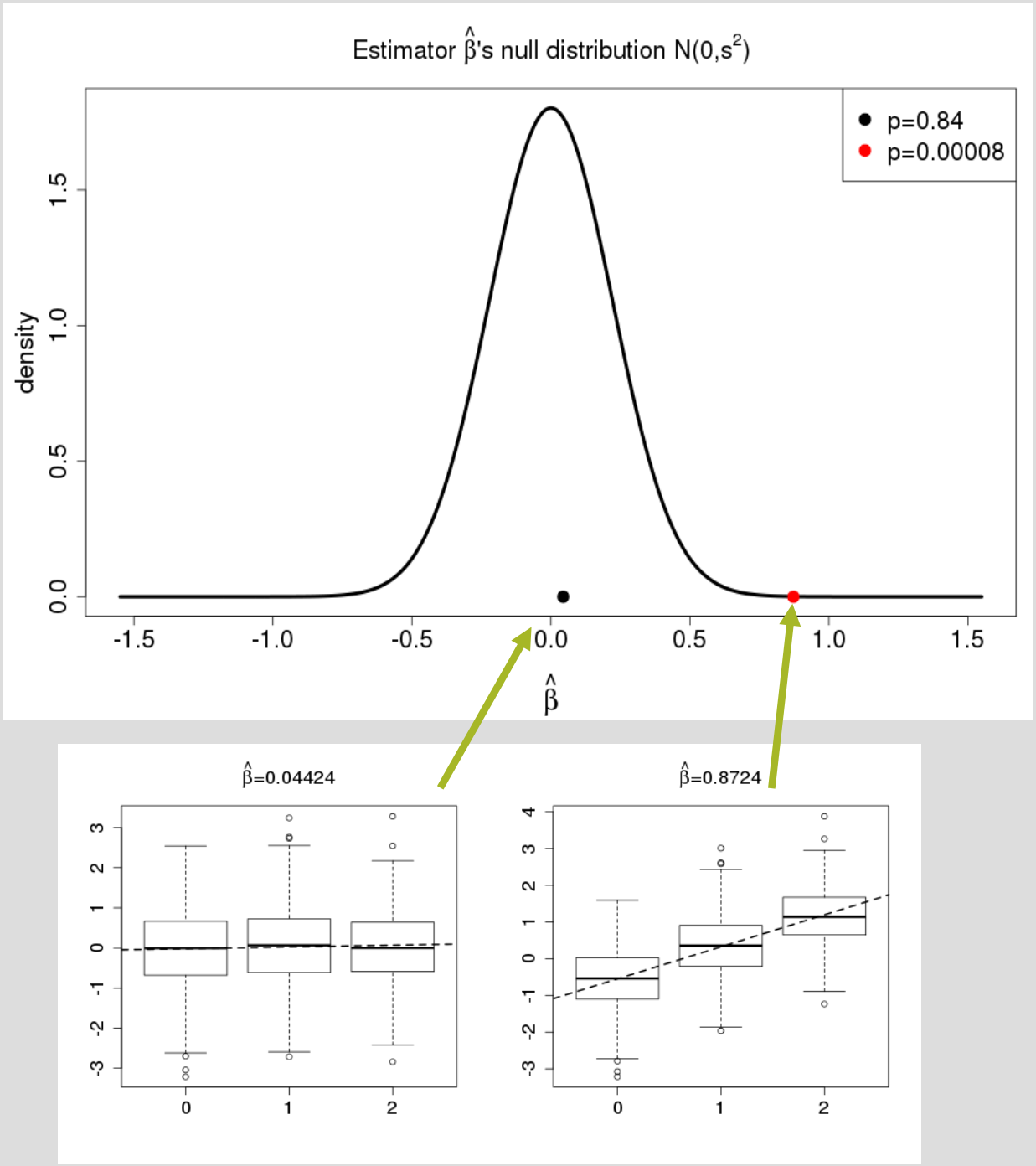
GWAS 3

Matti Pirinen
University of Helsinki
23.1.2019

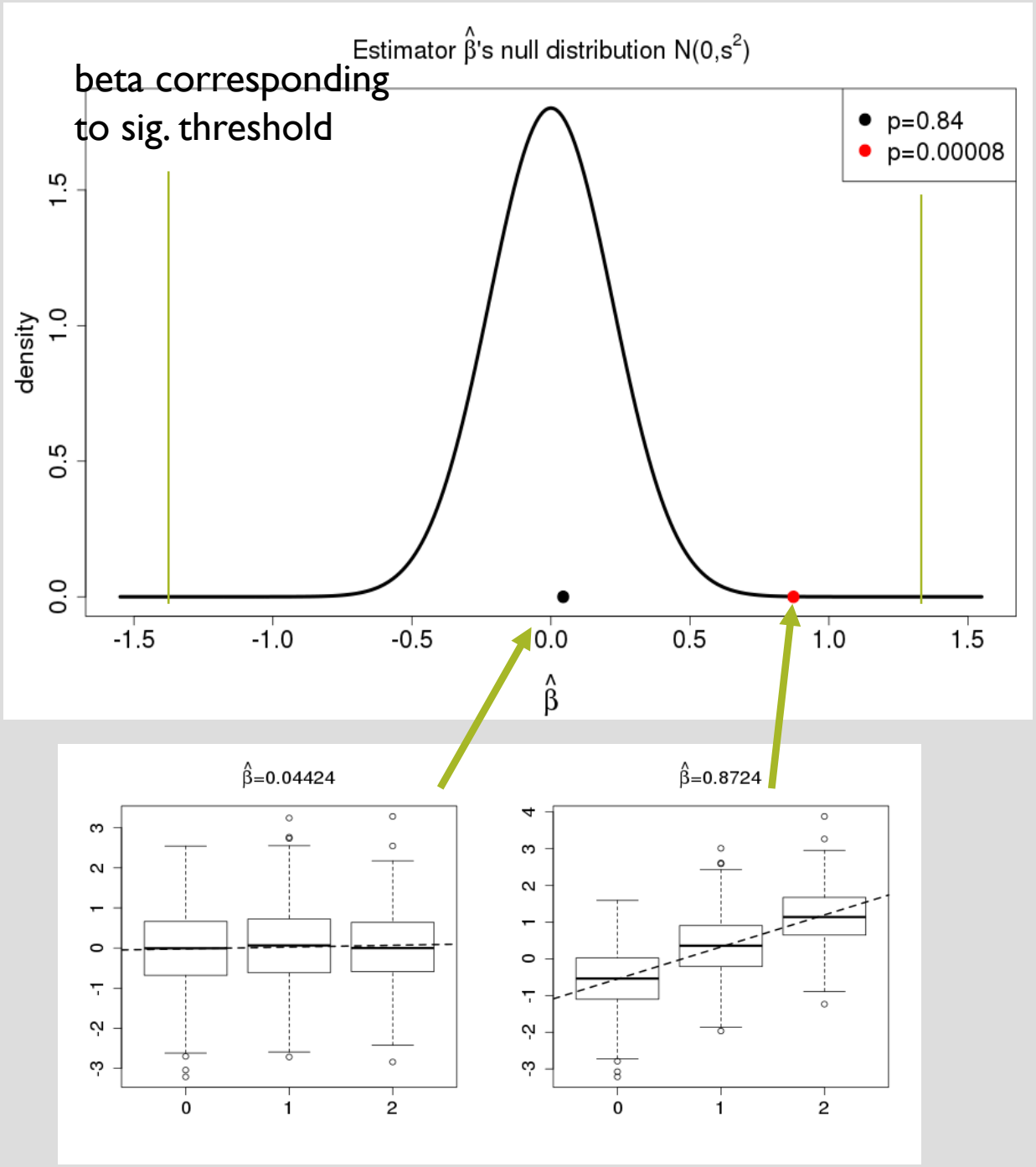
REMINDER: P-VALUE

- Is the observed slope plausible if true slope = 0 ?
- P-value: Probability that "by chance" we get as extreme value as we have observed
- P = 0.84: No evidence for deviation from null
- P = 8e-5: Unlikely under the null → maybe not null



SIG.THRESH & POWER

- Significance threshold α = Probability that a null variant will reach P-value below α
- What is probability that a non-zero variant will reach P-value below α?
  - Depends on the properties of variant and study
  - Is called statistical power of the significance test

