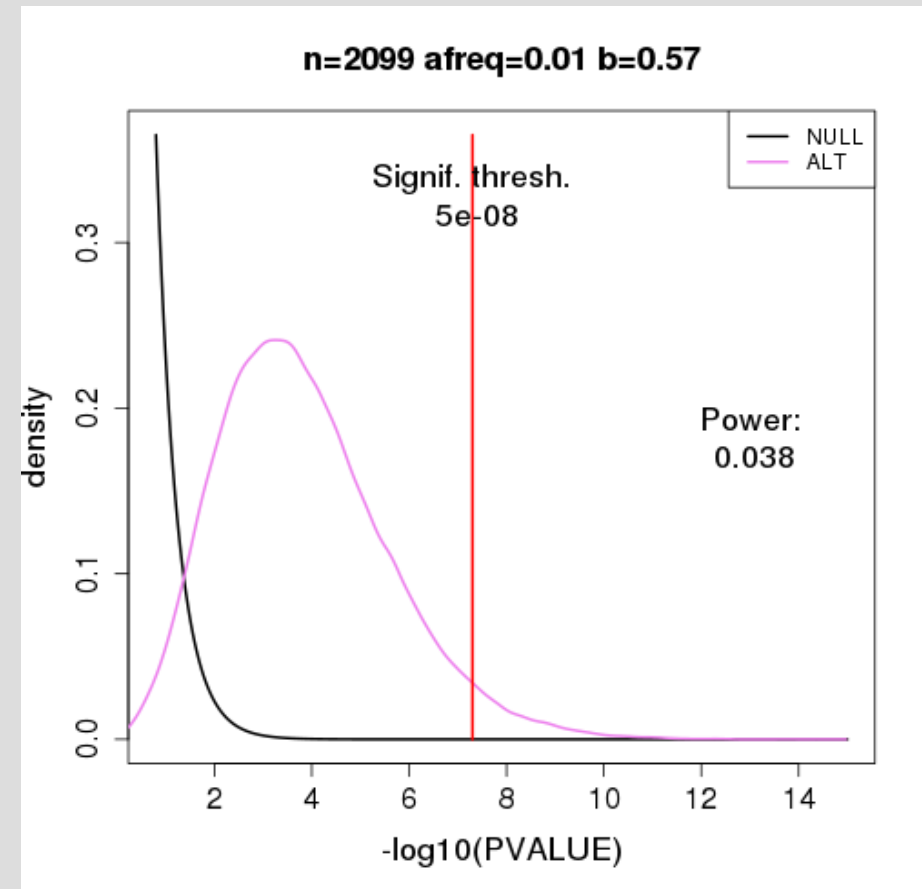
## WHY $\phi(1 - \phi)$ INCREASES POWER?

- If we have a lot of controls, we know the control frequencies very accurately
- But if we have only few cases, then we don't know the case frequencies accurately
- We cannot tell whether cases are different from controls unless we know accurately **BOTH** the case and the control frequencies
- Extreme setting: all samples are controls -- we learn nothing
- $N \phi (1 - \phi)$  is the **effective sample size** of a case-control study
  - To make it large, we should have large  $N$  and  $\phi$  close to 0.5
  - Often effective sample size is defined as  $4 N \phi (1 - \phi)$ , which is the total sample size of a study where counts of cases equals counts of controls and power is the same as the power observed in our study