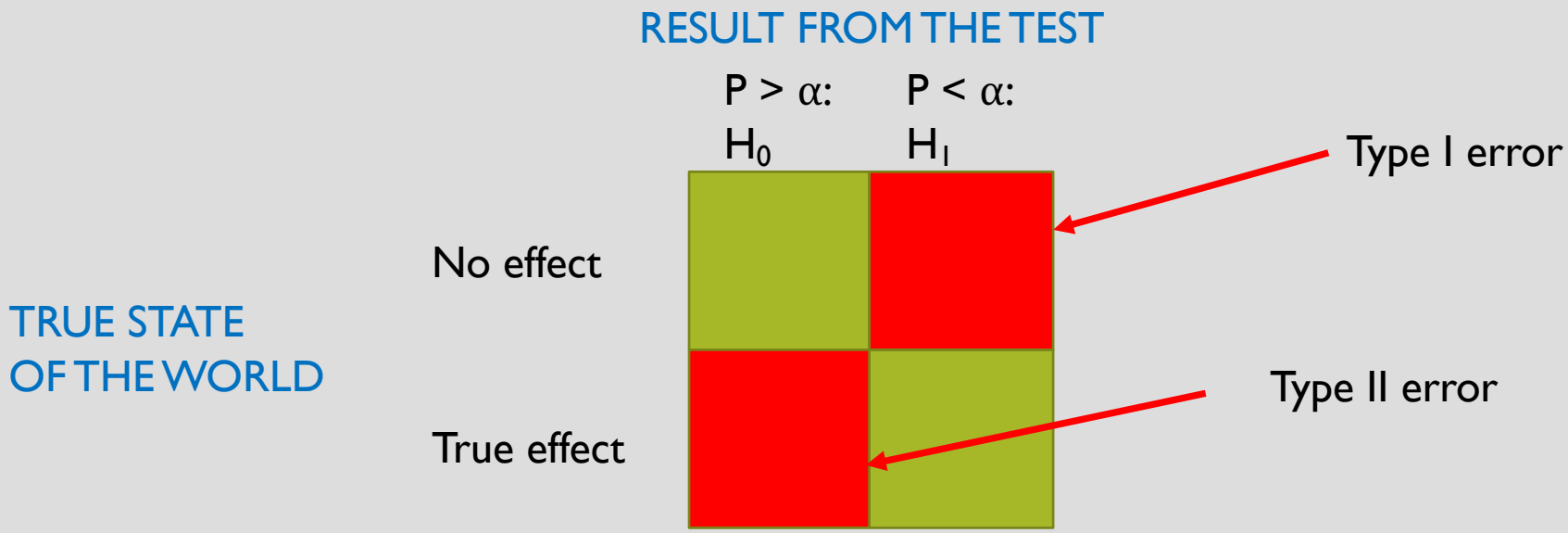


HYPOTHESIS TESTING

- H_0 (NULL HYPOTHESIS): Variant has no effect on phenotype
- H_1 (ALTERNATIVE HYPOTHESIS):Vriant has a non-zero effect on the phenotype
- Significance level α:"Reject H_0" and "accept H_1" if P-value (calculated assuming H_0) is $< \alpha$
  - If α is defined before the experiment, then the proportion of false rejection of H_0 would be α in repeated experiments
  - By making α small (say 1e-6) we can protect from false positive findings (Type I errors) but increase false negative findings (Type II errors)
  - By keeping α larger (say 0.05) we have more statistical power to reject H_0 but we are more likely to make a false positive finding (Type I error)