# PCSK9 VARIANT FROM MOTIVATION VIDEO



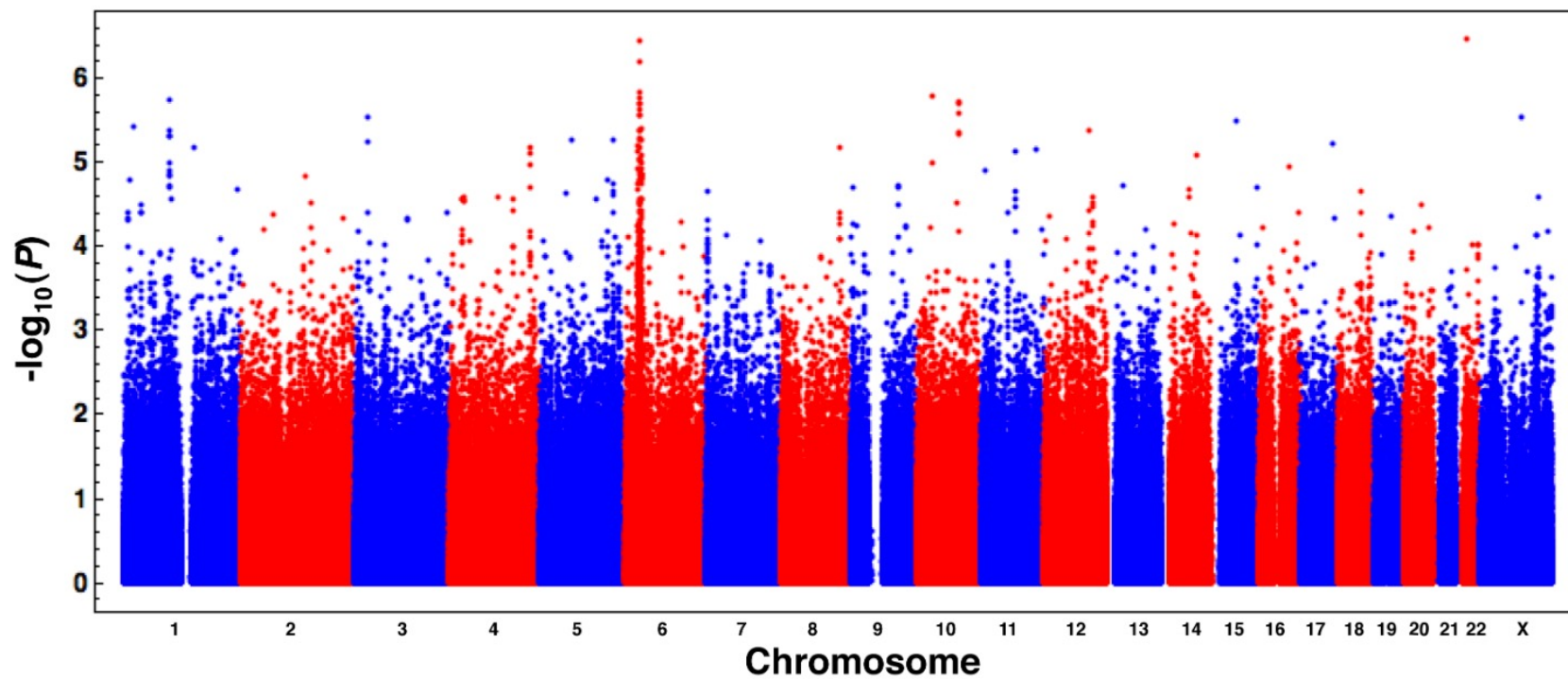
In Finland MAF = 4%:  
We are almost certain to detect it with 2099 samples



In Central Europe MAF = 1%:  
We are almost certain to not detect it with 2099 sample