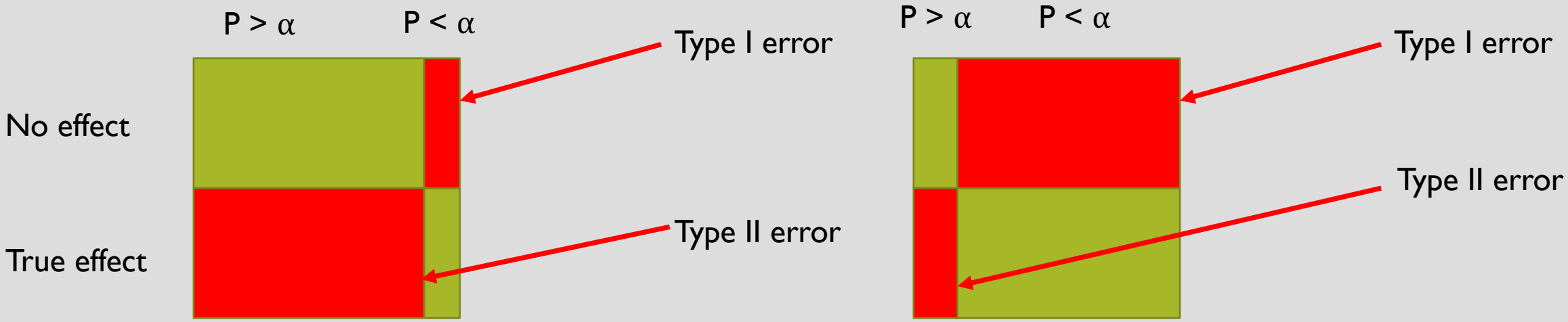
HYPOTHESIS TESTING FRAMEWORK

- Type I error: seeing an effect that is not real, a false positive
- Type II error: failing to see an effect that is real, a false negative
- Significance level α affects how likely different types of errors are to occur



HYPOTHESIS TESTING FRAMEWORK

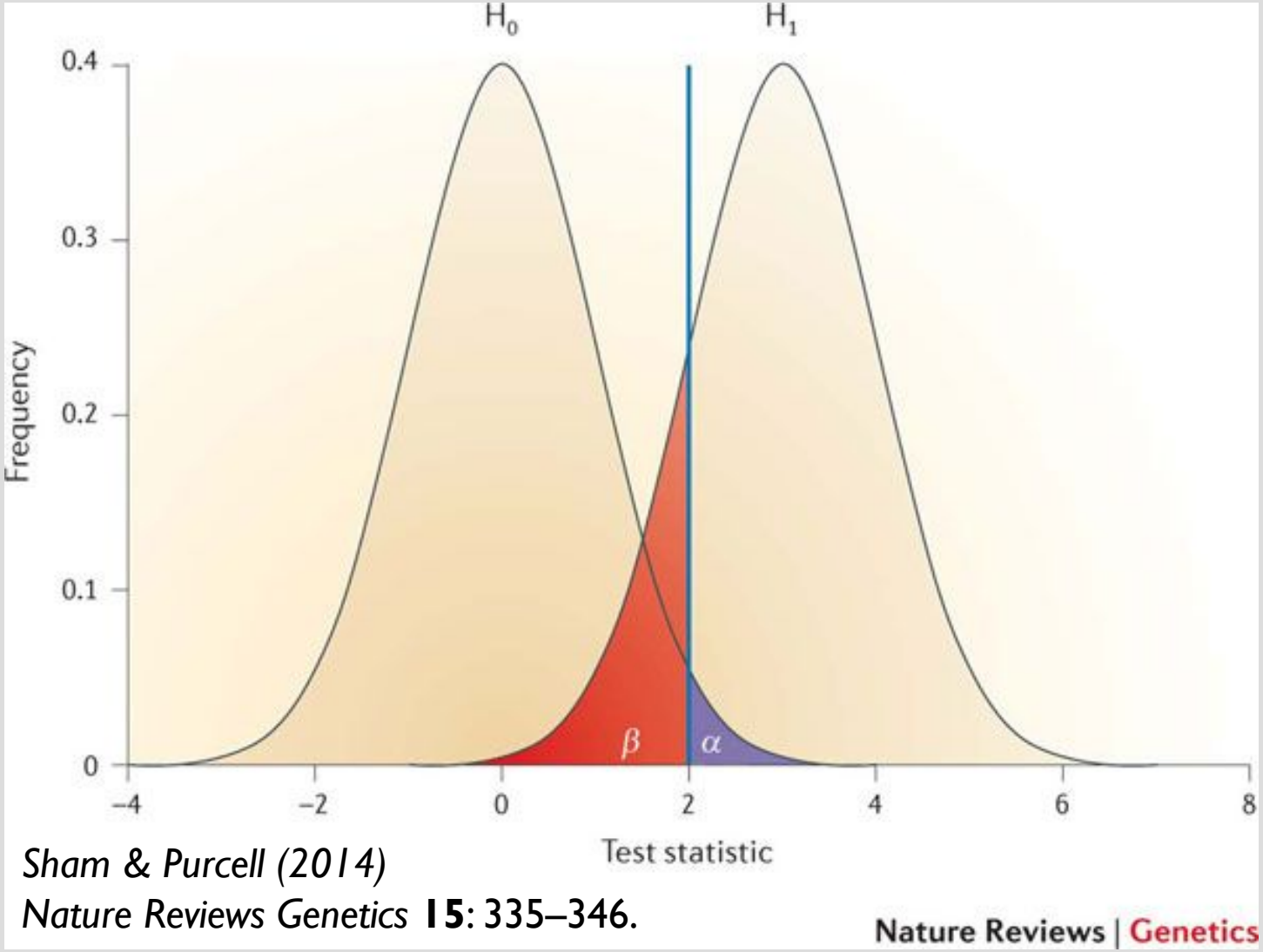
- Type I error: seeing an effect that is not real, a false positive
- Type II error: failing to see an effect that is real, a false negative
- With small significance level we rarely make Type I errors, but may often make Type II errors
- With large significance level we rarely make Type II errors, but may often make Type I errors