## SCHIZOPHRENIA GWAS I/3 2009

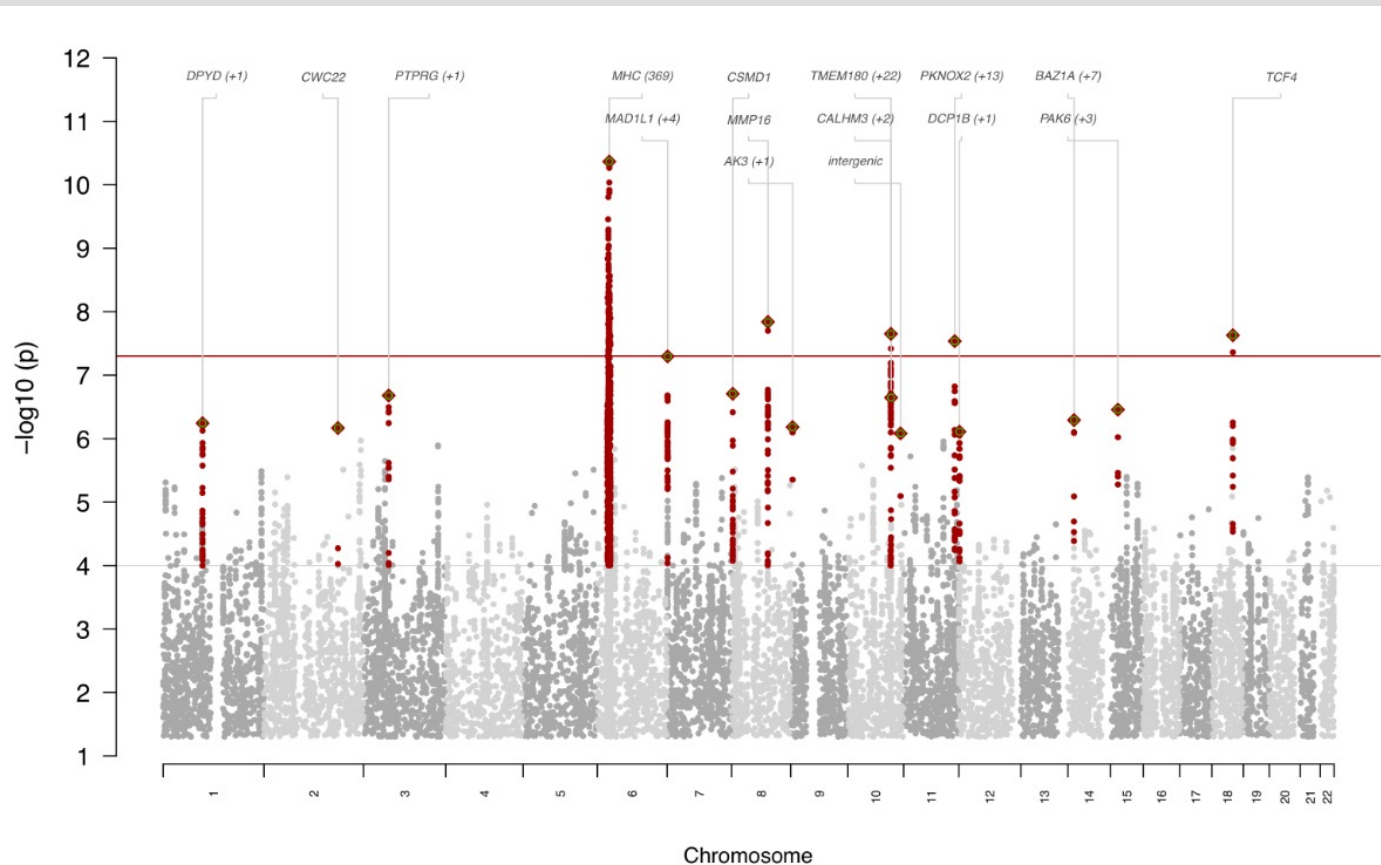
- 3,332 SZ cases and 3,587 controls at IM SNPs
- No genome-wide significant findings
- Suggestive evidence for HLA-region on chr 6