Small significance level (e.g. 5e-8 in GWAS)

Large significance level (e.g. 0.05)

TYPE I AND TYPE II ERRORS AND POWER



Sham & Purcell (2014)
Nature Reviews Genetics **15**: 335–346.

This schematic representation of the probability distributions of test statistic under H_0 and H_1 shows the critical threshold for significance (blue line), the probability of type 1 error (α; purple) and the probability of type 2 error (β; red).

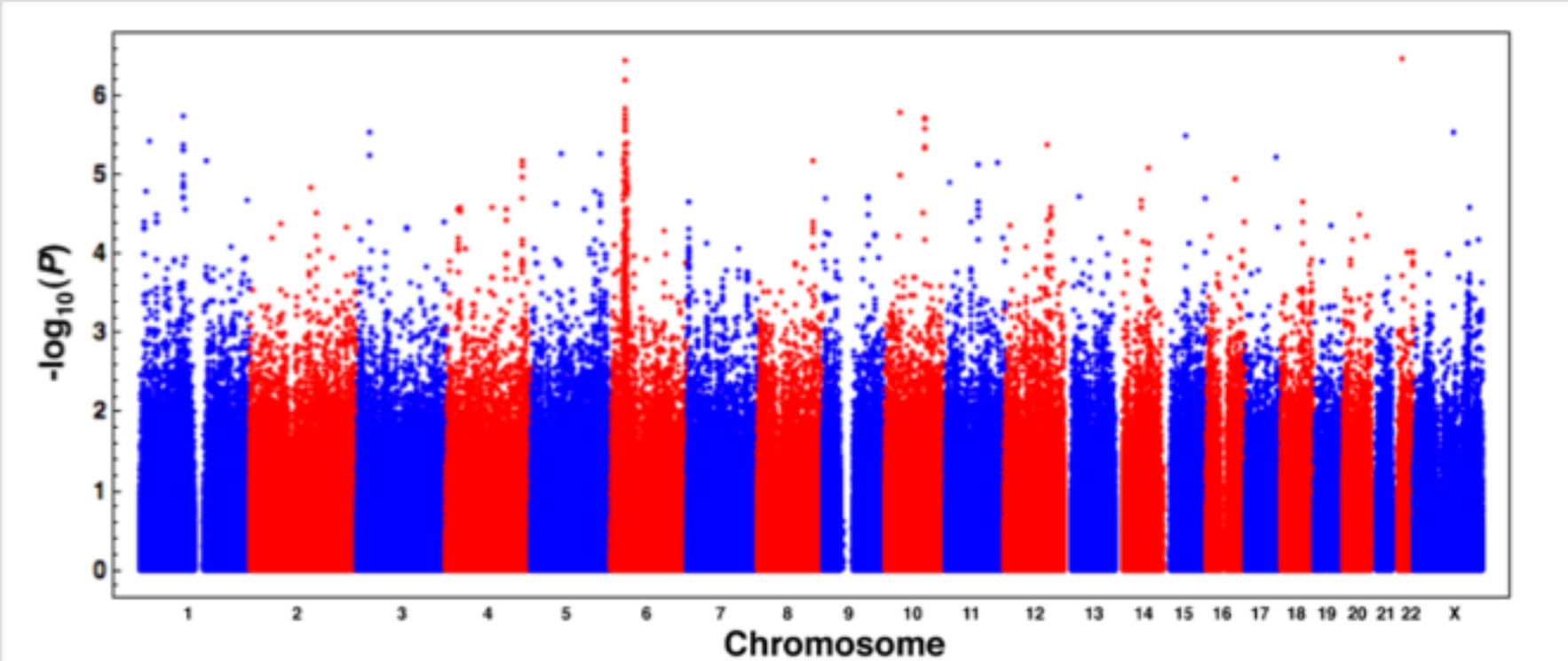
Type I error: "false positive", wrongly reject H_0 when H_0 would be true.
Making significance level very low **avoids** Type I errors.

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- β = P(reject H0 | H1 true).

SCHIZOPHRENIA GWAS 1/3 2009

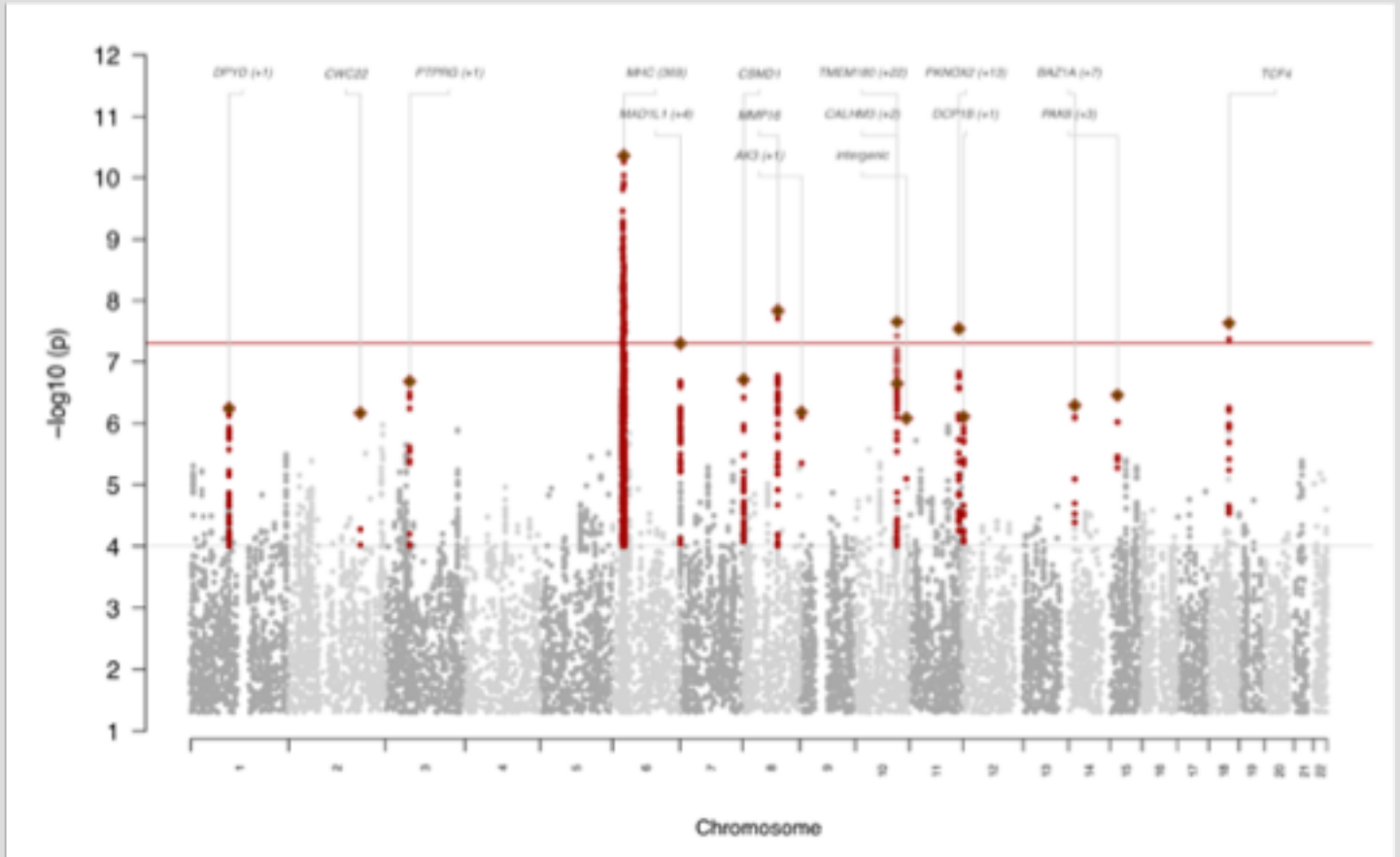
- 3,332 SZ cases and 3,587 controls at 1M SNPs
- No genome-wide significant findings
- Suggestive evidence for HLA-region on chr 6