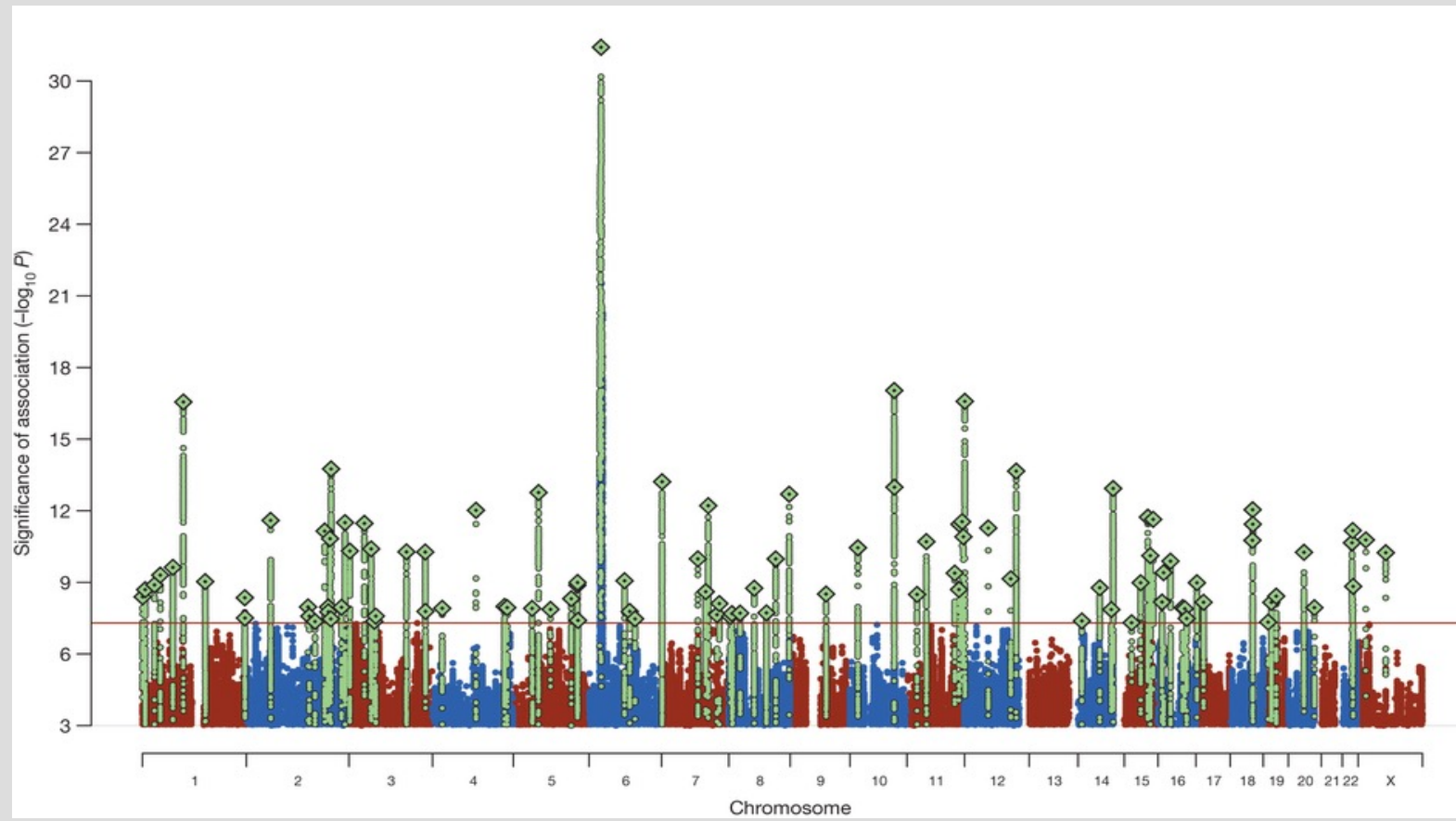
# SCHIZOPHRENIA GWAS 2/3 2011

- 9,394 SZ cases and 12,462 controls at IM SNPs
- 5 GWS loci



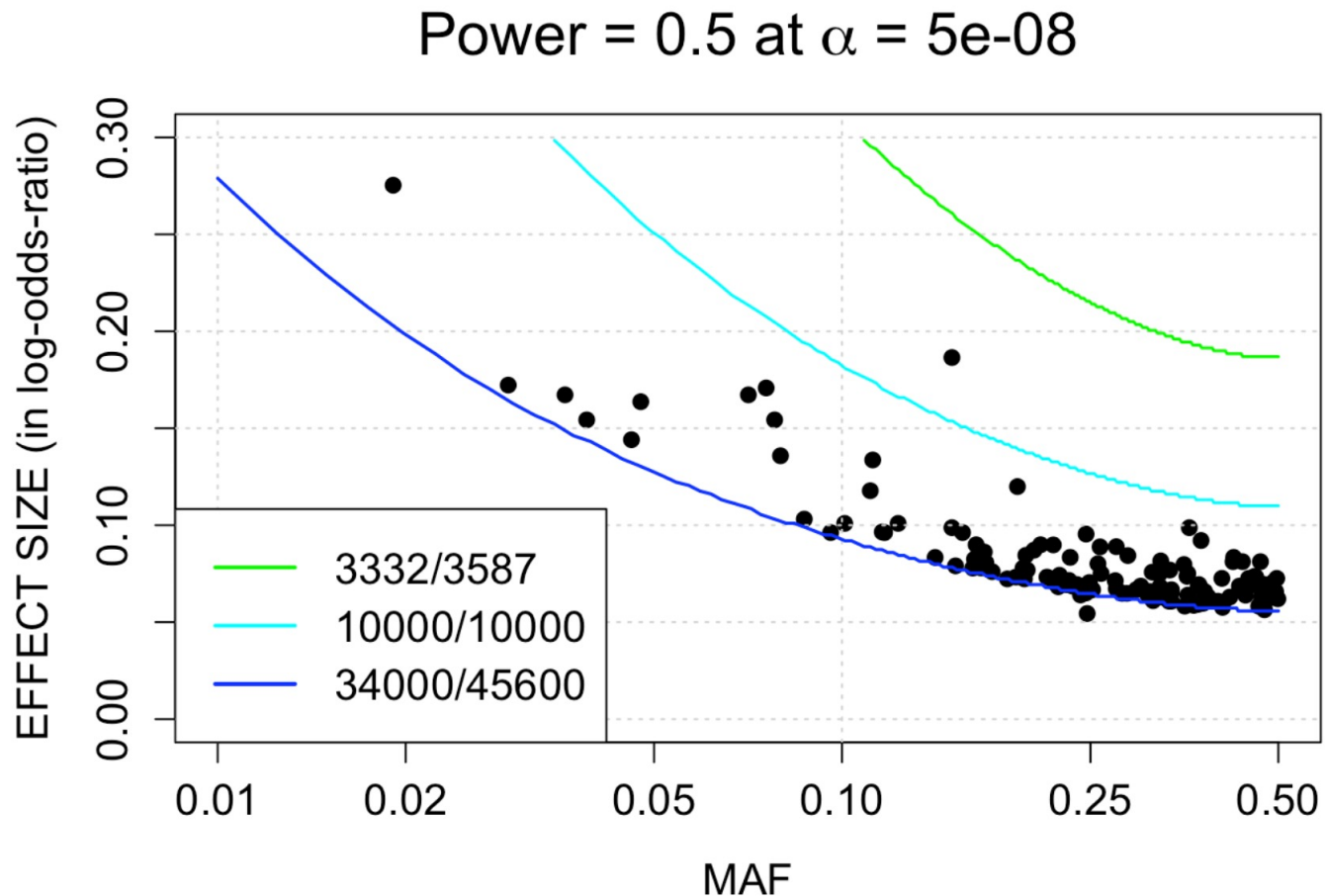
# SCHIZOPHRENIA GWAS 3/3 2014

- 34,000 SZ cases and 45,600 controls at 9.5M SNPs
- 108 loci



Psychiatric genomics  
consortium  
Nature 2014

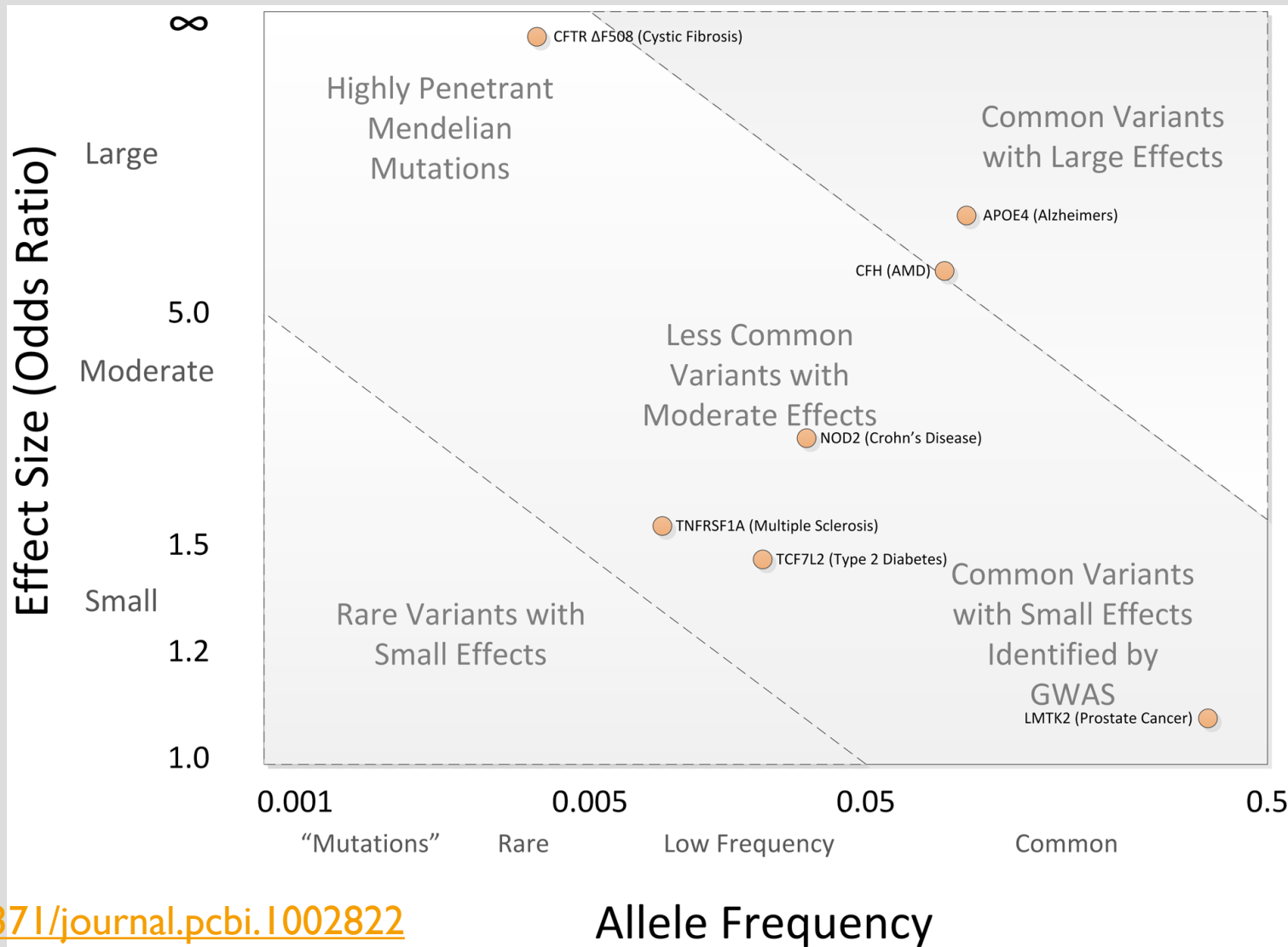
# ABSENCE OF EVIDENCE IS NOT EVIDENCE OF ABSENCE



- Non-significant P-value does NOT exclude the existence of non-zero effect, it only excludes the existence of so large effects for which the power would had been close to one.

Power curves for the 3 schizophrenia GWAS. The first 2 GWAS were underpowered for the effects that exist for SZ.

# EFFECT, MAF, AND REGION OF POWER



Disease associations are often conceptualized in two dimensions: allele frequency and effect size. Highly penetrant alleles for Mendelian disorders are extremely rare with large effect sizes (upper left), while most GWAS findings are associations of common SNPs with small effect sizes (lower right). The bulk of discovered genetic associations lie on the diagonal denoted by the dashed lines.