Int'l SCZ consortium
Nature 2009

SCHIZOPHRENIA GWAS 2/3 2011

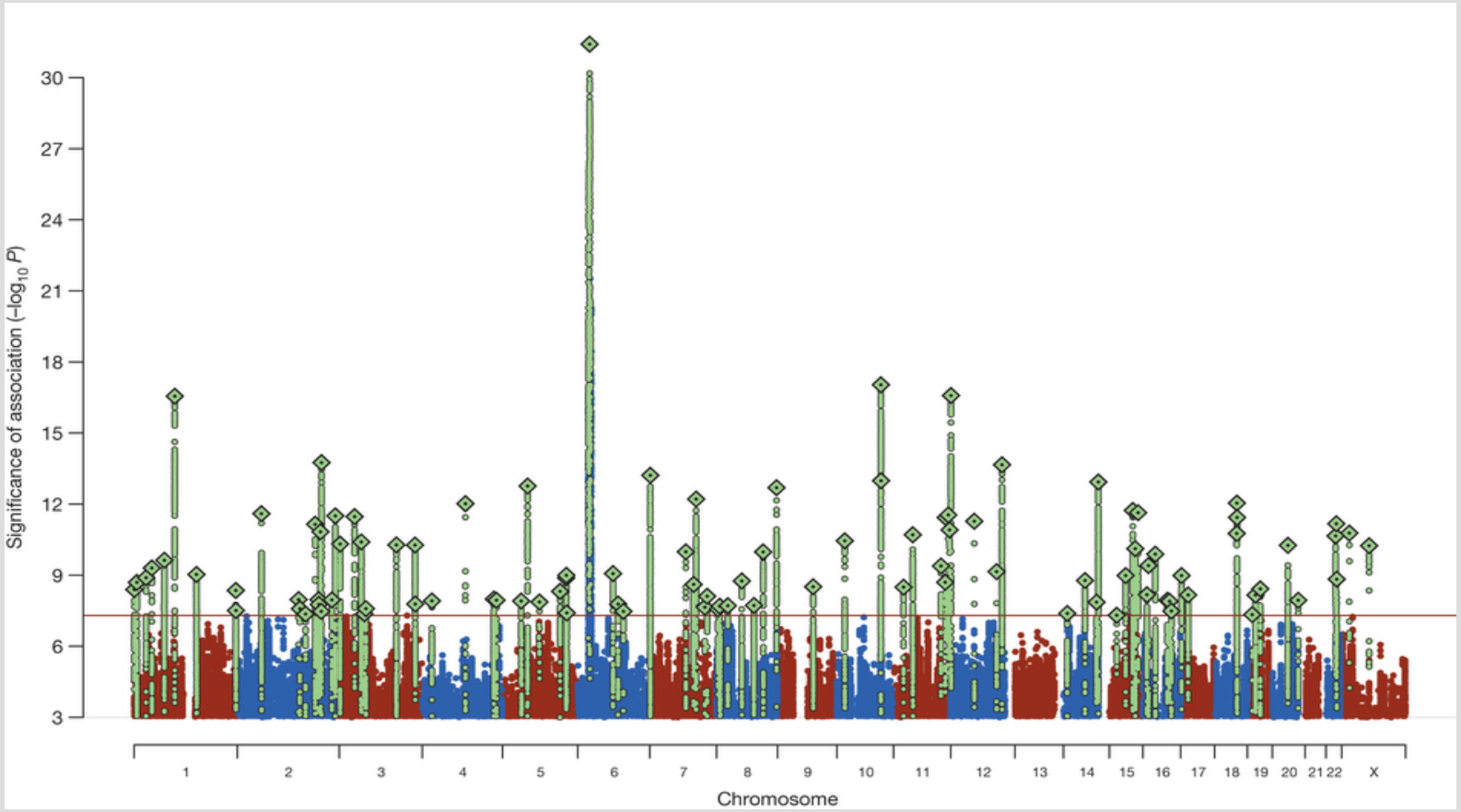
- 9,394 SZ cases and 12,462 controls at 1M SNPs
- 5 GWS loci



SCZ consortium
Nat Gen 2011

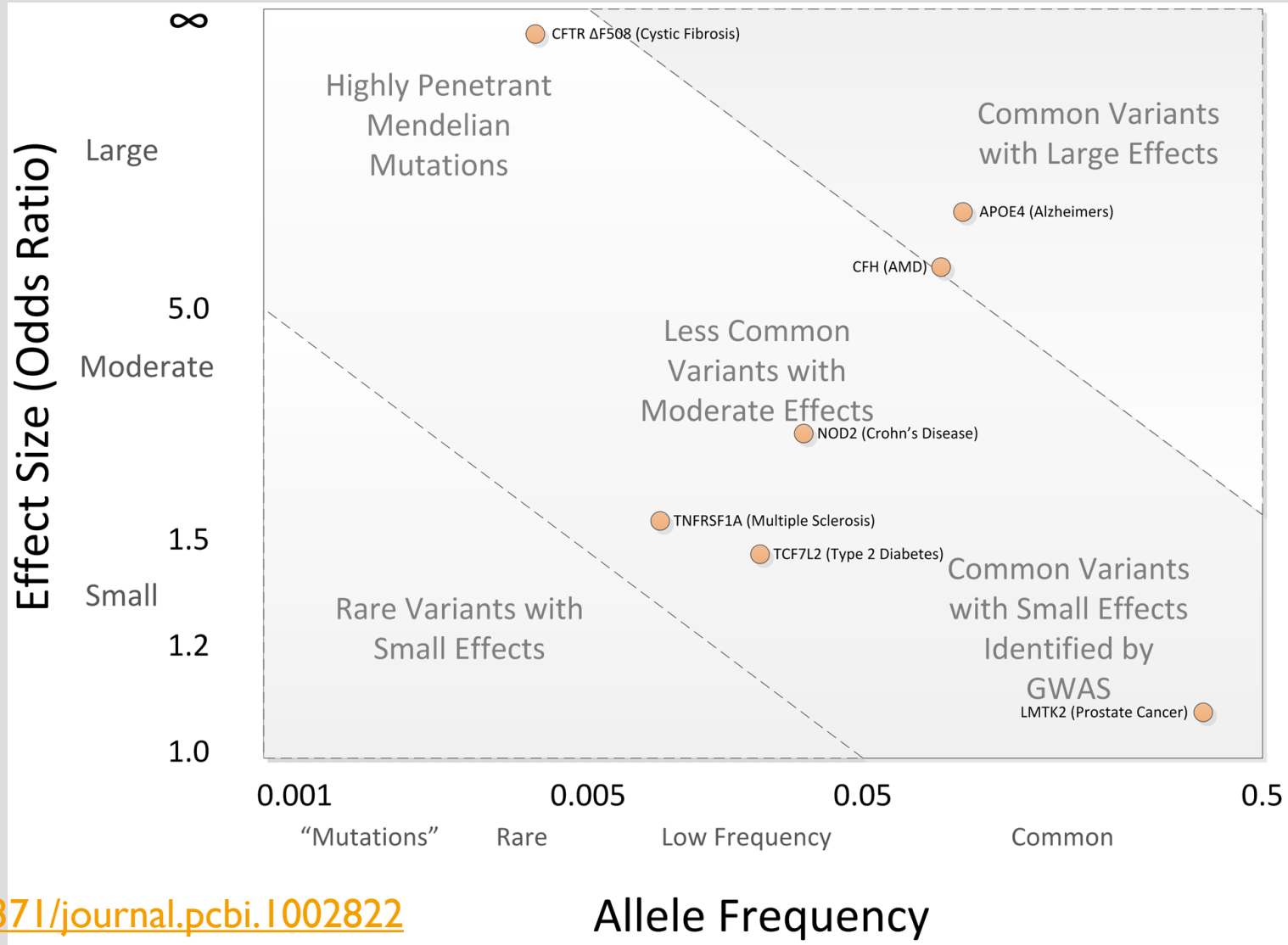
SCHIZOPHRENIA GWAS 3/3 2014

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



Psychiatric genomics consortium
Nature 2014

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.

Bush & Moore
PLoS Genetics
https://doi.org/10.1371/journal.pcbi.1002822